SCIENTIFIC AMERICAN

CAN THE CHEETAH SURVIVE?

$2.50

May 1986

For the Technical Professional:

The story behind IBM's new 32-bit RISC PCs.

These aren't your ordinary workstations.

For IBM RT Personal Computers—6150 and 6151—are built around a revolutionary 32-bit, virtual memory, RISC-based microprocessor. Developed and manufactured by IBM.

The significance? Far faster throughput. Because RISC (for Reduced Instruction Set Computer) architecture eliminates many of the unnecessary bells and whistles found in conventional chip design.

What's more, an IBM RT PC can directly address up to four megabytes (Mb) of real memory, over a trillion bytes of virtual memory and up to 210 Mb of DASD, or disk storage.

All of which makes IBM RT PC microprocessors particularly well suited for the total application requirements of technical professionals working in engineering, scientific, industrial and academic environments.

The IBM RT PC 6150 Model 20
Floor-standing unit

CADAM® and CIEDS[1] Software

One of the key programs available for the IBM RT PC—attached to the IBM 5080 Graphics System—is Professional CADAM. And it's compatible with established host-based CADAM.

There's also Computer-Integrated Electrical Design Series (CIEDS), which allows for the schematic capture of integrated circuit designs.

And there is a growing list of applications available from software companies in the following fields: mechanical and electrical CAD/CAM, petroleum production, artificial intelligence, software engineering, project management and graphics.

UNIX[2] System V

The primary operating system is derived from the UNIX System V—with some very important enhancements added by IBM. These enhancements make this operating system—Advanced Interactive Executive or AIX[3]—much easier to use, give it improved performance and reliability, and also add a virtual storage capability.

This operating system gives you multi-user and multi-tasking operations for up to eight concurrent local or remote users working on low-cost ASCII terminals such as the IBM 3161.

There is also a full-function IBM SQL/RT relational data base management facility available.

Understands Many Languages

The IBM RT PCs are highly literate. You'll be able to write and run applications in C, Fortran 77, BASIC, PASCAL and Assembler.

And as for general-purpose applications, you'll be able to run UNIRAS[4] (for engineering and business graphics), RS/1[5] (for data analysis and graphing functions), IMSL[6] (a library of Fortran subroutines), Interleaf's Workstation Publishing Software[7], Solomon III[8] (for accounting), SAMNA+[9] (for spreadsheet and word processing) and Applix IA[10] (for integrated office functions).

PC Compatible

The IBM RT PCs can incorporate as an optional feature an Intel 80286 coprocessor (in effect a PC AT on-a-board) that runs concurrently with the 32-bit RISC processor. That means the RT PCs can support many existing IBM PC applications and the interchange of their related data. It also means the RT PC I/O bus can accept standard PC cards.

Contact your IBM marketing representative. Or call 1 800 IBM-2468, Ext. CZ/428, for participating Authorized Dealers, or for free literature. Software companies interested in writing programs for the IBM RT PC should call 1 203 783-7141.

IBM®

IBM 5-86
DRM, Dept. CZ/428
101 Paragon Drive, Montvale, NJ 07645

☐ Please send me information on the IBM RT PCs.
☐ Please have an IBM marketing representative call me.

Name_____ Title_____
Company_____
Address_____
City_____ State_____ Zip_____
Phone_____

The IBM RT PC 6151 Model 10
Desk-top unit

CADAM® is a registered trademark of CADAM INC. 1. CIEDS is a trademark of the International Business Machines Corp. 2. UNIX is a trademark of AT&T Laboratories. 3. AIX is a trademark of the International Business Machines Corp. 4. UNIRAS is a trademark of UNIRAS Incorporated. 5. RS/1 is a trademark of Bolt, Beranek and Newman, Inc. 6. IMSL is a trademark of IMSL, Incorporated. 7. Workstation Publishing Software and Interleaf are trademarks of Interleaf, Inc. 8. SOLOMON III is a trademark of TLB Incorporated. 9. SAMNA+ is a trademark of SAMNA Corporation. 10. Applix IA is a trademark of Applix, Incorporated.

For gift delivery anywhere call 1-800-CHEER-UP (except where prohibited by law).
Product of France. Made with fine cognac brandy 80 proof. ©1986 Carillon Importers, Ltd., Teaneck, N.J.

Established 1845 **SCIENTIFIC AMERICAN** May 1986 Volume 254 Number 5

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.

GMC TRUCK

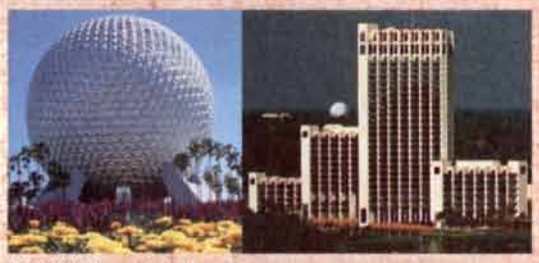
GREAT ESCAPE SWEEPSTAKES

At last, a sweepstakes for individuals who'd never dream of entering one.

Many contests try to give away things you probably wouldn't want at any price. But our GMC Truck Great Escape Sweepstakes offers something you can't put a price on: Escape. It's a chance to win the freedom to get away and experience the great American outdoors. And the opportunity to take off in a brand-new GMC truck. Even if you haven't entered the more mundane sorts of sweepstakes before, you should consider the unusually attractive opportunities presented by the GMC Truck Great Escape Sweepstakes:

GRAND PRIZE: A 1986 GMC SAFARI VAN, YOUR CHOICE OF A REGAL SEBRING 195XL BOWRIDER BOAT OR A 1986 STARCRAFT 20-FT TRAVEL TRAILER, PLUS A GREAT ESCAPE FOR FOUR.* You could win a stylish and accommodating 1986 GMC Safari with the towing capacity no front-wheel-drive mini-van can match, plus your choice of a Regal Sebring 195XL bowrider boat or a 1986 Starcraft 20-ft travel trailer, plus a ten-day, nine-night trip for four to the Grand Canyon, Yellowstone and Yosemite National Parks. Air fare, lodging, $1,000 spending money, and use of a vehicle at each site included.

FIRST PRIZE: 1986 GMC SAFARI VAN, PLUS YOUR CHOICE OF A REGAL SEBRING 195XL BOWRIDER BOAT OR A 1986 STARCRAFT 20-FT TRAVEL TRAILER.*



©1986 Walt Disney Productions.

TWO SECOND PRIZES: TRIP FOR FOUR TO THE WALT DISNEY WORLD® RESORT.*
Four days, three nights of sightseeing, adventuring, relaxing. Includes lodging at the Buena Vista Palace, inside the Walt Disney World Village, and three-day passports to the Magic Kingdom® and EPCOT® Center.

PLUS TEN THIRD PRIZES OF COLEMAN CAMPING OUTFITS AND THOUSANDS OF INSTANT WINNER PRIZES, TOO!

FOUR BONUS PRIZES FOR DEALERSHIP ENTRIES ONLY: A 1986 GMC LIGHT-DUTY TRUCK, PLUS YOUR CHOICE OF A STARCRAFT BOAT OR A 1986 STARCRAFT CAMPER.*
Only those who register at their GMC Truck dealer are eligible for this great bonus: a choice of a 1986 GMC Safari, S-15 Pickup, S-15 Jimmy or Full-Size Pickup, plus a choice of a 1986 Starcraft 24-ft Galaxy camping trailer or a Starcraft 16-ft Medalist boat and trailer.

*Prizes may vary from illustrations shown due to model changes, availability, and other circumstances. See your participating GMC Truck dealer for entry forms, official rules and details. No purchase necessary. Must be 18 or older to enter. Enter at dealer by May 31, 1986. Void where prohibited.

Let's get it together . . . buckle up.

GM

THE COVER

The painting on the cover is a portrait of the cheetah, which in spite of having evolved into the world's fastest land animal is in danger of becoming extinct (see "The Cheetah in Genetic Peril," by Stephen J. O'Brien, David E. Wildt and Mitchell Bush, page 84). Once there were several species of cheetahs; the present-day species, *Acinonyx jubatus,* ranged throughout the world. Now fewer than 20,000 cheetahs are thought to survive; their range is limited to central and southern Africa. One cause of the cheetah's decline appears to be its genetic makeup: new evidence indicates that the cheetah of southern Africa—and perhaps of all Africa—is an inbred species with remarkably little genetic variety. Populations that lack genetic plasticity are at a disadvantage when it comes to surviving periods of ecological change.

THE ILLUSTRATIONS

Cover painting by Marvin Mattelson

Page	Source
16–22	Ilil Arbel
38	Carol Donner
39–42	Edward Bell
43	Carol Donner and Edward Bell
45	Steven B. Levery, Fred Hutchinson Cancer Research Center, and Ronald E. Stenkamp and Keith D. Watenpaugh, University of Washington
46–50	Hank Iken, Walken Graphics
51	Steven B. Levery, Fred Hutchinson Cancer Research Center, and Ronald E. Stenkamp and Keith D. Watenpaugh, University of Washington
52	Yasuo Fukushi, Tohoku University, Japan
53	Hank Iken, Walken Graphics
55	U.S. Geological Survey
56–61	Ian Worpole
62	Michael H. Carr, U.S. Geological Survey
75–76	Quesada/Burke
77–80	Gabor Kiss
81	Sandia National Laboratories
85	Kathrine and Karl Ammann
86–87	Patricia J. Wynne
88–89	Stephen J. O'Brien, National Cancer Institute (*top*); Patricia J. Wynne (*bottom*)
90	Stephen J. O'Brien, National Cancer Institute
91–92	Patricia J. Wynne
96	California Institute of Technology
98	AT&T Bell Laboratories
99	Andrew Christie
101	National Optical Astronomy Observatories (*right*)
102	Thomas G. Phillips, California Institute of Technology
105	Timo Syrjänen, by permission of the National Museum, Helsinki
106–108	Andrew Tomko
110	John Moss
112	John Moss, Scarborough Borough Council
114–115	Andrew Tomko
117	Clifford D. Watson
118–119	Andrew Tomko
120	Sally Black
121	Andrew Tomko
122	Cambridge University Library
125–126	Michael Goodman

LETTERS

To the Editors:

Leon Kamin's ad hominem attack on our book *Crime and Human Nature* [SCIENTIFIC AMERICAN, February] exemplifies the selective use of citation and the political motivation that he wrongly accuses us of displaying. For example:

A major theme of the book is that individuals differ, in part for biological reasons, in their predisposition toward high rates of certain kinds of common crimes. The evidence comes chiefly from numerous studies of twins and of adopted children carried out here and abroad by many different scholars, presumably of many different political persuasions. Kamin dismisses this evidence with the airy phrase that criticisms of such work are "numbingly familiar" and make any findings about the genetic contribution to any form of human behavior (such as intelligence) "wholly inconclusive." Quite the contrary: it is only Kamin and a handful of other professors sharing his radical perspective who find such evidence inconclusive. The overwhelming majority of specialists in the various topics have long accepted the existence of genetic influences on intelligence, alcoholism, schizophrenia and some learning disabilities. Certain forms of aggressive and criminal behavior now seem to be liable to such influences as well. Kamin nowhere makes a serious argument against the evidence.

Another theme of our book is that biology is not destiny: innate predispositions are powerfully affected by early family experiences and probably also by schooling, peer associations and expectations of legitimate employment. The interactions are pervasive, and no one can measure the relative significance of each causal factor with the evidence in hand. Kamin refuses to believe we mean this; he insists on asserting that we really mean genes and environment are "radically separate sources of causation." Since we repeatedly say the opposite, and since we devote more than two-thirds of the book to nonbiological factors, why does Kamin misrepresent us?

The reason, of course, is that he believes we are agents of a Reaganite "ideological bent" inclined to "locate the causes of social tensions in genes." This is patent nonsense, for which he can adduce not a shred of evidence. Whatever Mr. Reagan's view may be, ours is that a better understanding of the sources of human behavior is liberating, not oppressive. If Kamin controlled the research agenda of social

science, we might still be blaming insanity on the devil and confusing dyslexia with stupidity. Informed by research findings of the kind Kamin denounces as supportive of a "muddled ideology of biological determinism," one of us (Herrnstein) has been engaged in an effort to improve educational attainment among children in a developing country by supplying their teachers with better instructional materials, and the other (Wilson) has been part of an effort to extend, improve and refine preschool education for disadvantaged children.

The rest of Kamin's criticisms involve, in part, the energetic picking of nits drawn from fewer than a dozen of the more than 1,000 studies cited in our book. The criticisms also involve errors, confusions and misrepresentations. His fantastic claim that there is no credible evidence that high-rate offenders have distinctive personality traits, for example, ignores entire literatures dealing both with standardized scores on personality tests and with clinical observations of psychopathy, the character disorder that most closely approximates the criminogenic personality. The very existence of personality traits is denied. "There is no evidence for (and considerable evidence against) a general personality trait," he asserts, as if, with a stroke of his pencil, he can refute not only the findings of legions of researchers but also the universal experience of ordinary people. His flippant dismissal of the evidence of a low or moderate, but statistically significant, correlation between body type and criminal behavior neglects the fact that no relevant empirical study violates the conclusion and numerous studies substantiate it, even when the finding was inadvertent.

Kamin's underlying thesis is that social science itself is largely worthless, a mere "stew of conflicting and often nonsensical experimental results" from which the effort to learn about the social world "is to make wine out of water." Hence, he concludes, our goal is not science but politics. He is wrong. Our goal is science. We believe, moreover, that politics informed by science is much to be preferred to science contaminated by politics.

Given space and time, we could dispose of all Kamin's criticisms, for they are groundless. The one thing we cannot fathom is why *Scientific American* asked him to review this book. It is not much different from having Trofim Lysenko review a textbook on modern biology.

JAMES Q. WILSON

RICHARD J. HERRNSTEIN

To the Editors:

I was pleased to learn from their letter that Professors Wilson and Herrnstein have each engaged in efforts to improve education for the disadvantaged. I should like to believe it was their absorption in such worthwhile activity that prevented them from taking the time to prepare a serious reply to my criticism of their book.

They point out that a "theme" of their book is that "biology is not destiny: innate predispositions are powerfully affected by early family experiences and probably also by schooling, peer associations and expectations of legitimate employment." I was not the only reviewer to miss that theme. The book was reviewed in the Winter 1986 issue of *Policy Review* by the mayor of New York, Edward I. Koch, under the all too evocative title "The Mugger and His Genes." Mayor Koch declared that the book's "central thesis is that certain individual biological—indeed genetic—traits, when combined with an uncertain moral environment, produce criminal behavior. Moreover, these traits can barely be changed, if at all. ... The authors find only marginal roles for schools, neighborhoods, peer group values, television violence, and job market conditions as causes of crime." Like Wilson and Herrnstein, Mayor Koch does mention the environment. But it is no accident that his review was not titled "The Mugger and His Moral Environment, As Provided by the Spectacle of New York City Officials on the Take"—nor is it accidental that the book was not called "Crime and Human Nurture."

I documented in my review a number of instances in which Wilson and Herrnstein miscited, misrepresented and misunderstood the research literature they claim to be summarizing for their readers. They now assert that my specific examples were a "picking of nits drawn from fewer than a dozen of the more than 1,000 studies cited in our book." Alas, my review had to be of finite length, and I could cite only a few of the errors and misrepresentations with which the book swarms. To prepare the review, I read a few hundred of the studies cited by Wilson and Herrnstein. It is hard for me to believe Wilson and Herrnstein have actually read those papers. The kinds of misrepresentations I did not mention in my review include the description by Wilson and Herrnstein of a control group that did not exist, the reporting of statistically nonsignificant results as if they were significant, the citing of a preliminary study as if it were definitive (without mention of later, larger and better-controlled studies from the same laboratory, which produced contradictory outcomes), and so on. I should be happy to supply precise details to any reader interested enough to write to me. Based on a sample of a few hundred cases, I soberly report my judgment that very few of Wilson and Herrnstein's citations are accurate, and that still fewer are adequate.

I should very much like to reply to Wilson and Herrnstein's claim that my criticisms of their work involve "errors, confusions and misrepresentations," but it isn't easy. They give no specific examples. I won't take the time to rebut such absurdities as their claim that I have argued against the existence of general personality traits, and that I flippantly dismissed evidence concerning a correlation between body build and crime. It seems sufficient merely to refer readers to my original review.

I don't much like being compared to Lysenko, but I'm not surprised, since Herrnstein wrote of me some years ago that I was "surely a cheat." Herrnstein arrived at that uncharitable judgment after considering work in which I demonstrated that data of the late Cyril Burt suggesting a high heritability of I.Q. were fraudulent. Herrnstein attacked the work and wrote that "Burt was ... a person of unimpeachable integrity." I think readers should know that Herrnstein's track record of judgments about data suggesting the inheritance of human behavior is not outstanding. Burt and his data, as is now universally agreed, were fraudulent.

It's bad enough to be likened to Lysenko, but I object even more to the flip side of the authors' analogy—the implication that their selective and inaccurate summary of a stew of conflicting and often nonsensical experimental results is a part of the genuine science of "modern biology." It isn't, and it isn't really social science either.

I did not argue that all social science is worthless, nor did I denigrate efforts to understand the social world. The fact is, however, that many of the studies cited by Wilson and Herrnstein to support their argument are at a level of banality nonspecialist readers cannot imagine. The future of social science will not be advanced by a publish-or-perish academic culture that encourages the continued production of potboilers, nor by pretending that an ideologically motivated summary sketch of such material amounts to a "magisterial survey of the now very extensive literature." The valid insights provided by social science include, among others, Gresham's law. We who practice social science should understand, above all others, that its currency is in danger of being debased.

LEON J. KAMIN

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MAY, 1936: "The 200-inch telescope to be installed on Palomar Mountain is being designed to reach all parts of the heavens seen from our latitude. In view of its great light-gathering power, one of its most important tasks will be the study of the structure and radial motion of the extra-galactic nebulae, especially in their bearing on the still open question of the nature of the 'expanding universe.' The nearer large spirals, such as the Great Nebula in Andromeda, offer many problems for detailed examination, while the Galaxy within which we live presents scores of questions awaiting their turn for study with powerful instruments."

"Welding is revolutionizing production technique. Within the past half-dozen years it has graduated from primarily a tool of repair to a foremost tool of production. In almost every enterprise that involves joining of metal, welding has been introduced because it permits faster, cheaper and better work and produces a stronger and lighter product. It is welding to which we owe the low-priced automobile without squeaks and rattles, the streamlined train and the railroad track without the clickety-click at rail joints that makes for passenger discomfort and maintenance expense."

"Electron optics is a comparatively recent branch of the science of electronics. This field of study shows that it is possible to shape electrodes in such a way that the electric field between them will act as an 'electron lens,' which is found to have properties almost identical with those of an ordinary glass lens. A typical image tube with a nine-inch viewing screen has a large-aperture lens mounted to image the scene onto the infra-red sensitive cathode. This image is in turn reproduced on the fluorescent screen of the image tube, thus enabling the observer to 'see' by infra-red radiation, even in darkness. Another use is in connection with infra-red microscopy."

"The lisping and hissing effects that have marred the screen voices of many talking-picture stars have been completely eliminated, and the quality of musical reproduction has been immensely improved, by means of a new sound-recording and printing method that utilizes ultra-violet light. The method is said to give such a sharp and clear definition of the peaks and valleys of the high-frequency waves that there are now no practicable limitations in recording and reproducing all the frequencies necessary for faithful reproduction from film."

"Some people think that hatchery trout are as far removed from a real trout as a Caspar Milquetoast is from Mussolini. They deem them nothing but caricatures of the fish that once filled our clean, cold streams and lakes. Anglers need no longer feel too downcast over the passing of wild trout. Hatcheries can, if they follow the right methods, rear trout that are capable of taking care of themselves on being planted and that will be pleasing to the anglers as game fish."

MAY, 1886: "The regular annual meeting of the National Academy of Sciences was held at the Smithsonian Institution on April 20 and for several subsequent days. Professor Samuel P. Langley presented results of his studies on invisible spectra. Professor Hubert A. Newton read a paper on Biela's comet. Professor Ogden N. Rood of Columbia College gave an account of a series of experiments on color contrast. Professor H. A. Rowland of Johns Hopkins University gave a valuable exposition of the absolute and relative wave lengths of lines of the solar spectrum. Professor A. Graham Bell gave a further contribution to the study of heredity in relation to deaf-mutism. Professor Wolcott Gibbs was elected Foreign Secretary in place of Alexander Agassiz, resigned."

"Those who by reason of constitution, habit or ill humor are continually opposing progressive ideas and inventions might learn a profitable lesson, if they would glance for a moment at the history of almost any of our important improvements, and study, through the perspective of several decades, the ungraceful position of those who then maintained a similar attitude toward the advanced projects of their times. When the gas flame was introduced, our ancestors can scarcely be called enthusiastic about the fluid, if a petition of 1833 against lighting Philadelphia with gas represents at all correctly the popular sentiment on the subject. The list of names appended to the petition includes men who were prominent among those most noted for their intelligence. Unreasonable as these apprehensions now appear, they were scarcely comparable with those excited a few years previously by Stephenson's newly invented locomotive."

"The report of Dr. John S. Billings, Surgeon of the United States Army, on the mortality and vital statistics of the population shows the death rate to have been larger in the colored than in the white population, and among the latter higher in the foreign element than among those of American parentage. The death rate was also greater in cities than in rural districts. The most important causes of disease and death were consumption, pneumonia, diphtheria, typhoid fever, malarial fever and the various ill-defined forms of attack to which children under one year of age are particularly subject."

"The consumption of mineral waters in the United States is annually increasing, and the utilization of our native springs is receiving greater attention each year. It is quite safe to predict that many a lonely spring, which is now tasted only by a passing frontiersman or a thirsty Indian, will some day become important."

"From a table recently prepared by the Director of the Mint it appears that during 1883 a total of $14,459,-464 worth of gold was utilized in the arts and manufactures of the United States. Of this amount, $7,905,163 was used in jewelry and watches; $3,598,-308 for watch cases; $1,084,824 in gold leaf; $827,000 for watch chains, and the remainder in smaller sums for dental supplies, pens, instruments, plate, spectacles, chemicals and jewelers' supplies. During the same period a total of $5,556,530 worth of silver was utilized for similar purposes."

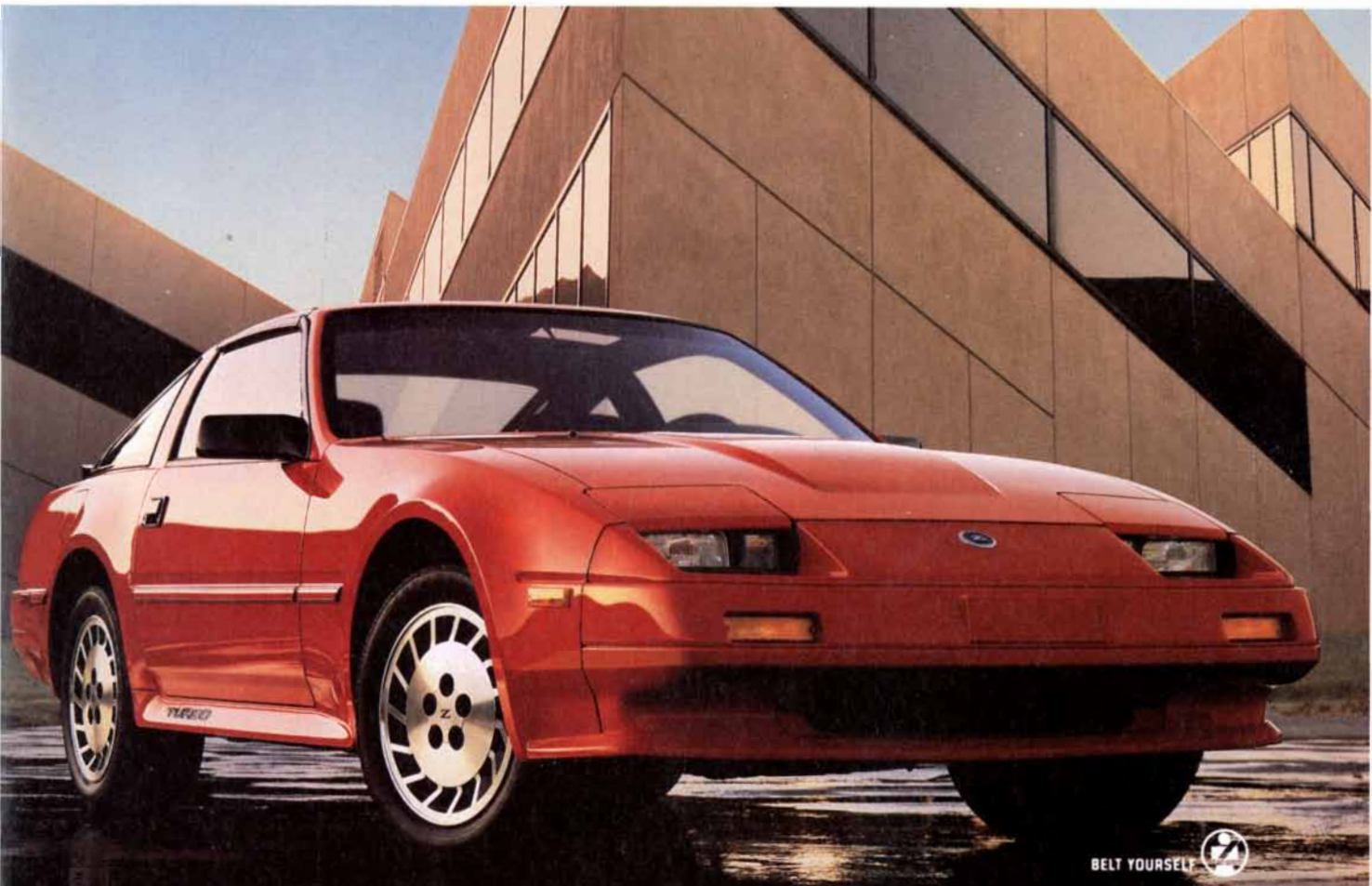
"If any one would accurately gauge his knowledge and learn how little he knows, let him try to teach something. However much he may have studied a subject as a science or practiced it as an art, the limitations of his knowledge will press him closely—all the more closely, perhaps, the more he may have learned, unless he be content to pour old ignorance from vase to vase and call it learning."

"On March 6 the active volcano of Kilauea in the Sandwich Islands sank from the bed of the crater, leaving a bottomless abyss about four miles in circumference. The volcanic activity was utterly extinguished."

THE NEW NISSAN 300 ZX

PACE CAR FOR THE PERFORMANCE GENERATION.

NISSAN



Very few automobiles have generated the excitement and emotional involvement of the Z-car. It was not surprising then, to read that Motor Trend Magazine called the 300 ZX, "the best all-around Z-car ever built."

And now, Nissan has taken one more step in the thoughtful evolution of this classic.

At the heart of this Z is one of the most technologically advanced, 3-liter, V-6 engines in the world. Track proven, this sophisticated design includes computerized electronic multi-port fuel injection and high-power, hemi-shaped combustion chambers. The turbo model puts out a rousing 200 horsepower. When you combine the rigid, lightweight powerplant of the 300 ZX with electronically adjustable shocks, the result is startling.

Outside, fender flares, housing wider tires, are integrated into the body. The air dam is extended and rocker panel extensions reduce air turbulence. All this, plus a wider track results in better handling than ever.

Inside, a choice of electronic or analog instrumentation 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.

THE NAME IS NISSAN

E MANY, NOT JUST THE FEW.

FLOATING POINT SYSTEMS PUTS THE POWER OF SUPERCOMPUTING AT THE DEPARTMENT LEVEL.

Productivity. That's why you're investing in computers at the department level. And why it makes sense to add an FPS 64-bit supercomputer to your existing system.

FPS supercomputers begin at well under $1 million and expand to peak speeds of over 340 MFLOPS.

They protect your FORTRAN investment. Provide vast applications software. And are ready on cue, not *in* queue.

Learn how Floating Point Systems has helped more than 300 international corporations turn on the power of high-speed computing. Contact FPS at 1-800-635-0938.

FPS FLOATING POINT SYSTEMS

THE PROVEN POWER IN ACCESSIBLE HIGH-SPEED COMPUTATION.

SCIENCE/SCOPE®

Electronics engineers and doctors have created an implant to improve the hearing of profoundly deaf people. The device uses a microphone and pocket-sized signal processor to convert sound waves to electrical impulses. A transmitter behind the ear sends these electrical signals through the skin to a nickel-sized receiver/stimulator implanted in the mastoid cavity. From there, signals travel along an electrode array threaded to the snail-like cholea of the inner ear. Auditory nerve fibers receive these electronic signals and transmit them to the brain—tricking it into thinking the ear can hear normally. The electrode of the electronic ear is protected by a plastic coating material, which must provide electrical reliability and long life. Since most polymer insulations tend to lose adhesion in body fluids after a period of time, the National Institutes of Health has asked Hughes Aircraft Company to subject three coating materials to intensive testing. One plastic was created by Hughes, originally for high-temperature electronic and structural uses; the others were developed by Union Carbide and Dupont.

The last in a series of three spacecraft for AT&T is carrying telecommunications throughout the United States. Telstar 3B can handle 21,600 long-distance telephone calls simultaneously or a mix of voice, video, and high-speed data services. It carries 24 transponder channels, which can be switched by ground command to various combinations of regional coverage across the continental U.S. and Hawaii, Alaska, or Puerto Rico. The cylindrical satellite measures 7 feet 2 inches in diameter and 22 feet 5 inches high and is based on the Hughes HS 376 model. It is the world's most successful commercial communications satellite, with 30 spacecraft ordered to date.

A cryogenic refrigerator designed to cool infrared sensors has passed a test equivalent to operating three years in space. The Vuilleumier cycle cooler, set in operation at twice its normal speed in order to simulate a design life of five years, has passed the year-and-a-half point of flawless operation. The device will be used with infrared sensors in space for applications such as defense and geological surveys. The sensors must be chilled to near absolute zero to maintain adequate sensitivity to low-temperature thermal radiation. The VM cooler, developed by Hughes, is believed to be the only one of its type to have performed this long at such low temperatures.

An experimental digital-to-analog converter chip is 10 times faster than the fastest conventional device. The chip, being developed at Hughes for advanced airborne radars, uses gallium arsenide as the substrate material. It has a settling time of 200 picoseconds, about an order of magnitude faster than a record-holding 6-bit Hughes silicon device. The new converter so far outdistances commercial devices that design engineers are developing special interfaces so that the device can be hooked up in data conversion systems for further testing and analysis.

Hughes Ground Systems Group is applying its airspace management experience to the exciting challenges of worldwide air traffic control. These systems will be designed to ensure service 24 hours a day, 7 days a week. They will support distribution of processing among multiple computers linked via local area networks. The many challenges include design and development of hardware and software to support advanced display and man-machine interface technology, and using satellite technologies for future ATC applications. To help design the next generation of air traffic control systems, send your resume to Hughes Ground Systems Group, Employment Dept. S2, P.O. Box 4275, Fullerton, CA 92634. Equal opportunity employer. U.S. citizenship required.

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

HUGHES
AIRCRAFT COMPANY

Subsidiary of GM Hughes Electronics

THE AUTHORS

MICKEY S. EISENBERG, LAWRENCE BERGNER, ALFRED P. HALLSTROM and RICHARD O. CUMMINS ("SuddenCardiacDeath") have been actively involved in emergency medical service programs in the Seattle area. Eisenberg, associate professor of medicine and adjunct associate professor of epidemiology at the University of Washington School of Medicine and director of the University Hospital's Emergency Medicine Service, is medical adviser to the emergency medical services division of the Seattle–King County Department of Health. He holds a bachelor's degree from the University of Michigan, a medical degree from the Case Western Reserve University School of Medicine and a master's degree and doctorate from the University of Washington School of Public Health and Community Medicine. Bergner has a B.A. from Brooklyn College of the City University of New York, an M.D. from the New York University School of Medicine and an M.P.H. from the University of California at Berkeley School of Public Health. From 1972 to 1980 Bergner was director of the Seattle–King County Department of Health; now he is in the cancer-control applications branch of the National Cancer Institute. Hallstrom is research professor in the department of biostatistics at the University of Washington. He studied mathematics at Brown University and has taught mathematics and statistics at a number of institutions. Since 1974 he has applied biostatistical techniques to study the epidemiology and treatment of cardiovascular diseases. Cummins is assistant professor of medicine at the University of Washington School of Medicine and an attending physician in the University Hospital's Emergency Medicine Service. He has a bachelor's degree from the University of North Carolina and a medical degree from the Case Western Reserve University School of Medicine. Cummins is medical director of programs in manual and automatic defibrillation for emergency medical technicians.

SEN-ITIROH HAKOMORI ("Glycosphingolypids") is head of the program in biochemical oncology at the Fred Hutchinson Cancer Research Center in Seattle; he is also both professor of microbiology in the School of Medicine and professor of pathobiology in the School of Public Health and Community Medicine at the University of Washington. His degrees are from Tohoku University in Japan: an M.D., earned in 1952, from its medical college, and a D.Med.Sci., granted in 1956 by its biochemical institute. After a research fellowship at the Massachusetts General Hospital and the Harvard Medical School, he joined the faculty at Tohoku. In 1963 he returned to Massachusetts General, and three years later he was appointed visiting assistant professor at Brandeis University. In 1968 Hakomori moved to the University of Washington, becoming full professor in 1971; he joined the Hutchinson Center in 1975.

ROBERT M. HABERLE ("The Climate of Mars") is a research scientist in the theoretical-studies branch of the space sciences division at the National Aeronautics and Space Administration's Ames Research Center. He came to science relatively late in life, having served in the infantry in Vietnam during the late 1960's before earning his B.S. and M.S. degrees at San Jose State University. After working for several years at Ames he left for the University of Washington, where he got his Ph.D. in 1981. Haberle then returned to Ames as a postdoctoral fellow and took his present post there in 1983.

JAMES H. AUBERT, ANDREW M. KRAYNIK and PETER B. RAND ("Aqueous Foams") work together at the Sandia National Laboratories. Aubert's undergraduate education was at the University of Michigan and his graduate training was at the University of Minnesota. After receiving his Ph.D. in chemical engineering in 1982, Aubert joined the staff at Sandia. Kraynik's degrees are a B.S. (1972) from Carnegie-Mellon University and an M.A. (1973) and a Ph.D. (1976) from Princeton University. Since completing his university education he has been at Sandia. Rand, who serves as consultant on foam applications, has more than 20 years of practical experience with polymeric and, more recently, aqueous foams. He writes that his chief interest is in developing "oddball" foams for a variety of uses.

STEPHEN J. O'BRIEN, DAVID E. WILDT and MITCHELL BUSH ("The Cheetah in Genetic Peril") bring their respective training in genetics, reproductive physiology and clinical veterinary medicine to their study of the cheetah. O'Brien, chief of the laboratory of viral carcinogenesis at the National Cancer Institute, earned his Ph.D. (1971) in genetics from Cornell

University. He joined the institute in 1972 to develop a program in the somatic-cell genetics of mammalian systems. Wildt, who has a Ph.D. (1975) from Michigan State University, is a staff physiologist at the National Zoological Park of the Smithsonian Institution, where he directs a research laboratory studying animal reproduction. Previously he taught at the Baylor College of Medicine and worked at the National Institutes of Health. Bush was graduated from the University of California at Davis in 1965 with a D.V.M. degree, and he served on the staff of the Johns Hopkins University School of Medicine before taking his present position as chief of the department of animal health at the National Zoological Park.

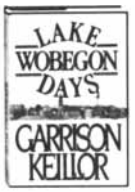
THOMAS G. PHILLIPS and DAVID B. RUTLEDGE ("Superconducting Tunnel Detectors in Radio Astronomy") are respectively professor of physics and associate professor of electrical engineering at the California Institute of Technology. Phillips was graduated from the University of Oxford with a B.A. in physics in 1961. He received his D.Phil. also from Oxford in 1964. In 1968 he moved to Bell Laboratories, where he remained (except for a year as a reader at the University of London) until 1980, when he joined the faculty at Caltech. Rutledge holds a B.A. (1973) from Williams College, an M.A. (1975) from the University of Cambridge and a Ph.D. (1980) from the University of California at Berkeley. After getting his doctorate he joined the faculty at Caltech.

MAREK ZVELEBIL ("Postglacial Foraging in the Forests of Europe") is lecturer in archaeology and prehistory at the University of Sheffield. He got his B.A. at Sheffield in 1974 and went on to pursue graduate studies at the University of Cambridge. He earned his Ph.D. in 1981 for research on Stone Age cultures of the northern Baltic region and joined the Sheffield faculty later that year. Zvelebil's main research interest is the archaeology of hunting-and-gathering cultures.

SANDRA HERBERT ("Darwin as a Geologist") is associate professor of history at the University of Maryland at Baltimore County. She has a B.A. in interdisciplinary studies from Wittenberg University and a Ph.D. in the history of ideas from Brandeis University. Before taking her current position she taught at the University of Maryland at College Park and at Princeton University and also served for a year as a visiting curator at the Smithsonian Institution.

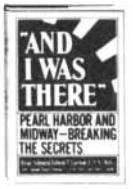
Here's an offer for everyone who has ignored every offer we've ever made.

354
Pub. price **$22.95**



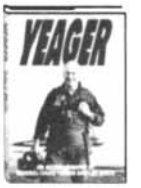
124
Pub. price **$17.95**

320
Pub. price **$27.50**

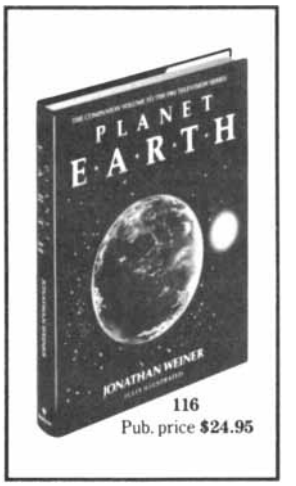


424
Pub. price **$19.95**

313
Pub. price **$19.95**



002
Pub. price **$17.95**



116
Pub. price **$24.95**

218
Pub. price **$18.95**

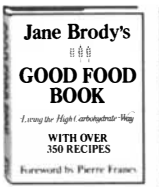
466
Pub. price **$29.95**

266
Pub. price **$18.95**

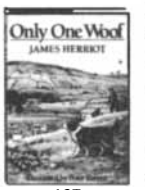
Choose any 2 books for only $1 each with no obligation to buy another book.

513
Pub. price **$24.95**

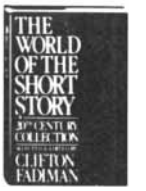
256
Pub. price **$17.95**



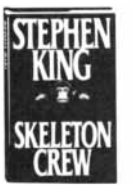
215
Pub. price **$22.95**



167
Pub. price **$9.95**

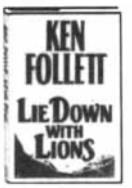


344
Pub. price **$22.95**



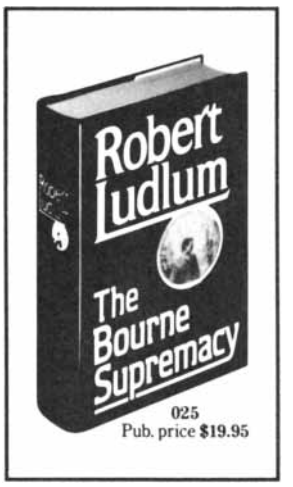
250
Pub. price **$18.95**

519
Pub. price **$19.95**

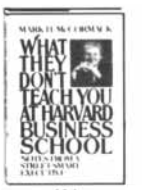


057
Pub. price **$18.95**

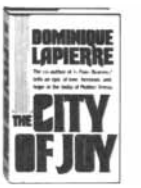
678
Pub. price **$25**



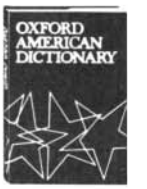
025
Pub. price **$19.95**



414
Pub. price **$16.95**



183
Pub. price **$17.95**

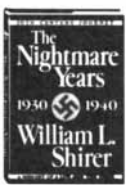


624
Pub. price **$16.95**

261
Pub. price **$18.95**

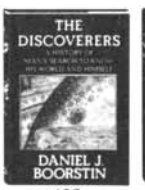
094
Pub. price **$9.95**

194
Pub. price **$16.95**



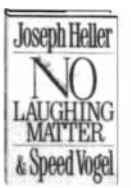
563
Pub. price **$22.50**

440
Pub. price **$17.95**

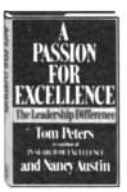


422
Pub. price **$25**

244
Pub. price **$18.95**



177
Pub. price **$18.95**

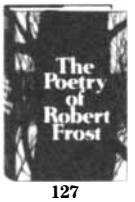


723
Pub. price **$19.95**

553
Price **$14.95**

603
Pub. price **$16.95**

262
Pub. price **$19.95**



127
Pub. price **$17.50**

Benefits of Membership. Membership in the Book-of-the-Month Club begins with your choice of 2 of today's best books for $1 each, plus shipping and handling charges, with no obligation to buy another book. As a member you will receive the *Book-of-the-Month Club News*® 15 times a year (about every 3½ weeks). Every issue reviews a Selection and more than 125 other books that we call Alternates, which are carefully chosen by our editors. If you want the Selection, do nothing. It will be shipped to you automatically. If you want one or more Alternates—or no book at all—indicate your decision on the Reply Form and return it by the specified date. *Return Privilege*: If the *News* is delayed and you receive the Selection without having had 10 days to notify us, you may return it for credit at our expense. *Cancellations*: You may cancel membership at any time by notifying us. We may cancel your membership if you elect not to buy and pay for at least one book in any six-month period.

Book-of-the-Month Club, Inc., **A170-5**
Camp Hill, Pennsylvania 17012

Please enroll me as a member of Book-of-the-Month Club and send me the 2 books I've listed below, billing me $1 each, plus shipping and handling charges. I understand that I am not required to buy another book. However, if I have not bought at least one book in any six-month period, you may cancel my membership. A shipping and handling charge is added to each shipment.

Indicate by number the two books you want. ☐ ☐

Name _____ **6-40**
(Please print plainly)

Address _____ Apt. _____

City _____

State _____ Zip _____

Prices generally higher in Canada.

BOOK-OF-THE-MONTH CLUB®
America's Bookstore® since 1926. © 1986 BOMC

COMPUTER RECREATIONS

Branching phylogenies of the Paleozoic and the fortunes of English family names

by A. K. Dewdney

In southern Ontario and western New York State are some of the finest and most fossiliferous exposures of middle Devonian rocks in the world. One of them lies near my home. There in a quarry that teems with ancient life I once met the compound gaze of a trilobite staring at me from a bed of shale. In the delicate process of removing it I heard a shuffling and whirled to discover a bearded man peering over my shoulder. "A marvelous specimen of *Phacops rana,*" he announced. "Too bad they're all extinct." He was a professor of paleontology, and his name was Smith. "Why," I wondered aloud, "did the trilobites become extinct?" Professor Smith gazed into space and then answered my question with another one: "Why is the name Smith so common?"

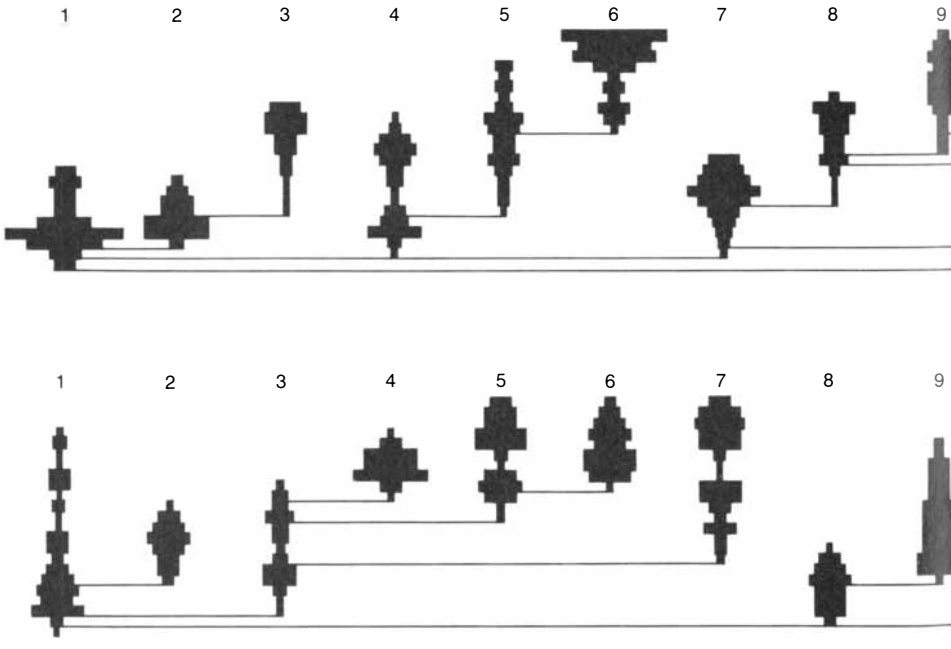
The real answers to both questions would trace a complex and irreversible interplay of genetic and environmental factors. Approximate answers, however, emerge from two computer programs that simulate the evolution of trilobites and names in similar ways. One program tries to mimic the great extinctions of the Paleozoic era. It was originally devised by David M. Raup, a well-known paleontologist at the University of Chicago, and it randomly traces the development of a phylogenetic tree: with each tick of the geologic clock the end of each branch of the tree grows, branches or dies with a certain probability. The random evolutionary pattern generated by the program bears a surprising resemblance to the fossil record, and both show mass extinctions.

The second program imitates the proliferation and extinction of English family names from 1350 to the present. The program is the brainchild of Christopher M. Sturges and Brian C. Haggett, who are both employees of the British Ministry of Defence. Sturges and Haggett trace genealogies in their spare time, and they were puzzled by the gradual disappearance of many names from the available records. Could such a disappearance be a chance event? The two sought to find out by writing a simulation program. They discovered that as many as three-fourths of the family names common in 1350 would disappear by chance alone in the next 636 years.

Any reader who wants to travel in a probabilistic time machine can write a simplified version of each program that nonetheless captures the essential features of the original. In the first program you will traverse hundreds of millions of years as you watch genera or entire phyla of unknown and unnamed creatures diversify or die out. In the second program mere hundreds of years will pass, but the interest is more human. Which family names will triumph, the Smiths or the Smedleys?

The program I call PALEOTREE begins with a single primordial creature. In order to preserve the paleontological flavor of the simulation it seems appropriate to give the creature a scientific name: *Paleoplasmus radiculus.* Readers are free to abuse Latin with equal abandon. For a given, arbitrary time PALEOTREE represents the total population of *P. radiculus* as a single node, or point, on a tree. Every population of creatures belonging to the same biological category as *P. radiculus* and descended from it is also shown as a node; the program makes no attempt to count the number of living specimens in the biological category that are represented by the node at any given time.

In an instant millions of years are allowed to pass, and the program inspects the evolutionary results. Perhaps the old population of creatures retains its biological integrity for another few million years. If it does, the original node gives rise to a new one, connected by a line to its progenitor. Perhaps instead the original population is successful and diversifies into a new biological category. That outcome can be represented as a Y-shaped branch at the original node. One branch of the Y leads to a new node representing the continuation of the original biological category, and the second branch leads to a new node representing the beginning of a new category. The final possibility is that the original population becomes extinct. In that case the original node



Evolutionary development, branching and speciation

simply terminates and no branches lead away from it.

PALEOTREE randomly determines the fate of each node it generates in the tree: is the population to continue unchanged, is it to give rise to a new biological category or is it to become extinct? Taken at face value, however, the three possibilities do not really reflect the rich structure of biological evolution. As I have described the program so far, all the creatures descended from a common ancestor are monophyletic; in other words, they are related closely enough to be put in the same broad biological category, and so they belong to the same tree. Nevertheless, evolutionary development occasionally leads to a new broad category or to a mass extinction when entire categories die out. Taxonomists recognize many such categories, from varieties and subspecies to phyla and kingdoms. For the sake of simplicity PALEOTREE recognizes only two categories, the species and the genus; any other interpretation of the two categories would do as well.

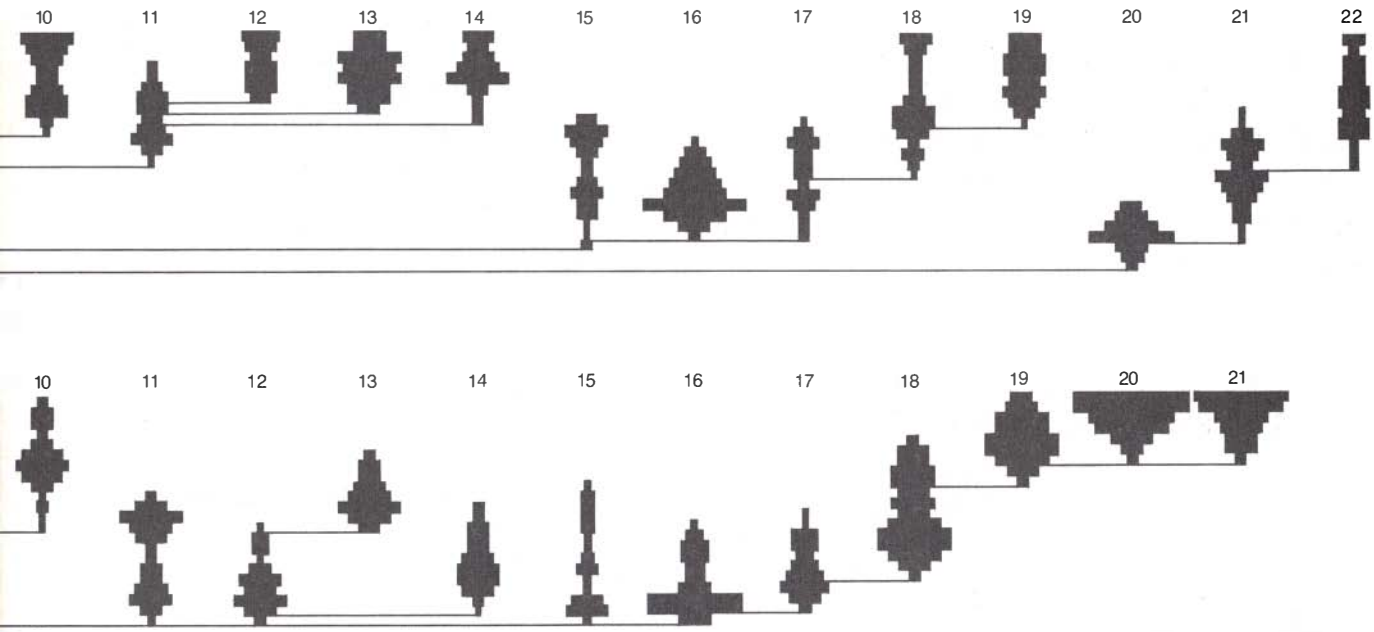
In PALEOTREE a genus can give rise to an entire new genus. The mechanism is simple: a new species branches away from the old genus to found a new genus. What is really just another branch in the evolutionary tree is viewed as a tree in its own right. Of course, a tree (or genus) can also become extinct: all its species die out. The number of species that make up a genus at a given time is called its diversity, which measures not only its genetic richness but also the likelihood that it will survive. There is a pretty method for graphing the results of an experiment with PALEOTREE: each genus is represented by a symmetrical stack of horizontal bars that resembles a spindle. Each bar represents a new epoch, and the length of the bar at a given level of the stack gives the diversity of the genus at the corresponding epoch. The resulting stack reminds me of a bizarre home-lighting display [*see illustration on these two pages*].

It is easy and instructive to generate a tree and its associated spindle by hand. Make a dot on a piece of paper and roll a pair of dice. If the sum of the numbers showing on the dice is 2, 3 or 4, the species immediately becomes extinct. *P. radiculus* is stillborn and no branches emerge from the dot. If the sum is 9 or 10, the species diversifies. Two short lines branch upward from the dot, and a dot is made at the end of each line. If the sum showing on the dice is any other number, the species does not change. A single line sprouts upward and terminates in another dot. For each epoch the experiment is repeated for each species that is not yet extinct. The procedure gives rise to a tree, and the number of dots at any given level of the tree can then be plotted as the length of a bar in a spindle diagram [*see illustration on page 21*].

When the dice are rolled, the probability that a species will become extinct in the next epoch is 1/6, or .167, the probability that it will diversify is 7/36, or .194, and the probability that it will remain unchanged is 23/36, or .639. These probabilities approximate the ones assumed in PALEOTREE, and they give rise to similar trees. Since the probabilities of extinction are somewhat smaller than those of diversification, one might expect some trees to go on forever. Do they? I should be happy to hear from readers willing to stay up all night to find out.

In order to draw a spindle diagram it is not necessary to maintain a tree for each genus. PALEOTREE merely counts the number of species that make up a genus at each epoch, and it keeps a record of the branching between the genera. Both functions can be handled by a double array called *history*. The columns represent genera and the rows represent epochs. Thus *history* (i,j) is the number of species present at time i in the jth genus. When PALEOTREE has simulated evolution for the desired number of epochs, one can print the array *history* and use its entries to draw the spindle diagrams by hand. Readers with some experience in graphics programming might try adding a program that draws the spindles automatically, but a detailed description of that exercise would require a column in its own right.

To run PALEOTREE in its simplest form one must type the number of epochs, or "generations," at the beginning of the program. The number is preserved by the program under the name *numgen*, and it limits the value of the index i for the array *history*. The value of i increases in unit increments from 1 to *numgen* as the program executes its outer loop. A second loop inside the first one is indexed by the val-

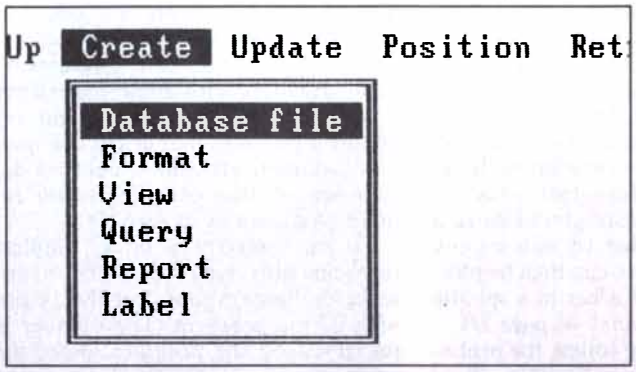


of biological genera, simulated by David M. Raup's version of the computer program PALEOTREE

The database used now be used

Introducing dBASE III® PLUS.

The PLUS stands for all the improvements we've made to the world's number one selling database management software.



The Assistant helps beginning users accomplish day-to-day data management tasks without programming.

Mind you, dBASE III PLUS still has the powerful dBASE programming language, dot prompt, and all the features that have made dBASE III the standard of the industry.

We've simply raised the standard.

And just as dBASE III introduced more power to the people, our new dBASE III PLUS introduces more people to the power.

People who aren't all that crazy about programming, for example.

The Assistant feature in dBASE III PLUS now provides them with new easy-to-use pull-down menus for creating, using and modifying multiple databases.

So now anyone who can manage a simple cursor can manage day-to-day data management tasks. Without programming.

And by using our new Screen Painter, anyone can create custom screens. Without programming.

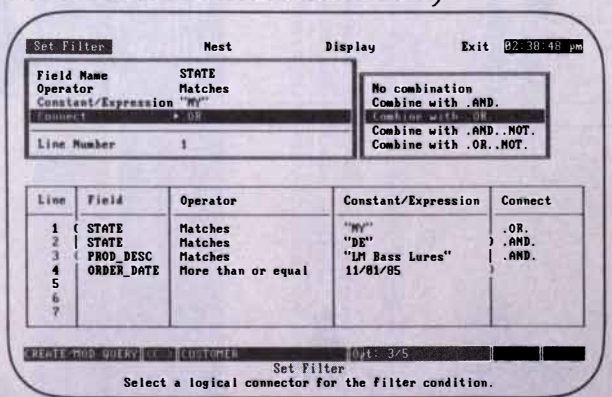
Or using View, access related information in several databases at one time. Without programming.

With Advanced Query System, another new non-programming feature, any user can build complex queries just by selecting from the dBASE III PLUS pull-down menus.

For rapidly creating entire programs, there's even a new Applications Generator.

And for all those who wish to learn to program, the Assistant can be of further assistance. By teaching you programming commands as you go along. Without disrupting your work flow.

These are only a few of the dBASE III PLUS features that can help new users quickly get up to speed. And experienced users quickly increase their speed. (Sorting, for example, is up to two times faster and indexing up to ten times faster than dBASE III.)



Advanced Query System lets you set up and answer complex queries without programming.

by more people can by more people.

And it's the fastest way to network those users, too. Because now, true multi-user capabilities for local area networking are built right in.

dBASE III PLUS can also help put developers in the fast lane. With a new Data Catalog and more than 50 new commands and functions. Plus code encryption and linking, improved debugging aids, assembly language calls and much more.

For the name of the Ashton-Tate® dealer nearest you,* call 800-437-4329, Extension 2844.**

And get your hands on dBASE III PLUS.

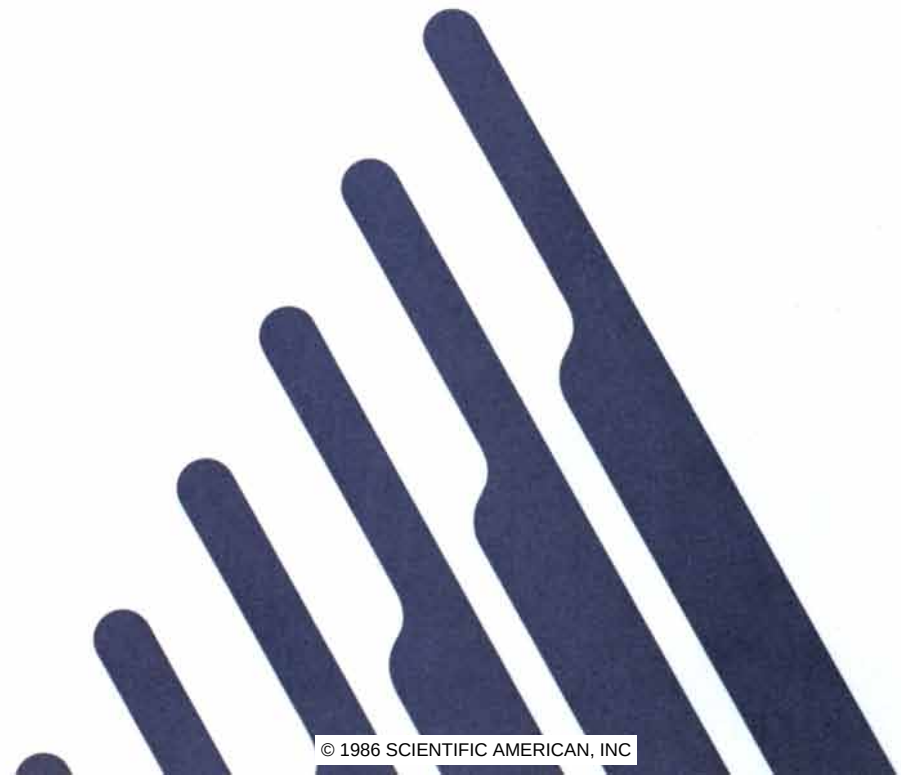
It's the software more people can look forward to using.

*Upgrades are available to all dBASE III owners. **In Colorado call (303) 799-4900, Extension 2844. Requires IBM® PC or 100% compatible. Trademarks/owners: Ashton-Tate, dBASE III PLUS/Ashton-Tate; IBM/International Business Machines Corp.

ASHTON·TATE

dBASE III PLUS

The data management standard.



Buckle up—together we can save lives.

1986 Mercury Sable. You've reshaped your thinking. Obviously, so have we.

Sophisticated new shape. The 1986 Mercury Sable.

The new Mercury Sable is truly a sweeping automotive advance. One that shows we've kept up with the times as much as you have.

You see it in Sable's wind-slippery shape that moves through the air more efficiently than any other domestic four-door sedan.

You see it in Sable's dramatic new "laser" lightbar extending between the flush-mounted aerodynamic headlamps.

Sable has front-wheel drive, of course. And a 3.0-liter V-6 that's new right down to the dipstick.

Inside, everyone will think of Sable as a driver's car. Except the passengers. Controls are positioned right where your hands and feet would like them to be. Backlighted instruments tell you just about everything you ever wanted to know about a car's condition.

Perhaps most important of all, Sable is a Mercury. And Mercury's commitment to quality isn't just something written on paper. It's built into the car.

You're ready for a car that suits today and you, beautifully. And the car you're ready for is ready for you. The 1986 Mercury Sable.

For more information, call toll-free 1-800-822-9292

LINCOLN-MERCURY DIVISION Ford

MERCURY. The shape you want to be in.

ue of j in the array *history;* each value of j corresponds to a genus. A variable called *max* is also defined in order to calculate the limit of the index j. Because only one genus is present at the beginning of a run, *max* is initially set equal to 1. As the outer loop is repeatedly executed the number of genera waxes and wanes. For each new epoch *max* is set equal to the number of genera in the previous epoch. A variable called *lmax* is then set equal to the value of *max* at the beginning of the j loop, and j runs from 1 to *lmax*.

Within the second loop PALEOTREE looks up the number of species currently included in the jth genus, and it stores the number in a temporary variable called *temp*. The program then enters its third and innermost loop, which determines the fate of each species in the jth genus. The index k of the inner loop can be regarded as a counter for the species. It runs from 1 to *temp,* and the value of *temp* is also transferred to the entry *history*$(i + 1,j)$ in the array, which is destined to give the number of species in the jth genus in the next epoch $i + 1$.

For each species, or in other words for each value of k, a random number between 0 and 1 is selected by means of the random-number instruction of one's programming language. If the random number is less than .15, the kth species becomes extinct and *history*$(i + 1,j)$ is decreased by one unit. If the random number is greater than .8, the kth species produces an evolutionary offshoot and *history*$(i + 1,j)$ is increased by one unit. The procedure ensures that species become extinct with probability .15 and that they give rise to new species with probability .2.

The program written by Raup, on which PALEOTREE is based, determined new genera by a form of retrospective analysis that is a bit too complicated to describe here. Raup has noted, however, that reasonably similar results can be obtained more simply: assume each species gives rise to a new genus with the rather slender probability .02. The assumption must be incorporated into the innermost loop. If the random number governing the fate of the kth species is greater than .98, PALEOTREE turns aside for a moment to create a new genus instead of a new species.

Several housekeeping details are important in this part of the program. First *max* is increased by 1 and the ith entry in the new column of the array *history,* namely *history*(i,max), is set equal to the index j of the genus that gave rise to it. In this way the array can store information at the beginning of the newly created genus about the identity of the genus that gave rise to it. The reader (or a plotting program) can thereby determine the origin of each genus. In epoch $i + 1$, its first stage of real evolution, the newly created genus includes just one species. Accordingly the entry *history* $(i + 1,max)$ is set equal to 1, and evolution continues normally thereafter.

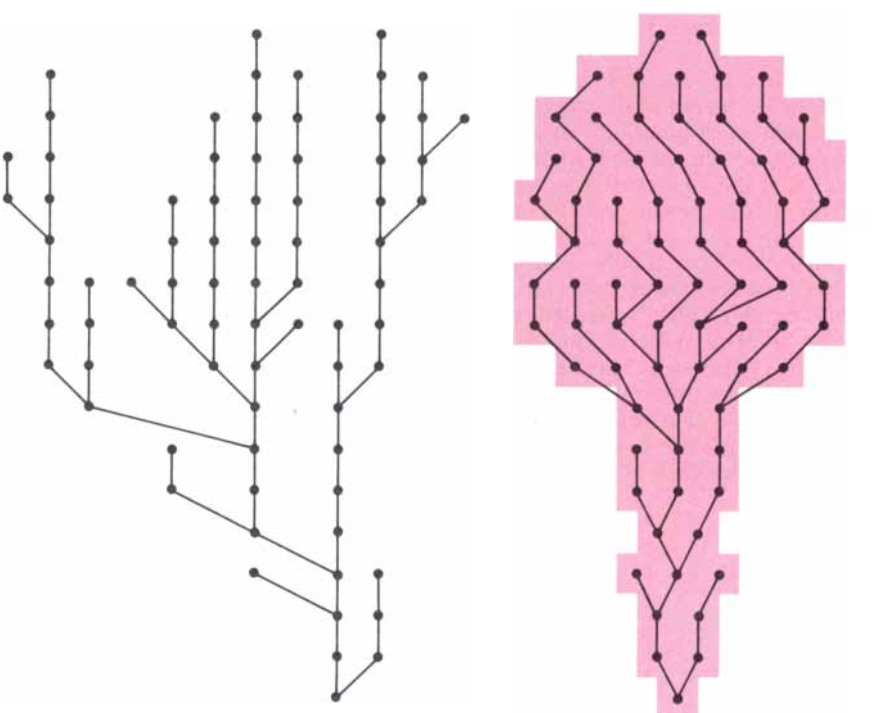
Given the probabilities set up in the program, some genera almost inevitably become extinct. In that case the column of *history* corresponding to a newly extinct genus becomes a 0 and *max* must be decreased by 1. In some programming languages this circumstance can lead to difficulties. Instead of effectively ignoring such a column in the array, the innermost loop may execute once or twice in spite of the fact that *temp* is 0. Some languages may assume such a loop is running decrementally, from 1 to 0, and so the genus may have one or two chances for a miraculous rebirth. To avoid the situation one must insert a test into the beginning of the innermost loop: if *temp* is 0, the rest of the loop must be skipped.

As usual, readers may want to tinker with the programs suggested in this department. For example, how is diversity affected if the probability of extinction within a genus is increased from .15 to .2? Will equal probabilities for branching and extinction cause all life to die out? It will be interesting to hear from readers who encounter mass extinctions or slender survivals.

It should surprise no one that the second program is quite similar to PALEOTREE in structure and function. Remember that it mimics the extinction and survival of family names. Indeed, it can be obtained from PALEOTREE by removing the procedure that creates new genera. I call it NEOTREE.

In NEOTREE an array called *history* also keeps track of events as time goes on. When the program starts, however, *history* holds 1,000 entries instead of only one. The entries represent 1,000 English surnames prevalent in the year 1350. It would be foolhardy (but not without a certain antiquarian charm) to incorporate 1,000 real names into the program. Instead each column of *history* represents a single family whose name unfortunately must be the numerical index of the column. The initial entry in each of the 1,000 columns is 1, which represents one English nuclear family: a mother, a father and a number of children. Actually the program keeps track only of the males, the family members who bear the surname to the next generation.

It would be highly unrealistic to imitate the program PALEOTREE and limit a given line of descent to two new families in each generation. Hence in NEOTREE each family is allowed to generate from zero to six new families. The innermost loop of NEOTREE operates on family names in the same way



Speciation within a genus, determined by the roll of a pair of dice

as the innermost loop of PALEOTREE operates on genera. When NEOTREE considers the *j*th name, for example, the variable *temp* is set equal to the number of families currently bearing the *j*th name. Each family is considered in turn by the innermost loop. A random number then determines the number of marrying males a family will produce.

Sturges and Haggett employ a table derived from a statistical analysis of genealogical records to determine the probabilities for the seven possible fates of family names [*see illustration below*]. For example, the statistical probability is .317 that a family will produce no males destined to marry. The probability is .364 that it will produce one such male, and so on.

The random numbers generated by the program are readily converted into a table of possible outcomes reflecting the probabilities that a family will give rise to various numbers of marrying males. For example, if the random number lies between 0 and .317, the family will have no marrying males. The family (if not the name itself) will become "extinct." If the random number lies in the range between .317 and .681 (the sum of .317 and .364), the family will give rise to one marrying male. In other words, the random number will lie in the second range approximately 364 times out of 1,000, and that proportion matches the probability obtained from the statistical analysis. The rest of the table is constructed in a similar way.

Once NEOTREE has determined the fate of a given family for a new generation it acts exactly like PALEOTREE. At the beginning of the *i*th generation the number of families with surname *j* is stored in the variable *temp*. The index *k* of the innermost loop then runs from 1 to *temp* and *temp* is stored in *history*$(i+1,j)$, just as it is in PALEOTREE. For each family with name *j* that produces no males who marry, *history*$(i+1,j)$ is decreased by one unit, thereby reflecting the fact that one less family will bear the name *j* in the next generation. For each family with name *j* that produces one marrying male *history*$(i+1,j)$ is left unchanged. For each family with name *j* that produces two marrying males the value of *history*$(i+1,j)$ is increased by one unit, and so on.

In a sense NEOTREE and PALEOTREE lead to opposite results. In NEOTREE, for example, count each generation as 28 years; the period from 1350 to 1986 thus requires 22 or 23 generations of computation. What proportion of family names vanish in that time? Sturges and Haggett found that the number of family names steadily diminishes and that nearly three-fourths of the names disappear over the 636-year period. Perhaps some readers will have the tenacity to run NEOTREE long enough to predict the year the original 1,000 names will be reduced to one!

The results of PALEOTREE have been quite different: as one would expect, and in spite of extinctions of the genera, life becomes gradually more diverse. As time passes the net number of genera tends to increase. One would expect some of them to become so filled with species that they would be virtually immune to chance extinction before the universe collapses. Nevertheless, in Raup's initial version of the program there were also an unusually large number of extinctions, and that result was puzzling.

PALEOTREE was developed in the 1970's, when a number of scientists were drawing attention to the apparent evidence for mass extinctions in the fossil record. If the evidence could be believed, such mass extinctions could have been caused by geologic or astrophysical catastrophes. Working with Stephen Jay Gould of Harvard University and others, Raup proposed that a fast simulation program such as PALEOTREE could serve as a null hypothesis: if it generated mass extinctions similar to the ones observed, there would be less urgency to explain them as catastrophic effects.

In fairness to Raup I must point out that before he wrote the program he had no reason to doubt the existence of mass extinctions. He expected to find little or no correlation between the results of his program and the broad fluctuations in the fossil record. Actually, however, he found the correlation to be quite high. Some phylogenetic lines generated by the program branched and broadened, whereas others petered out, just as they do in the fossil record. There were even occasions when several lines became extinct at the same time. What need was there to invoke an ice age or the collision of an asteroid if chance alone could explain the record of abrupt disappearances in nature?

Steven M. Stanley of Johns Hopkins University made a telling criticism of Raup's interpretation of his results soon after they were published. Stanley pointed out that the number of species the program allowed in a genus was too small to be realistic. The smaller the number of species in a genus is, the more likely it is that probabilistic fluctuations will lead to the genus's extinction. In a real genus the number of species is generally high enough to withstand such a fluctuation. Hence in spite of the mass extinctions that arose in Raup's simulation, the ones in the fossil record might not be a statistical artifact. Raup immediately recognized the validity of Stanley's criticism and now supports the view that mass extinctions were probably not the result of random change.

One reason for the low species diversity found in Raup's results was that the program initially declared rather small arrays; the arrays artificially limited each genus to at most 40 species. As Raup puts it, "We didn't realize at the time how important size would be." His program did not lump more than one species into a single array entry as PALEOTREE does because he wanted to be able to reconstruct the entire evolutionary tree for each species and then inspect the process of random speciation closely.

Another reason for the low species diversity is that both Raup's program and PALEOTREE begin with a single species and build up from it. Such a model might best approximate the evolution of very early life, when the diversity of species was low. Readers who wish to see Stanley's criticism in action should initially supply *history* with 100 genera, each having 100 species. Are mass extinctions still encountered? I shall report unusual events, whether for trilobites or for Smiths.

NUMBER OF MALES WHO MARRY	0	1	2	3	4	5	6
PROBABILITY	.317	.364	.209	.080	.023	.005	.001

Statistical summary of English marrying males since 1350

In last month's column I mentioned Magi's four-dimensional dementia. My algorithmic assistant professed to have slipped into the fourth dimension and to have discovered some remarkable properties of hypercubes. How long, he asked, is the diagonal of an *n*-dimensional hypercube? What happens when an *n*-dimensional hypersphere is inscribed in an *n*-dimensional

BREAKTHROUGH: ADD SOUND TO JET ENGINES TO MAKE AIRPLANES QUIETER.

Two airstreams exit from a jet engine in operation–a hot, fast central core and a cooler, slower outer layer. When they mix, they become chaotic. This turbulence causes an unpleasant low-frequency rumble.

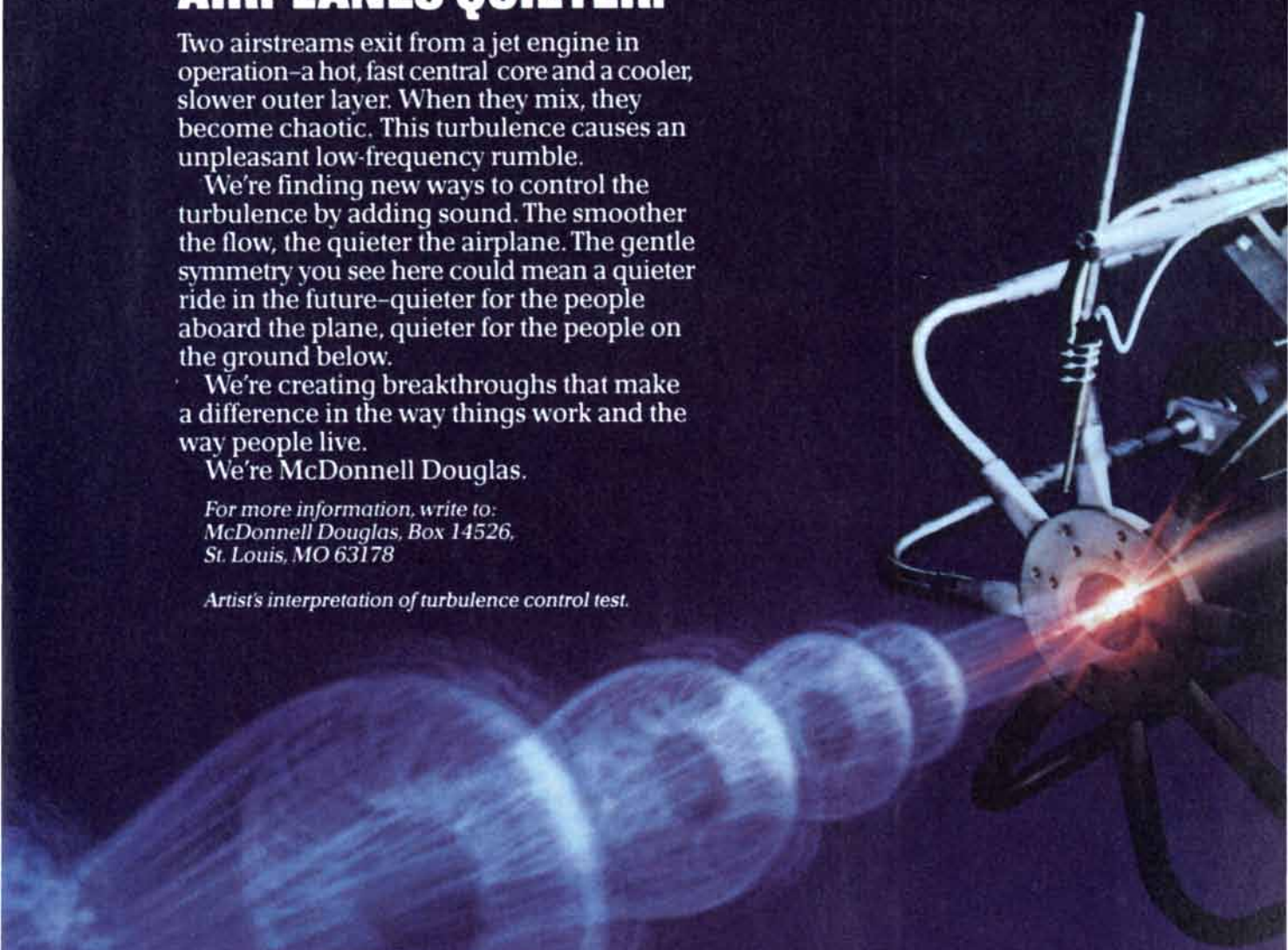
We're finding new ways to control the turbulence by adding sound. The smoother the flow, the quieter the airplane. The gentle symmetry you see here could mean a quieter ride in the future–quieter for the people aboard the plane, quieter for the people on the ground below.

We're creating breakthroughs that make a difference in the way things work and the way people live.

We're McDonnell Douglas.

For more information, write to:
McDonnell Douglas, Box 14526,
St. Louis, MO 63178

Artist's interpretation of turbulence control test.



MCDONNELL DOUGLAS

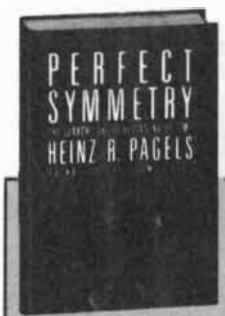
COMMERCIAL & MILITARY AIRCRAFT HEALTH CARE ENERGY SPACE & MISSILES INFORMATION SYSTEMS HELICOPTERS FINANCING

Take any 3 books for only $1 each!

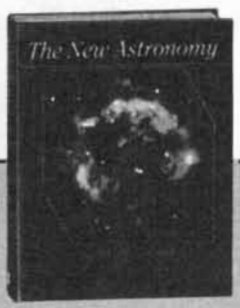
(values to $75.00) *when you join the*

Library of Science

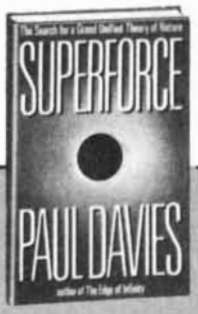
(You simply agree to buy 3 **more** books—at handsome discounts—within the next 12 months.)



67900 $16.95



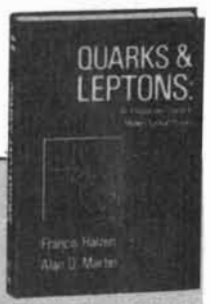
64137 $24.95



81398 $18.95

78211-2 $37.50
(Counts as 2 choices)

85130 $14.95



73185-2 $29.95
(Counts as 2 choices)

Physics, Astronomy and Cosmology

37353 $17.95

36360 $15.95

37205 $19.95

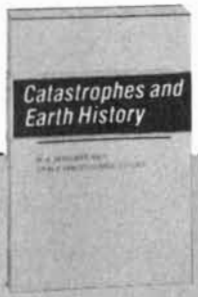
81477 $15.00

79275 $16.95

81445 $15.95

55459 $15.95

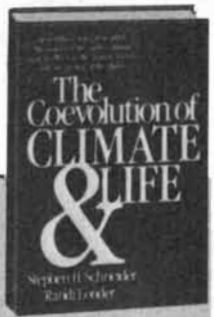
85759 $25.00



37355 $19.50

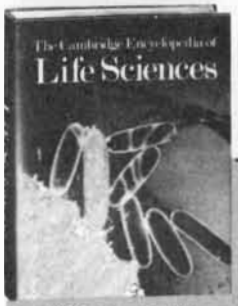
66200 $17.95

50225-2 $34.50
(Counts as 2 choices)

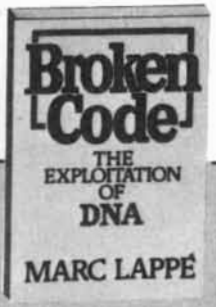


38970 $25.00

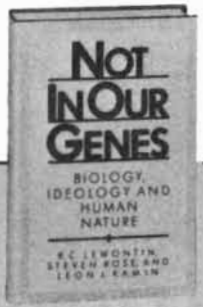
Evolution of Life Sciences



37203-2 $45.00
(Counts as 2 choices)



36890 $17.95



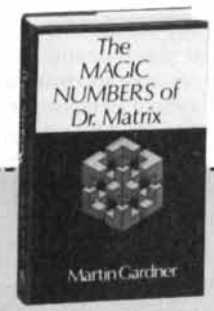
64866 $21.95

85857-2 $32.95
(Counts as 2 choices)

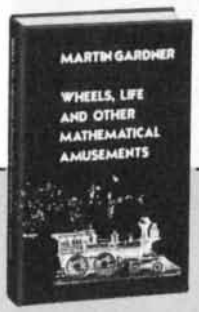
62708 $22.50

76925 $24.95

84395 $14.95

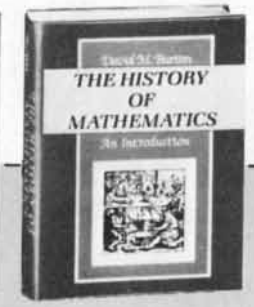


59500 $19.95



84847 $15.95

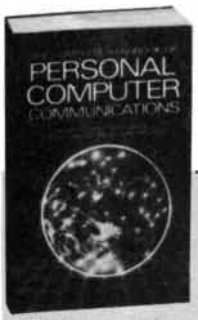
60775 $20.95
(Softbound)



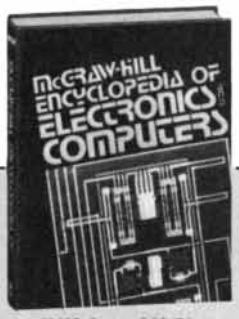
53109-2 $40.00
(Counts as 2 choices)

61510 $19.95

Mathematics

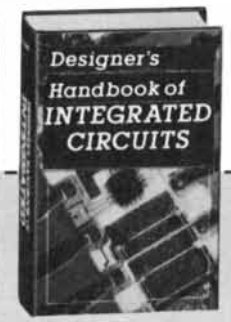


39796 $14.95

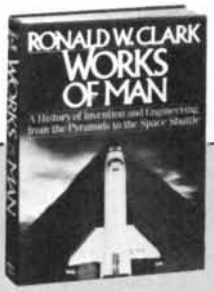


61613-3 $63.50
(Counts as 3 choices)

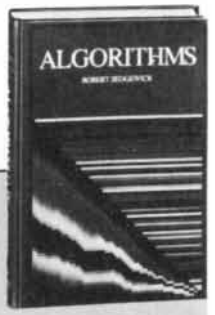
45200 $16.95



42042-3 $69.50
(Counts as 3 choices)

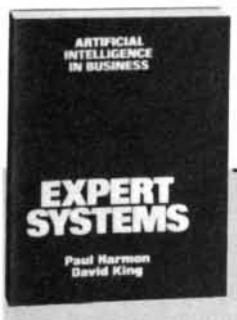


87525 $25.00



32486-2 $28.95
(Counts as 2 choices)

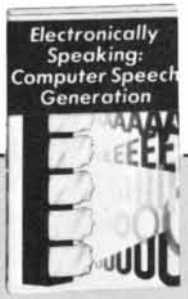
Computer Science and Technology



47100 $24.95

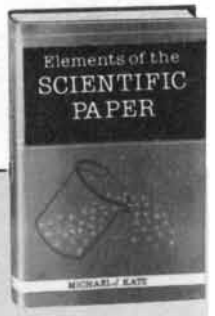
74259 $19.95

42215 $24.95

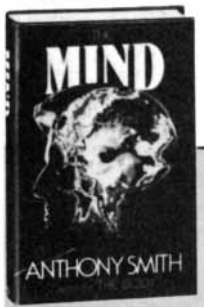


44127 $14.95

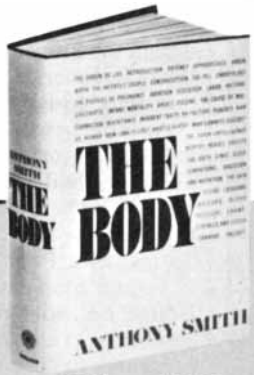
44126 $13.95



44314 $20.00

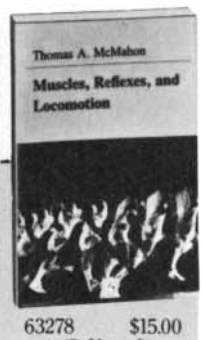


62719 $17.95



36474 $24.95

36789 $16.95



63278 $15.00
(Softbound)

Brain, Mind and Body

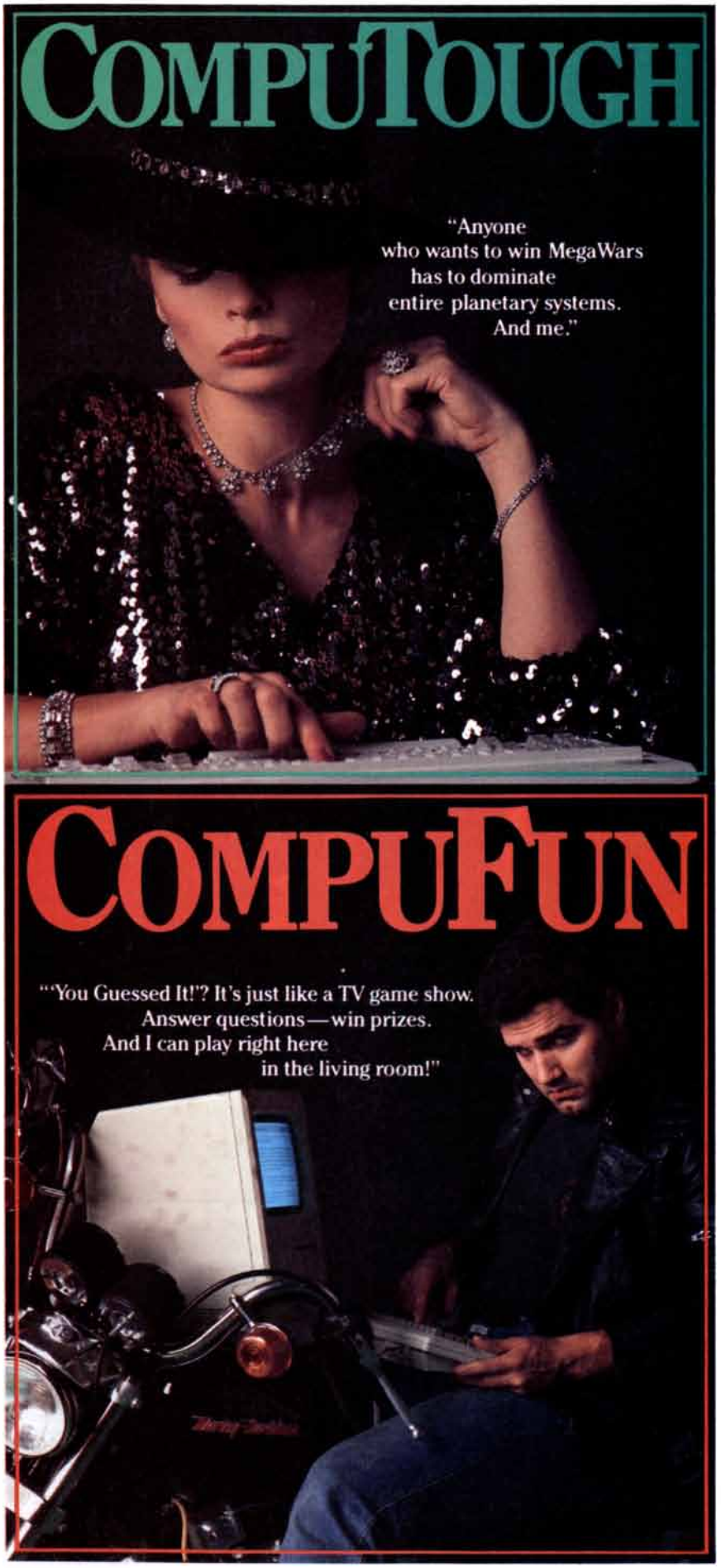
MEMBERSHIP BENEFITS • In addition to getting 3 books for only $1.00 each when you join, you keep saving substantially on the books you buy. • Also, you will immediately become eligible to participate in our Bonus Book Plan, with savings of up to 70% off the publishers' prices. • At 3–4 week intervals (16 times per year), you will receive the Library of Science 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.

The Library of Science is the nation's oldest, most respected science book club.

Dedicated to serving both professionals and serious amateurs, it surveys the entire field of published sources on scientific subjects to make the very finest works available to club members at substantial savings.

If reply card has been removed, please write to: The Library of Science, Dept. 2-BK9, Riverside, N.J. 08075, to obtain membership information and an application.

Scientific American 5/86



hypercube? I thought his answers were insane because they appeared to run counter to ordinary intuition. The diagonal gets longer without limit as the dimension of the hypercube increases. At the same time the volume of the inscribed hypersphere becomes progressively smaller!

The first answer is easy to derive by extending the Euclidean distance formula from two dimensions into higher ones: the length of the diagonal of the n-dimensional hypercube is simply the square root of n, and that number grows without limit as n becomes larger. I do not know of any elementary way to demonstrate the shrinkage of a hypersphere, but Thomas Banchoff of Brown University, whose graphic displays of a rotating hypersphere were featured in last month's column, gives a good plausibility argument.

The volume of the inscribed hyperoctahedron roughly approximates the volume of the hypersphere. But how does one characterize the inscribed hyperoctahedron? Each vertex of an ordinary octahedron inscribed in a cube touches a face of the cube. Similarly, each vertex of the hyperoctahedron lies in the center of each hyperface of the hypercube. Given that hint, the volume of the hyperoctahedron is not difficult to figure out, and it gets smaller as the dimension decreases.

All April fooling aside, I am indebted not to Magi for these intriguing problems but to Ronald L. Graham, the redoubtable mathemagician of AT&T Bell Laboratories.

In the February column I described the chess-playing machine constructed in 1890 by Leonardo Torres y Quevedo, a Spanish engineer. The machine would never fail to checkmate the black king of its human opponent with its rook-king combination. How did Torres y Quevedo construct his machine? An old but excellent description can be found in a book by David Levy, *Chess and Computers*. How might one design the moves of such a machine? I invited readers to submit algorithms that were equally inexorable.

Approaches to the problem, not to mention algorithmic styles, were so different I had to select more than one solution as being most succinct. Three readers, Scott K. Liddell of Silver Spring, Md., Paul Canniff of Marlton, N.J., and Stephen J. Perris of Baton Rouge, La., not only used similar algorithms to get checkmate but also wrote programs to confirm their efficacy.

Liddell's algorithm, the best of the three, applies a rotate-and-divide strategy: "It rotates the board to find any configurations in which the black king is at least two ranks ahead of the white

king. There is always at least one such configuration, but most of the time there are two of them. If the algorithm finds only one such configuration, that configuration is selected. If two configurations are found, the program selects the one in which the black king is closest to the far end of the board. Once a position is selected the program then cuts off the board so that the lone king cannot advance. If the white rook is in a position that blocks it from cutting off the advance of the black king, the rook moves to the far side of the board. No matter what Black does, the program is then able to cut off the board on the next move. The entire procedure requires at most two moves by White."

By the term "entire procedure" Liddell refers to a basic cycle of moves in which Black is forced incrementally toward one corner of the board. A great advantage of the underlying algorithm is its use of rotational symmetry; if the black king is not "ahead of," or above, the white king, the program merely rotates the board until it is.

Several strategies submitted by various readers applied a key concept of chess: opposition. If the two kings are only two squares apart and one of them must move, the moving king cannot move closer to the other king. It must either maintain the same distance or move farther away. Perhaps the best strategy employing continued opposition was the one proposed by Alexis A. Gilliland of Arlington, Va. Here are the rules (in slightly paraphrased form) that he adopts:

1. The rook always occupies the first rank or the eighth rank. When it is attacked on one rank, it is moved to the other one.

2. The white king moves toward the center of the board and attempts to push the black king toward the file of the rook.

3. White seeks a position of opposition in which the two kings are one file apart on the same rank and it is White's turn to move. To arrive at opposition, it may be necessary for White to play a wasting move with its rook, temporarily violating rule 1.

4. When White obtains opposition, it checks the black king with its rook by moving the rook to the white king's file, thereby forcing the black king one file closer to the edge of the board.

5. Steps 3 and 4 are repeated until the black king arrives at the edge of the board. Step 4 then produces mate.

Gilliland maintains his strategy always produces mate. He also notes that an algorithm for mate in the minimum number of moves is not likely to have minimum length itself.

Exceeds the minimum daily adult requirement for driving excitement.

Introducing Grand Am SE. Aerodynamic. Monochromatic. With a 3.0 liter multi-port fuel injected V6 and a rally tuned suspension that knows what to do with a road. Drive one often for fast relief from boredom.

PONTIAC GRAND AM

WE BUILD EXCITEMENT

LET'S GET IT TOGETHER BUCKLE UP.

GM

BOOKS

How the Chinese created paper and printing; curious Cambrian fossils of the Burgess Shale

by Philip Morrison

SCIENCE AND CIVILISATION IN CHINA, Joseph Needham. VOLUME 5, CHEMISTRY AND CHEMISTRY TECHNOLOGY, PART 1: PAPER AND PRINTING, by Tsien Tsuen-Hsuin. Cambridge University Press ($89.50). Two billion of us are literate in one or more of the languages of the earth. Each day we readers enjoy on the average a quantity of new hard copy sufficient to fill an area 10 to 20 times the size of this page. Nearly all of that text and image appears on white paper in ink derived from carbon black. For 1,000 years it was imprinted by a flat surface cut as a mirror image in relief of the desired form.

All three essentials, paper, ink and letterpress, are well-known Chinese inventions. Here is an up-to-date account of their history, the newest addition to the splendid shelf that began to appear under Joseph Needham's signature a generation ago and that continues to please and inform readers, now with a little help from his learned friends throughout the world. Professor Tsien Tsuen-Hsuin is an eminent authority on the matter who has enjoyed a long sinological career at the University of Chicago; his book continues the enterprise in the high comparative style established by Needham. A couple of hundred illustrations, fine footnotes (as a rule less ambitious than Needham's most magisterial asides) and grand bibliographies from East and West buttress the text.

Of the elements of printing, paper came first. Writing is much older. In China it appeared on the scapulas of oxen and the carapaces of tortoises. Stone, clay, wood slips, palm leaves, woven silk and split sheepskin then became widely used mediums. Paper is a matted sheet of fibers taken from water suspension onto a fine screen. It was the ability to unite random linear fibers provided by the disintegration of plant tissues or by textile scraps into a new dimensionality, large flat sheets, that gave paper economic advantage over all its predecessors. Those included natural laminar materials from the stems of the sedge papyrus and from many woody sources of pounded bark. Cellulose-rich fibers are required for good bonding (there are now synthetic exceptions to this rule); no paper is made from silk fiber, although some may be added to give luster. The best natural sources are those without much binding substance that are high in yield of long cellulose fibers.

The oldest specimen of paper we have—it is shown here—was found near Xi'an. Manufactured in the second century B.C., probably from hempen rags, it was coarse paper used for packing bronze mirrors. The official histories of the Eastern Han concur in celebrating the eunuch Cai Lun as the inventor of paper for writing, centuries later, in about A.D. 100. The historians observe that silk was too costly and bamboo slips too heavy. It may be that what Cai hit on was using fresh fibers from tree bark and hemp plant instead of fibers from secondhand textiles. By Cai's time better papers are indeed found in use for writing.

The handicraft grew steadily, in villages near streams and by wooded hillsides (fuel for boiling the raw plant material was important). One by one new fiber sources were tapped. At first hemp and cotton were used, but they were too valuable in textile production; their rags are a limited if superior source. Then came the slow-growing climbing rattan; by A.D. 986 a scholar could lament satirically that worthless writers were killing off the rattan with their millions of words. He was right. The paper mulberry shrub and the young bamboo have been cultivated ever since in most parts of China as the main fiber sources for handcrafted paper. Of course, pulped wood fibers are now the dominant source for the industrial process everywhere.

The oldest paper known was used for careful wrapping. Today too only about four-tenths of the paper made worldwide is intended to receive ink. Old China also used handmade paper variously; paper currency, "flying money," began to supplement coins about 1,000 years ago, and everyone knows of China's ubiquitous paper lanterns, kites, fans, umbrellas, windows and screens. Funeral offerings simulated worldly wealth in paper. Paper effigies of caparisoned horses, camels, servants, armor and money "in great quantities" were, according to Marco Polo, all burned along with the dead man. Less familiar are light paper armor, practical for foot soldiers in difficult terrain, and decorated wallpaper. The latter touch, which caught on for good in Europe earlier than 1700, was regarded in China at the time as being in rather low taste. Nevertheless, one photograph here shows a delightfully intricate round geometric design, the oldest paper-cut known, found recently in Xinjiang. It was made in the fifth or sixth century. A few pages later we see the proverbial paper tiger itself, satisfyingly striped and toothy but hardly frightening.

The humblest and most intimate use is also old; literary sources are primly silent, but in the sixth century a scholar, Yan Zhitui, wrote for his household: "I dare not use for toilet purposes" paper bearing the names of sages. An Arab traveler of the ninth century, accustomed to meticulous ablutions, was offended by what he regarded as the careless Chinese habit. Toilet paper was made cheaply in very large quantities from rice straw; in 1393 the imperial court alone bought 750,000 newspaper-size sheets, reserving a special quality for the imperial family. At the other end of the aesthetic axis, we encounter the small, reddish notepaper designed by the ninth-century courtesan-poet Xue Tao. Made of hibiscus skin mixed with the powdered petals of hibiscus, it remained in vogue for many centuries.

Paper led to printing. Hand copying of manuscript onto silk and paper was well known. Seal cutting on wood blocks of inscriptions in mirror image, sometimes consisting of as many as 100 characters, was certainly a forerunner; so was the wide use of inked paper squeezes or rubbings made from ample carved inscriptions on bronze and stone. The astonishing Mountain of Stone grotto in Hebei has preserved for a dozen centuries 7,000 steles, each 30 meters in height, bearing in all four million words of the Buddhist canon carved in relief. The complex ideographs meant accurate copying by hand was costly and difficult. Yet both the standardization of Confucian texts needed for national civil service examinations and the zeal of believers enjoined to multiply repetition of the powerful formulas of Buddhist sutras provided motive enough for the development of printing.

We hold sure evidence of the consequences. In 1966 a miniature scroll two inches wide printed from a dozen 20-inch-long wood blocks was found in southeast Korea. It is the oldest printed document known. The text is a Buddhist sutra done into Chinese. A known translation made by a monk of Chang-An (now Xi'an), the temple construction itself and the form of certain new characters place the date of production early in the eighth century. Only slightly later is the famous A.D. 764 edition of a million copies of brief prayers from the same sutra. They were placed by order of the empress Shōtoku in 10 Japanese temples; each printed charm was stored in a tiny wood pagoda. Many copies and some of the pagodas still exist; one is shown in an illustration. Although the edition of one million was certainly printed, it cannot be established from what material the blocks were cut. These examples produced outside China follow exactly the detailed methods of all subsequent Chinese printing. It can hardly be doubted that printing had begun somewhat earlier within China, had followed the widening usage of Chinese characters and had finally spread beyond that symbolic frontier.

The art flourished. Sir Aurel Stein found in the monastery at Dunhuang as long ago as 1907 what is still the oldest complete printed book we know, a five-meter paper scroll of the *Diamond Sutra*. Its colophon dates it "reverently . . . for universal distribution" on the 15th day of the fourth month of A.D. 868. Its frontispiece showing the Buddha and his disciple and its rather refined printing are eloquent. Although the Bible of Johann Gutenberg would not appear for six centuries, there are quite a few printed Buddhist texts known to be only a little younger than this sutra.

The craftsmen and the craft of Chinese bookmaking are described here in some detail: the calligrapher whose nimble brush put the information on fine paper, pasted ink side down to the smoothed fruitwood block; the skillful engravers who then cut away with their elegant arsenal of chisels and gouges all surface that was not black; the printer who inked the block with a horsehair brush, pressed the paper to the inked surface and brushed a pad lightly over the back. By the Ming time, Matteo Ricci reckoned that in a day's work a skilled printer produced 1,500 double-page sheets. The engraving process was proofread at four stages; inlaid local corrections to the block are commonplace. Page format and binding have long since been standardized, and they support a fine trade jargon, adorned by such terms as elephant trunk and book ear, applied to the functional features of the traditional paperbound, thin Chinese book.

The 81,258 magnolia wood blocks carved on both sides to print the *Tripitaka Koreana,* which was completed in the year 1251, are still neatly shelved in southern Korea. They capture the flavor of the golden age of Chinese bookmaking during that period, the Song dynasty. Such grand canons were rare, but there were a great many widely used secular works and many private printers as fine as those of the state and the monasteries. By the year 1500 some 20 million books had been printed in Europe, copies of perhaps 10,000 or 15,000 titles. From that period China yields registers of 50,000 titles, half a million distinct, thin volumes. It seems likely that until about 1700 it was text in Chinese that could be read from the majority of the world's printed pages.

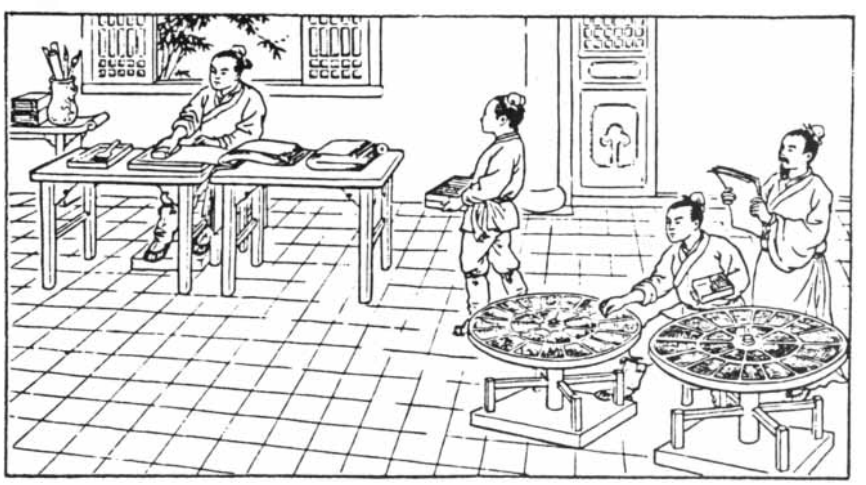
Of course, we associate movable type, not whole-page blocks, with Gutenberg. Such type suits well the alphabetical languages of Europe. Yet Gutenberg holds no priority even on movable type. Earthenware type for individual characters was certainly made and used in China in about 1050. The contemporaneous scholar-scientist Shen Kuo tells us about it in terse detail. "If one were to print only two or three copies, this method would be neither simple nor easy. But for printing hundreds or thousands of copies, it was marvelously quick." A font in a Western language needs about 100 species of type; given a few thousand individual pieces, any spread can be set. A full Chinese font requires from 2,000 to 400,000 pieces, including the necessary repetitions. The hardworking early compositor could put most characters within arm's reach only by the use of a number of big lazy Susans. Eventually he was kept on the run. In about 1300 the characters of a Chinese font were arranged by their rhyme.

Type was made from wood, from enamel and from bronze in China before the European invention of printing. Korean prints made use of cast bronze type as early as 1234. The impetus came from the book by Shen Kuo; therefore it seems possible such type may have been used earlier. There are extant pages printed from cast bronze type in Korea in 1403, 50 years before the German success. A book or two were printed there with movable wood type even earlier. In Korea the casting of type derived from the casting of coins. The character was cut in beechwood; the die was pressed into a clay mold, which was baked and used to cast the bronze. The type was finished by polishing. In premodern China lead was not used for type.

Historical queries crowd into the mind, and Tsien has addressed them soberly. Paper clearly entered Europe from China by way of Samarkand, Baghdad and Damascus by means of Arab intermediaries. The oldest European paper manuscript known comes from 10th-century Spain. By 1150 Arab and Jewish papermakers there were manufacturing an excellent product from flax.

Printing took three or four centuries more to appear, although for a century before any type was set by Gutenberg many products were printed from wood blocks: textiles, playing cards, religious images and even block books. All of these have earlier Chinese counterparts; yet there is no clear evidence one way or another for the diffusion even of the notion of movable type for text from the world of 10,000 characters to the world of the short Latin alphabet along the river Main.

Even more striking is the contrasting



Chinese printshop circa 1300, with type cases at the right

Quality. Our Magnificent Obsession.

This is as close as we can come to showing you Lands' End Quality in black and white. In our free catalog—and please send for one—living color does it more justice.

Finally, though, you'll need to feel the fabric itself in this Pinpoint Oxford shirt to understand the outer limits of Quality we insist on in Lands' End products, and why we're capitalizing the word in this advertisement.

In this shirt, Lands' End Quality is revealed in the material itself, in the construction of the shirt, and in the generous proportions of the finished garment.

The material is woven in Japan of the exceptionally fine cotton yarn it is possible to spin there. As to the make of the shirt, it is characterized by the fact that all seams are fully single-needle stitched, the collar and cuffs are double-track stitched as they should be, and the collar is non-fused, the only way to give it the natural roll that makes buttondowns what they are.

Finally, we grace the shirt with a box pleat and locker loop, extra long tails, a 7-button placket, gauntlet buttons, even an English-style split back yoke. State-of-the-art tailoring.

All this Quality with a capital Q and our Lands' End Pinpoint is yours for just $29.50, whereas $45 might not buy a comparable shirt in your favorite men's shop.

We're obsessed with Quality at Lands' End. It's a matter of first principle:

Principle 1:

We do everything we can to make our products better. We improve material, and add back features and construction details that others have taken out over the years. We never reduce the quality of a product to make it cheaper.

Clip the coupon, won't you? Better still, give us a call (800-356-4444) any hour of the next 24. Ask for a free catalog and get to know Quality in our definition of the word in everything we make or offer.
All of it GUARANTEED. PERIOD.

Please send free catalog.
Lands' End Dept. Q-30
Dodgeville, WI 53595

Name _____

Address _____

City _____

State _____ Zip _____

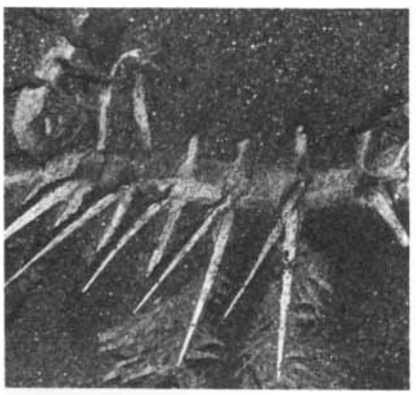
Or call Toll-free:
800-356-4444

impact of printing on the two worlds. Printing on paper is democratic, we traditionally hold; much cheaper than hand copying on scarcer surfaces, it reaches many more readers. It fed a great expansion of learning in Europe, characterized by national diversity and social mobility. At the same time, printing guarantees authentic spread of a central text (and figures), and easy multiplication enables many to ponder the same canonical truths. In Europe print capped deep change within a century or two; in China it shared for 1,000 years the punctuated stability of a unified culture.

This fascinating book is a worthy updating of its celebrated predecessor of 60 years ago, *The Invention of Printing in China and Its Spread Westward,* by T. F. Carter, revised in 1955 by L. Carrington Goodrich. Let us offer the courtesy of the last words to an Elizabethan versifier, old Thomas Churchyard: "Though parchment duer a greater time and space, / Yet can it not put paper out of place: / For paper, still, [from] man to man doth go, / When parchment comes in few men's hands you knowe."

THE BURGESS SHALE, by Harry B. Whittington. Yale University Press, in association with the Geological Survey of Canada ($25). The unexpected bones of our own ancestors lie quietly in Olduvai Gorge; the first of all the birds left unique impressions of its plumage in the satiny limestone of Solnhafen. This persuasive and visual little book by an authoritative University of Cambridge paleontologist takes the reader to another fossil site richer, if less intimate, than Olduvai, to see the marks of creatures stranger, if humbler, than *Archeopteryx.*

A few miles west of the tourist-filled meadows at Lake Louise in the Canadian Rockies, a certain Mount Burgess looks across to a steep ridge. Halfway up that ridge—the scene is pictured in the book—an area of fine-grained dark shale is exposed. It was named the Burgess Shale in about 1909 by Charles D. Walcott, newly the secretary of the Smithsonian Institution. The Burgess yields no trace of bone or feather; those are biochemical inventions that came much too late to have entered the 540 million years of storage here. But it is a wide window on Cambrian life in the sea—the dry land was yet unquickened—that we find nowhere else.



Hallucigenia sparsa, *a Burgess Shale fossil*

Walcott was not only a scientific administrator of talent but also a brilliant and untiring field-worker. He was a world authority on the Cambrian System, whose rocks he searched out and split during more than 50 years. His greatest discovery, the Burgess Shale repository, was made with his wife. They paused to split just one more loose block of shale, and before them shone silvery images against the dark rock; they dubbed the intricate filmy creatures "lace crabs."

Walcott returned the next summer with his two sons. The three patiently searched the slope high above the place where the loose block had been found, until they came on the narrow band rich in fossils. Summer after summer they came back to pry at that rock. During eight years of work—Walcott was 67 the last year he spent at the site—their excavation removed a six-foot layer from an area the size of a narrow garden. From that shale came 60,000 fossils, including 15,000 thumb-size lace crabs, formally *Marrella splendens.* They are much more abundant in this fauna than any of the trilobites, the forms that typically characterize the Cambrian.

Harry B. Whittington's compact book is a model of concise exposition and presentation of the evidence. The reader will see 100 photographs, drawings and models of the animals of the Burgess, all carefully marked for scale, many exhaustively described. The gallery makes a curious impression on a reader who is no marine biologist. Unfamiliar enough, these varied soft-bodied organisms, spiny, tentacled or segmented, would attract anyone spotting them in a rock pool, but they would come to a casual stroller as no grand surprise.

Yet the expert paleontologists, from Walcott to today, betray surprise. The very names they gave stand witness: *Hallucigenia, Anomalocaris, ... inexpectans, ... dubia, ... rara....* The reason is plain: inconsistent architecture. Take *Hallucigenia.* You and I look at its inch-long print with interest. The little animal stands on seven pairs of pointed spines, and its back bears a set of tentacles. Its head is only vaguely known. It looks like an unusual but plausible enough small scavenger of the shallow sea floor, but to the author "it would be difficult to imagine an animal as bizarre." Walcott thought it a strange kind of worm; the moderns, who know 40 prints of it, some from the 1930's, see that it is a puzzle, without clear affinities, without any basis for higher classification. Seven spines? There are a couple of dozen such forms, most possessing a variety of traits that are plausible in themselves but are so strange in combination that the overall assemblage cannot be placed in any large group now known, whether continuing or extinct.

These are creatures from another world of time, the seas of half a billion years ago. Among their kind they are neither primitive nor simple. They played subtle and well-adapted roles in a highly organized ecology; some of those roles can be made out clearly today by their diverse means of feeding, for example. Their strangeness consists only in that they neither left descendants nor followed the architecture of more fortunate forms that did. The reasons for their extinction could be manifold, even chancy. We have every reason to regard these creatures as a fair sample of their time and place, reminders that evolution is an ongoing experiment on a large scale, its many repeated runs far more numerous than our capricious samples.

How did this unique paleontologists' paradise come about? That account is the most fascinating part of the narrative. There is no space to follow the arguments in detail; moreover, the geological diagrams here are indispensable for clarity. The conclusions derive from the work of a wonderful little international expedition to the Burgess, sent there in 1966 by the Geological Survey of Canada, Whittington among them.

What they untangled was context. Walcott's fauna are so strange because nowhere else were so many soft-bodied Cambrian creatures preserved. Within the past few years, though, a number of other sites have turned up, some only a few miles from the original quarry, some as far distant as Utah, that yielded a few specimens resembling the Burgess Shale creatures. What happened undersea on Mount Burgess long ago was that a small section of a sloping mud fringe around a Cambrian reef (a limestone structure, too early to have been the work of the polyps that build modern coral) suddenly slumped. The turbulent mud carried down with it the creatures of a shallow, oxygenated, well-lighted patch of sea bottom below the high

reef wall, burying them no great distance away in a quieter, muddier bottom where seawater circulation was poor and the oxygen sparse.

All these circumstances can be pretty well inferred from the details of the rocky context, examined for miles around, and from the key fossils themselves. They are preserved as mineralized film (best photographed in ultraviolet) lacking original hard parts, compacted and flattened tenfold during their long mineralization. The microdrill the moderns use to dissect the fossils makes it plain that the animal had been tumbled and infiltrated with fine mud before its compaction; mineral grains still separate the delicate layers of the fossil itself. So constrained an event is evidently rare. Without some version of it we would find no well-preserved soft-bodied animals, even though they were dominant in that time and place.

The largest animal of the Cambrian swam there too, a formidable if soft-bodied hunter, particularly of the still-unhardened large trilobite species of the time, all forms known only from the Burgess Shale. The monster was about 20 inches long. It undulated its way close to the bottom. Up front it bore a hardened but not mineralized ring jaw and a single pair of strong spined limbs it could dangle to probe for and then seize its crawling prey. Isolated parts of the predator were known earlier, but the whole creature has been assembled, understood and restored only very recently by the author and a colleague, Derek E. G. Briggs. Any daring scuba diver on Cambrian holiday from another star would have been delighted to spear such a prime trophy, unclassifiable in any phylum we know.

DESCRIPTIVE CHEMISTRY, by Donald A. MacQuarrie and Peter A. Rock. W. H. Freeman and Company (paperbound, $12.95). A platinum crucible three times as large as life glows sunset orange in its porcelain triangle above the red-hot mesh and blue flame of the big Meker burner. The colors of this dramatic photograph are vivid and convincing; anyone whose memories of beginning chemistry lab hold that sight will all but feel the heat. The thin volume continues in a visually striking mode, although the many text figures in excellent color do not flaunt the panache of the glossy cover.

These two chemists of the Davis campus of the University of California have written a thick new textbook on general chemistry, overflowing with the powerful valence models, molecular structures and simple thermodynamic accounting held fit for today's beginning student in chemistry. What is here is not their full text; it is a reprinting of a dozen selected chapters. The book still bears the textbook's tags, chapter-end questions and all, but it has a simpler goal and a much wider appeal. The authors set out a "systematic discussion of the chemistry of the elements and their compounds," period by period. This is not the fastidious grammar of introductory chemistry but rather its diverse vocabulary, here given fresh and novel photographic support. (Many of the photographs were made specially for the book by Chip Clark.)

About 40 elements appear onstage in purest form. We see a blue-violet beakerful of liquid oxygen, and later three graceful long-necked flasks displaying green-yellow chlorine, red bromine vapor and violet iodine. (Unstable astatine is too rare and fluorine too irascible to pose in glassware.)

Metals abound, of course. The listing here reckons 83 of them among the 108 elements entered in the periodic table. One can see golden cesium and white phosphorus safeguarded against the corrosive atmosphere, gleaming bricks of noble gold itself, the pure wafers of zone-refined silicon, potassium reacting explosively in water, and sodium, a fresh knife-cut section shiny within its rind of tarnish. Gallium lies liquid in the hot palm. Here too are the flame colors and discharge tubes one expects, from the red strontium flare (rather overexposed) to the argon glow tube. Each element is grouped with its electron-shell analogues in a compact textual account of main reactions and behavior, along with simple tables of the compounds and their uses, usually also with sources and methods of extraction of the element.

Celebrated minerals are shown, cinnabar and beryl, orpiment and hematite, spodumene and crystalline sulfur, even routine gypsum and quartz. An ammonia plant is viewed by night, and anhydrous ammonia is seen jetting into the soil straight from the plowshare, its most important application today. We see the natural alternative, nitrogen-fixing bacterial nodules on the roots of a legume. A mouse, submerged in clear fluorocarbon, is presented breathing dissolved oxygen; it looks rather uncomfortable, but it is safe enough. The microcrystals of xenon tetrafluoride are a curiosity, unfortunately shown without a scale marking. The transition ions of vanadium, chromium and manganese are exhibited in varied solutions for color comparison of the differing states.

The text is fundamental; it touches matters such as the three isotopes of hydrogen, the chemical-engineering processes of Frasch, Solvay, Bayer, Hall and Raschig, and basic oxygen steelmaking. (A note reminds the reader that soda ash in the U.S. is now the legacy of an old volcano, only a tenth of it still the product of the brothers Solvay.) In sum, the book outlines and evokes the indispensable chemical experiences that arise best in museum, field trip and beginning lab. The laboratory is brought a little nearer by a few facsimiles of current cautionary labels for bottles of acid, just what a novice ought anxiously to study before opening the reagent stopper.

Many a general reader will want the book as reminder and reliable first reference; eager would-be chemists en-

Sodium, a soft metal, being cut with a knife

chanted by substance and change, and caught by the specificity of smells, colors, sparks and pops, will find it a useful supplement to the experience itself. Faute de mieux, to be sure, but somehow proper in today's world of cheap yet rich visual representations.

It would be a treat to see a companion work on organic chemistry, a topic not included at all. Carbon appears here in its mineral avatar alone, as diamond, carbides and the oxides. Still we gain the unfamiliar sight of the red crystals of carbon tetraiodide. A similar simple book with pictures of thready masses of DNA, the bright colors of dyestuffs, the patterns of layer chromatography and the less tangled flow sheets of the petrochemical industry would make an admirable sequel, perhaps with a page or two of aromatic samples to be scratched at will. The top 50 chemicals produced in 1983 in the U.S. are tabulated here; of these a score are true organics.

The few omissions and missing citations in the reliable text are small flaws arising from the spin-off process that generated this pleasant and inexpensive abridgment. Most missed is an account of the lanthanides and the actinides; they appear here only in the periodic table.

EDWARD S. CURTIS: THE LIFE AND TIMES OF A SHADOW CATCHER, by Barbara A. Davis. Chronicle Books ($45). Edward Sheriff Curtis was a gifted and energetic young photographer and engraver who lived in Seattle at the turn of the century. Out of that frontier town and busy port the Yukon and the Nome gold rushes gave birth to the brief Harriman Expedition, whose brilliant scientific staff reconnoitered western Alaska for a site that would accommodate a trans-Bering railroad connection by bridge or tunnel. Curtis went along as a photographer; he returned with luminous and dreamlike landscapes of the Arctic world. Soon the life of a city studio no longer fulfilled him.

In a well-documented 100-page biography Davis follows the shadow catcher from youth to his death in Los Angeles in 1952, at 84. From about 1904 to 1930 Curtis carried out against all odds his grandiose scheme to publish 20 volumes of text and image on *The North American Indian*. About 300 copies of the original work with its expensive and enormous photogravures exist, and the text has been reprinted. Most of the biography treats his struggle to complete his work and then to save it from remaindered obscurity. The story is a pageant of the American early 20th century; its characters range from TR and J. P. Morgan to

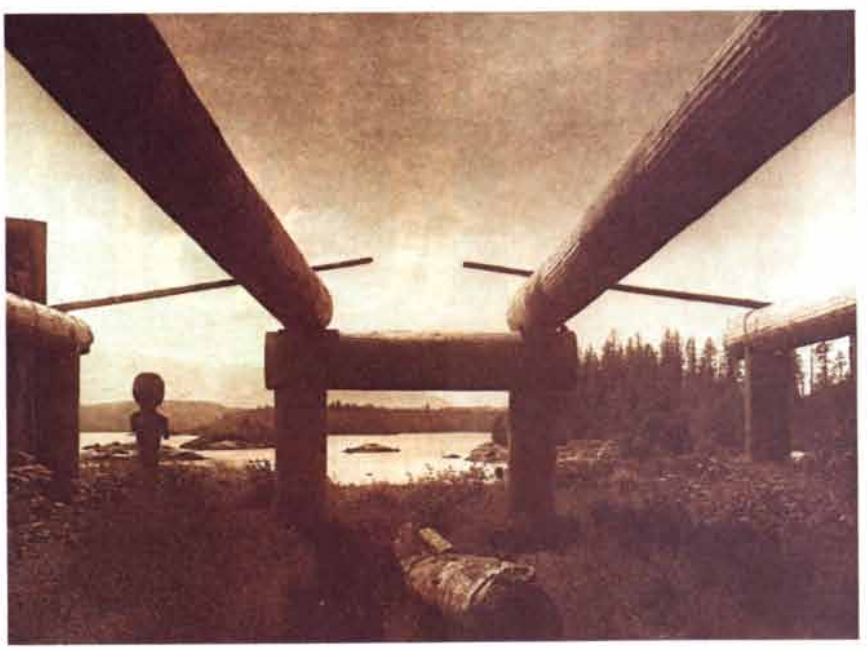
Geronimo, Buffalo Bill, Franz Boas, Robert Flaherty and Cecil B. de Mille. Curtis knew them all, the powerful as erratic patrons, the others as fellow artists, scientists and strivers. He was a cameraman for de Mille on *The Ten Commandments* and an adviser to Flaherty for *Nanook of the North*. He entered the lives of many subjects easily and sincerely; his principles of direction were built around the pronoun *we*, not *I* and *you*.

The volume offers about 150 plates, some of them reproduced from prints made anew from the original copper gravures, some not published before, many of them once celebrated. It is these shadows that establish Curtis as a kind of second John James Audubon, two men seized with a dream that lay between science and art. Audubon worked with brushes, a shotgun and his own unflagging energy; Curtis needed dozens of expert helpers in the field and at home. Audubon pictured the birds of forest and meadow; Curtis, the men and women of 75 cultures over an entire continent during an era of defeat and travail. Both achieved the wide reproduction of outsize images. Audubon painted for the lithographic stone; Curtis, a century later, took his big glass plates for printing by photogravure.

Curtis could not fully succeed. Yet he did not fail. These enhanced shadows are not caught just as they fell. They were seen by an ethnographic eye that sought romance in every image but only rarely allowed the excess called sentimentality. These people and the works of their hands look more beautiful, prouder and more powerful here than in any other sustained evocation.

Of course that is not the whole truth; they were then in real peril and poverty, and there they mostly remain. Yet it is part of the truth all the same. Here is the Canyon de Chelly, riders of the People before the towering pinnacles. Here are the kayaks of Noatak hunters, like seaborne birds in the dusk. Here is a timbered tomb of a chief of the Haida at Yan, a monument looking as ancient and strong as a tumulus on the plains of Europe. The weaknesses of this artist are all too easy to see, as Beaumont Newhall remarks, but his strengths lie deeper.

The collection (it includes a wide variety of Curtis' work beyond the Indian studies) and the text are fine to see and to ponder. It is worth recalling that Curtis completed one motion-picture film, made in the Northwest among the Nootka and the Kwakiutl between 1912 and 1914. It was shown briefly but soon disappeared from view. One print was rescued and released not many years ago by the University of Washington; the story is told in a 1980 book from that press by Bill Holm and George Quimby. The events of the film are embedded in a staged and costumed drama no subtler than its period, but while masked Eagle and huge Bear dance grandly in the prows of two speeding war canoes the viewer for a moment shares in full the hunter's beliefs. There too we feel the power of this man's art, shadows caught in an unflinching pursuit of a vivid and loving dream.



Kwakiutl house frame, photographed by Edward Curtis in 1914

THE GREAT SPACE RACE

Beginning May 7 on PBS

With Jo Franklin-Trout

"The Great Space Race" is an unprecedented four-part series that will examine man's quest to conquer space. Never before has a television series offered such a close-up view of the people, the stakes and the goals of international space activities. Tune in on Wednesday, May 7, and join us as we explore space. Funding for "The Great Space Race" has been provided by The Boeing Company. **BOEING**

Established 1845 SCIENTIFIC AMERICAN May 1986 Volume 254 Number 5

Sudden Cardiac Death

The heart can stop pumping virtually without warning. Immediate aid, provided either by emergency medical service personnel or by trained bystanders, can mean the difference between life and death

by Mickey S. Eisenberg, Lawrence Bergner, Alfred P. Hallstrom and Richard O. Cummins

There is no more dramatic personal medical emergency than sudden cardiac death. An individual's heart all at once ceases to function. The victim, often in apparently good health, collapses either entirely without warning or with at best fleeting premonitory symptoms such as chest pain, shortness of breath, sweating, nausea and overwhelming fatigue. Because no blood flows to the brain once the heart stops pumping, the victim loses consciousness in a matter of seconds. Seizures, gasping for breath and incontinence follow as death rapidly approaches.

In the past there was little hope for those struck by sudden cardiac death. Today circulation and respiration can be artificially sustained so that the victim may then be resuscitated by literally shocking the malfunctioning heart back into its normal rhythmic contractions. These procedures are futile, however, unless they are applied within minutes after the victim collapses. Each minute that passes without proper treatment precipitously reduces the chances of resuscitation.

Because of the need to have medical care reach the victim as soon as possible, services that quickly dispatch trained specialists to the scene of a reported cardiac arrest have been established throughout the U.S. as well as in many other countries. Emergency medical personnel have been able to react speedily enough to snatch life back, allowing sudden cardiac death victims to return to normal, socially productive lives.

Without question, wider adoption of such programs would save more lives. Evaluations of emergency medical service programs have shown that staffing specially equipped vehicles with trained personnel can be an effective and efficient way to deal with the problem of sudden cardiac death. Complementing major investments in personnel and equipment are less expensive, practicable steps that could also markedly decrease the mortality of sudden cardiac death victims.

The term "sudden cardiac death" refers to the unexpected cessation of breathing and circulation caused by an underlying heart disease such as atherosclerosis of the coronary arteries. Unless breathing and circulation are restored within minutes after it strikes, permanent biological death, ushered in by irreversible brain damage, cannot be avoided.

Sudden cardiac death is a major subset of the broader problem of coronary heart disease. Exactly how many people suffer from coronary heart disease is not known. Autopsy studies indicate that early signs of the disease are present in virtually every American male by the third decade of life. Unfortunately the mere presence of factors recognized as contributing to the risk of heart disease does not accurately identify persons liable to sudden cardiac death. Indeed, most cases of coronary heart disease are generally not discovered until serious complications have arisen from it, such as angina pectoris (pain produced by a lack of oxygenated blood circulating to the heart muscle), myocardial infarction (the death of heart-muscle tissue) or, in the worst case, sudden cardiac death.

These and other heart-disease complications manifest themselves every year among Americans as approximately 1.5 million heart attacks. Some three-fourths of those who suffer a heart attack are admitted to a hospital, where they can be given advanced coronary care. These patients have excellent survival rates: more than 80 percent of those admitted survive the attack and are discharged.

More than 350,000 Americans per year are not as fortunate. Their heart attacks often come entirely without warning; when symptoms do manifest themselves, they precede the attack usually by less than an hour, giving the victim little opportunity to reach a hospital before collapse. These are the victims of sudden cardiac death. The majority of them have a known history of coronary heart disease, which may include previous myocardial infarctions and angina. In about 20 percent of the cases, however, sudden cardiac death is the first indication to the victim that his heart and coronary arteries are in fact unhealthy.

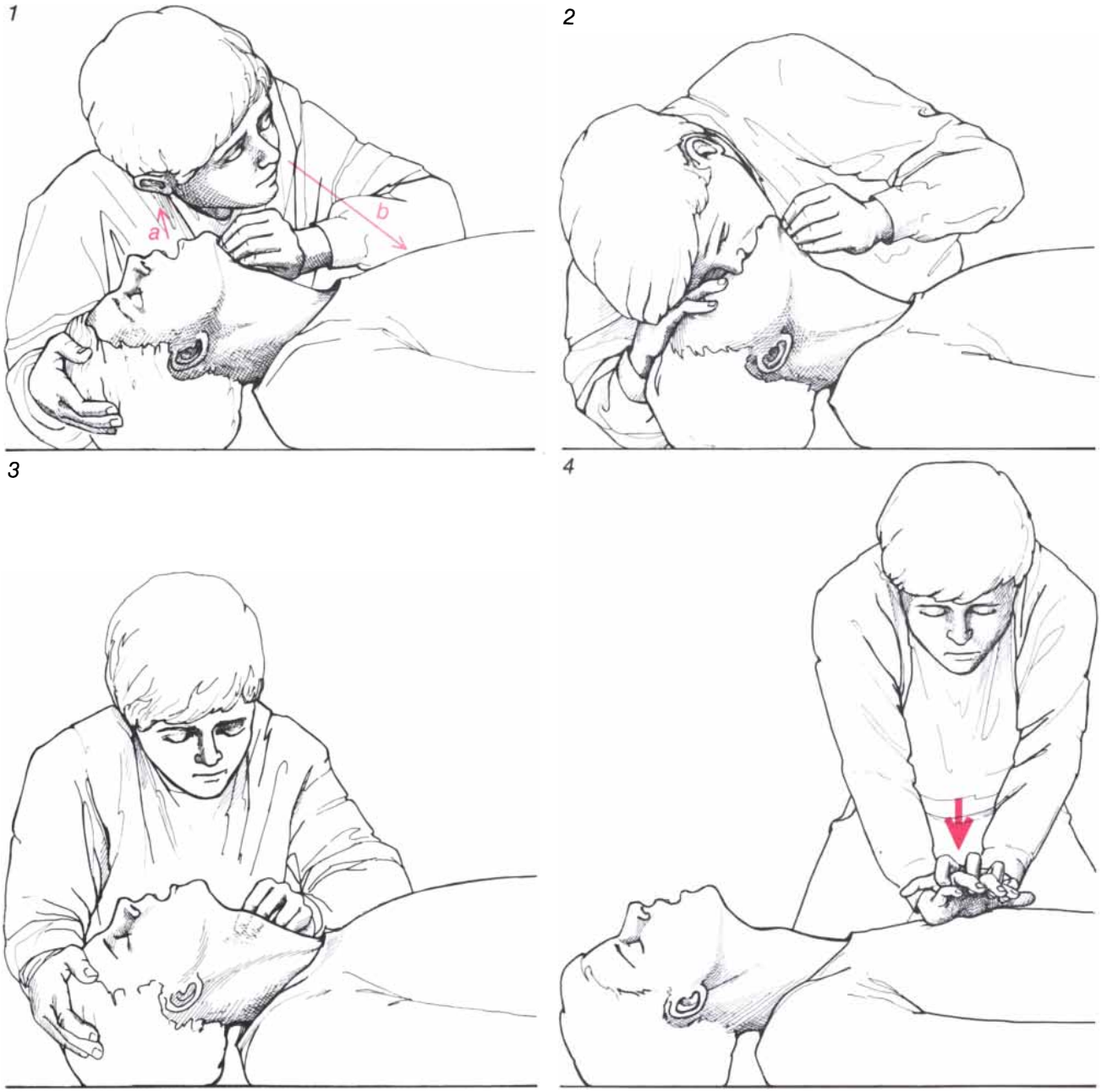
The chambers of a healthy heart normally contract and relax rhythmically, their action being coordinated by electrical impulses emanating from two nodes in the heart and transmitted to the muscle along embedded nerve fibers. Although the heart can experience many kinds of arrhythmia, or abnormalities of cardiac rhythm, for a variety of reasons, only a few arrhythmias are associated with sudden cardiac death. The commonest of these is ventricular fibrillation, which is observed in about two-thirds of the cases.

During ventricular fibrillation the

impulses that control the contraction of the ventricles, the two large, lower chambers of the heart, become chaotic. The muscle twitches but there is no coordinated pumping action, so that blood is not expelled from the chambers. As a result the blood pressure and pulse rate immediately fall to zero.

The mechanism that triggers ventricular fibrillation is unclear. Suggested causes include myocardial ischemia (reduced blood flow to the heart muscle), electrolyte imbalances or platelet abnormalities in the blood and psychological stress. Acute myocardial infarction is a precipitating factor in only 20 to 30 percent of the cases of ventricular fibrillation.

The absence of consistent findings among patients resuscitated from sudden cardiac death suggests there may



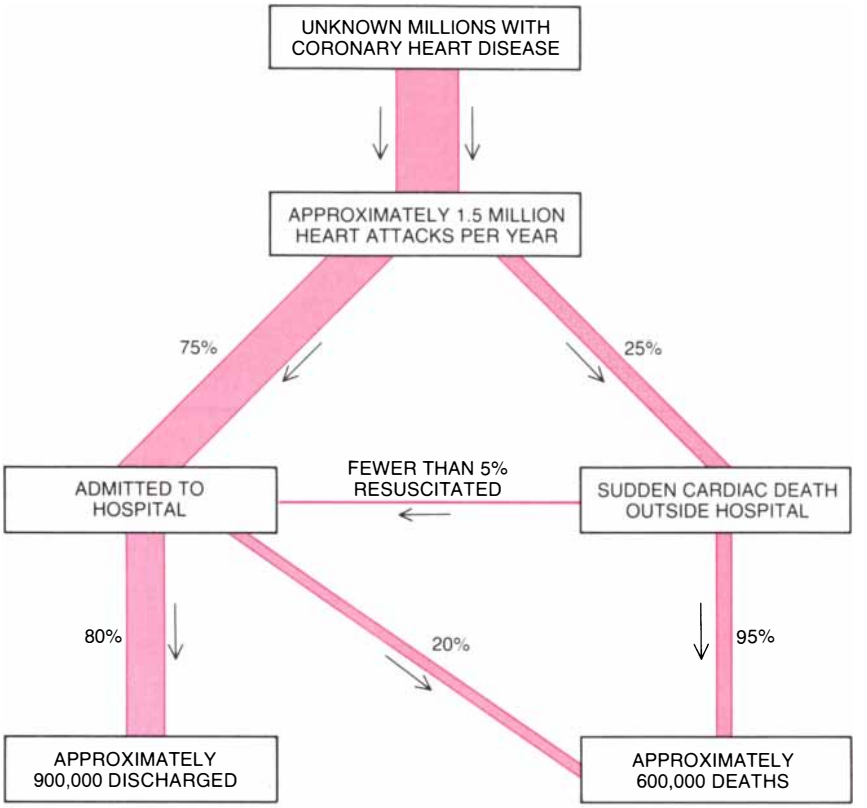
CARDIOPULMONARY RESUSCITATION (CPR) is the appropriate immediate treatment for a cardiac-arrest victim. When it is determined that the victim has indeed lost consciousness, the head and neck are positioned to ensure that the airway is not being blocked by the back of the tongue (*1*). This is done by tilting the head back and lifting the chin. Absence of breathing is confirmed by listening or feeling for exhalation (*a*) and looking to see if the victim's chest is rising and falling (*b*). If there is no breathing, artificial mouth-to-mouth respiration (*2*) is started immediately. The victim's nostrils are pinched shut, the rescuer seals his mouth over the victim's and blows into the victim's lungs. After two initial breaths, given in the space of four to seven seconds, the pulse is checked by feeling for the pulsation of the carotid arteries in the neck, on each side of the Adam's apple (*3*). If no pulse is detected, chest compressions (*4*) are initiated at once. The heel of one hand is placed on the victim's sternum, two inches from the sternal notch. The fingers of both hands are interlocked, and pressure is applied in a rocking motion by thrusting firmly downward with elbows locked. The chest of the victim is depressed just under two inches deep 80 to 100 times per minute. If the rescuer is working alone, he must alternate between chest compressions and ventilation: for every 15 compressions, two inflations are delivered. Victims of cardiac arrest who receive CPR immediately after collapse have roughly twice as much chance of survival as those who do not. The CPR procedure is shown here only sketchily; definitive instructions should be obtained through a public-health organization or the fire department.

be a number of different etiologic syndromes. Nevertheless, it is possible to divide sudden cardiac deaths into three broad groups according to the immediate cause of the rhythm disturbance: those associated with myocardial infarction, those associated with myocardial ischemia and those associated with poorly understood primary cardiac-rhythm disturbances.

Regardless of the contributing or immediate causes of sudden cardiac death, modern emergency resuscitation procedures can hold off biological death long enough so that life can be restored. The speed with which these procedures are applied directly determines the victim's chances of survival.

The immediate treatment of sudden cardiac death is cardiopulmonary resuscitation (CPR), a repeated series of mouth-to-mouth respirations and chest compressions that circulates a small amount of oxygenated blood (approximately 5 to 30 percent of normal cardiac output) to the brain, heart and other vital organs of the body. CPR is then followed by definitive medical treatment, which seeks to restore normal circulation and respiration. The treatment includes insertion of a breathing tube into the trachea, delivery of drugs and defibrillation.

Defibrillation interrupts the chaotic twitching of the heart muscle and gives the heart's normal electrical activity a chance to take over the timing of cardiac contractions. It is achieved by applying an electric shock across the victim's chest. Such a shock depolarizes all heart cells simultaneously and essentially "resets" the pacemaking nodes of the heart.



CORONARY HEART DISEASE manifests itself every year in the U.S. as approximately 1.5 million heart attacks caused by partial or complete blockage of the coronary arteries and arrhythmias (cardiac-rhythm disturbances). About a fourth of the heart attacks can be classified as out-of-hospital sudden cardiac deaths: instances in which coronary heart disease has precipitated the cessation of circulation and breathing with little or no warning. Currently fewer than 5 percent of sudden cardiac death victims are successfully resuscitated.

As recently as the early 1960's community emergency medical services were seriously deficient in providing that kind of cardiac care. Prehospital emergency care was mainly given by poorly equipped private ambulance and mortuary companies, whose vehicles often doubled as hearses and whose personnel were only haphazardly trained in first aid. Reports from the National Academy of Sciences and the President's Commission on Highway Safety confirmed the inadequate levels of emergency medical care and the unevenness of personnel competency.

It was more the desire to reduce mortality from traffic accidents than the desire to save heart-attack victims that stimulated improvements in emergency care during the mid-1960's. Nevertheless, cardiac-arrest victims benefited from the speedier emergency treatment. Emergency medical services were encouraged by the Federal Government through the National Highway Safety and Traffic Act of 1966. An official 81-hour training course for emergency medical technicians (EMT's) was soon developed to instruct such personnel in basic life-support techniques, including CPR.

Although it was a known fact that more than half of the deaths related to heart disease occurred outside the hospital within one or two hours of the onset of symptoms, advanced coronary care was not incorporated as a specific component of emergency medical service until the latter half of the 1960's. Mobile intensive-care units capable of handling incidents of sudden cardiac death were first set up in Belfast, Northern Ireland, by J. Frank Pantridge and J. S. Geddes of the Royal Victoria Hospital in 1966. These units were staffed with a physician and a nurse who were familiar with defibrillation and the administration of antiarrhythmic drugs, as well as treatment for myocardial infarction. Similar emergency-care vehicles soon were seen on U.S. streets, but instead of being manned by hospital staff the vehicles were manned by paramedics: lay personnel who have had 400 to 1,500 hours of instruction on how to treat out-of-hospital cardiac arrests as well as other acute illnesses and medical emergencies. The first such paramedic programs, those in Seattle, Miami, Los Angeles and Columbus, were respectively established by Leonard A. Cobb of the University of Washington School of Medicine, Eugene L. Nagel of the University of Miami School of Medicine, J. Michael Criley of the University of California at Los Angeles School of Medicine and James V. Warren of the Ohio State University College of Medicine.

The development of paramedic programs and the expansion of EMT training gained substantial support in 1973, when Congress passed the Emergency Medical Service Systems Act, which was renewed in 1976 and again in 1979. The original legislation provided $185 million for the development of EMT and paramedic programs in 300 regions nationwide; it also encouraged an integrated systems approach to emergency medical care.

As a system, emergency medical services are a complex integration of various elements: emergency-access telephone numbers, trained personnel, specially equipped vehicles, supervising physicians and telemetry links, by

which vital physiologic data can be transmitted to hospitals. Although emergency medical services are designed to deal with all kinds of medical crises—not just cardiac arrest—their ability to respond quickly makes them a critical part of any program to aid victims of sudden cardiac death.

Currently no more than 5 percent of sudden cardiac death victims in the U.S. are resuscitated and survive until they are discharged from the hospital. The low national survival rate reflects the fact that in the vast majority of incidents emergency medical service arrives too late, if at all. The percentage of survivors can be substantially raised through a concerted national effort. Because the absolute number of cases is very large, even minor improvements in the speed with which CPR and defibrillation are begun on cardiac-arrest victims would result in a large number of saved lives.

Many circumstantial factors, such as the victim's age, whether the collapse was witnessed and whether he was fortunate enough to be given CPR by an onlooker, have great bearing on the probability of successful resuscitation from sudden cardiac death. Such factors are unpredictable, however, and can only be altered with difficulty. Other factors that determine a victim's chances of survival are directly related to the structure of the emergency cardiac-care program, and the structure can often be significantly improved. It is primarily differences in emergency care that account for the wide discrepancy among communities in reported survival rates from out-of-hospital sudden cardiac death: from a low of 3 percent to a high of 30 percent. The controllable factors that can produce the most immediate change in the percentage of successful resuscitations are the time intervals between collapse and the start of CPR and between collapse and the initiation of definitive therapy (namely defibrillation).

Communities have diverse demographic structures, geographies and financial resources, and they have developed different types of emergency medical services. Emergency services can be provided either by units of regular EMT's, specialized EMT's equipped for on-site defibrillation or paramedics, or by combinations of these units.

Not surprisingly, studies have revealed that communities with only ba-

CHANCE OF SURVIVAL RESUSCITATION (*colored bars*) **from sudden cardiac death depends critically on the particular circumstances surrounding the cardiac arrest and on the speed with which definitive care can be provided by paramedics or physicians. The nature of the arrhythmia that results in the victim's loss of circulation and the presence of bystanders who witness the victim's ensuing collapse cannot be controlled. The response time of paramedic units and the likelihood that a bystander will be able to perform CPR can be improved by increasing the number of emergency-cardiac-care vehicles and personnel and educating the public in CPR.**

sic EMT units experience poor survival rates for sudden cardiac death cases. Because EMT's can provide only CPR at the scene of the collapse and must transport the victim to a hospital for more advanced care, it may be more than 20 minutes before a sudden cardiac death victim can receive defibrillatory shocks.

Programs that provide the technicians with defibrillators improve the survival rate significantly. The addition of defibrillators to EMT units produced a nearly fivefold increase in survival rates in a Seattle suburb that had no other advanced emergency care. A study also indicated that this single step resulted in similar benefits for rural Iowa communities.

Emergency-care programs that rely entirely on paramedics achieve higher survival rates than programs that rely exclusively on basic EMT's. Yet the cost of the extensive training and more specialized equipment (including defibrillators) has limited the number of paramedic programs.

The highest survival rates are recorded in communities employing a combined system of EMT's and paramedics such as those found in Seattle, Miami and Columbus. The technicians arrive three or four minutes after collapse and start CPR; several minutes later paramedics arrive to give more advanced care, primarily defibrillation. Even better survival rates are possible if the EMT personnel are also equipped with defibrillators. The EMT units can then start CPR and administer defibrillatory shocks if the victim's condition warrants such treatment. When the paramedics arrive sometime thereafter, they can administer medications, insert airway tubes and deliver additional defibrillatory shocks as needed. Such a combined program, begun in 1979 in Washington's King County, has proved to be successful: the survival rate of out-of-hospital cases of ventricular fibrillation in which more than four minutes elapsed between the arrival of EMT's and the arrival of paramedics was increased from 18 to 38 percent.

In general, adding more vehicles, ancillary equipment and trained personnel will shorten the time between collapse and the initiation of CPR and the time between collapse and the provision of definitive care. After a certain point, however, there are diminishing marginal returns: further improvements in response time may require substantial investment in vehicles and personnel with little or no gain in survival rates.

For example, an average response time of four minutes has been realized by utilizing 100 fire-department emergency-care vehicles distributed throughout the 500 square miles of King County. We calculated that lowering the response time by an additional 30 seconds would require the addition of 31 new vehicles, each costing about $43,000 per year (assuming no additional personnel are necessary). Adding paramedics is even more expensive: in King County it costs $500,000 to keep a paramedic unit in operation for one year. A cost-effective option in such a situation is to upgrade basic EMT units by equipping the vehicles with defibrillators. Because no additional personnel would be required, more rapid defibrillation could be achieved at a relatively modest cost.

	MYOCARDIAL INFARCTION	MYOCARDIAL ISCHEMIA	PRIMARY RHYTHM DISTURBANCE
PERCENTAGE OF SUDDEN CARDIAC DEATHS	20-30	?	?
EVIDENCE UNCOVERED THROUGH AUTOPSY	CORONARY ARTERY OCCLUSION; SEVERE CORONARY ARTERY DISEASE	CORONARY ARTERY DISEASE	APPARENTLY NORMAL HEART IN A SMALL PERCENTAGE OF CASES, OTHERS SUFFERED FROM CORONARY ARTERY DISEASE
DURATION OF WARNING SYMPTOMS PRIOR TO COLLAPSE	MINUTES TO HOURS	MINUTES	SECONDS; NO WARNING AT ALL
LONG-TERM PROGNOSIS AFTER SUCCESSFUL RESUSCITATION	GOOD	GOOD	POOR

THREE CAUSES OF HEART MALFUNCTIONS that lead to sudden cardiac death are myocardial infarctions (the death of heart-muscle tissue due to lack of oxygen) and ischemia (reduced supply of oxygen and nutrients to the heart muscle) and poorly understood cardiac-rhythm disturbances. Sudden cardiac deaths are grouped according to these causes.

Cost is not the only factor that may make the addition of EMT or paramedic units prohibitive. In many communities geography, high-rise buildings and the layout of roadways put limits on response times that cannot be substantially lowered regardless of the number of additional units. Nevertheless, there are other strategies, some of them readily implemented and inexpensive, that can shorten the response time of emergency care.

For example, simply improving people's access to emergency-communication centers and increasing efficiency in dispatching can save valuable seconds in response time. Many communities have therefore improved access by setting up a single, easy-to-remember telephone number (911 in the U.S.) through which all emergencies can be reported. Computer-assisted dispatching systems have also been developed that can rapidly pinpoint addresses and identify the closest available emergency-care vehicle.

Another measure that yields some benefit with relatively little investment is the establishment of CPR training programs. Such a program was first set up in 1971 by Cobb in Seattle. Since then many other communities throughout the country have instituted similar programs, with training provided either as a public service or for a small fee. The assumption is that large numbers of trained individuals dispersed through a community will improve the chance that a victim will receive early CPR. Victims who receive CPR immediately after collapse have roughly twice the chance of survival of those victims who do not. The American Heart Association and the American Red Cross have actively supported these training efforts. The sessions, which last for from three to nine hours, also present an ideal opportunity to teach participants emergency telephone numbers, risk factors of coronary heart disease, preventive health care and warning signs of myocardial infarction.

Community-wide CPR instruction has achieved only limited success, however. Our experience in the King County area is illustrative: although the percentage of the population trained in CPR has increased sixfold in

the past decade, the percentage of cardiac arrests in which bystanders initiated CPR has leveled off at a fairly constant 30 percent in the past few years. One reason is that the majority of those who voluntarily seek CPR training are young (the average age is 33 years) and are evenly divided between males and females. The majority of cardiac-arrest victims, however, are men over the age of 50, who usually collapse at home. The spouses of cardiac-arrest victims, generally women over 50, are therefore the ones who could make the most use of CPR skills.

Clearly, new approaches are needed to reach those who are most likely to witness a cardiac arrest. In recent years alternatives to traditional CPR courses have been instituted in an effort to give as many people as possible some information about what to do in the event of sudden cardiac death. CPR demonstrations have been shown during nationally televised baseball games, and serial public-service spots have been developed for national television. It has long been proposed that families of coronary-heart-disease patients be targeted as prime students for CPR instruction. In spite of these commendable efforts and the national CPR campaigns run by the Heart Association and the Red Cross, only a small percentage of the population most likely to witness a cardiac arrest have received adequate training.

One alternative that aims CPR information at those who could apply it most immediately was adopted by the fire department of Phoenix, Ariz., in 1974: telephone instruction in CPR by emergency medical service dispatchers. The person who reports an incident of cardiac arrest is in the best position to come to the victim's immediate aid. If the dispatcher can guide the caller through the CPR procedure step by step while professional medical help is on the way, precious time can be saved. The Phoenix program has not been formally evaluated, but anecdotal reports suggest it has been possible to give adequate guidance over the telephone.

William B. Carter and his colleagues at the King County Emergency Medical Services Division of the Seattle–King County Health Department have developed a CPR-instruction message that can be transmitted by telephone to a caller reporting a cardiac arrest. All King County dispatchers were trained to give the instructions. The message is in two parts: the caller is first told how to perform artificial respiration, and he must then return to the telephone for chest-compression instructions. The message is delivered in simple and direct phrases designed to give a clear picture of the action to be taken.

Of the more than 400 cases of telephone CPR studied to date, no serious adverse effects have been observed either in CPR-related injuries (determined by a review of medical and autopsy records) or in the psychological impact on the recruited caller (determined by in-depth interviews). There appears to be a positive psychological effect associated with helping a cardiac-arrest victim, regardless of the outcome. Telephone CPR programs have recently been endorsed by the American Heart Association and have been instituted elsewhere in the U.S. as well as in England and the Netherlands.

There is no question that the sooner CPR is initiated, the greater the likelihood of survival for the victim is. Yet CPR is no more than temporary aid; the victim needs the more definitive care that paramedics and physicians can give for a better chance of survival. Although defibrillation is only one component of definitive care, it is probably the most important; when it is provided rapidly, it may be all that is necessary to save the victim's life. Increasing the general availability of defibrillators in the community can therefore result in dramatic increases in survival rates.

Sudden cardiac death associated with ventricular fibrillation usually does not take place near paramedic vehicles or in hospitals, where defibrillators are standard equipment; it occurs primarily in the home. It is not feasible to supply conventional defibrillators to residences of coronary-heart-disease patients because such machines require the ability to interpret cardiac rhythms, a skill that can be acquired only through special training. The need for extensive training and the potential harm to the operator of a defibrillator if an accident occurs has kept defibrillation a specialist's skill.

Newly developed automatic defibrillators may soon change the situation,

EMERGENCY-VEHICLE PERSONNEL	NUMBER OF VEHICLES RESPONDING TO EMERGENCY CALL	AVERAGE TIME BETWEEN COLLAPSE AND CPR (MINUTES)	AVERAGE TIME BETWEEN COLLAPSE AND DEFIBRILLATION (MINUTES)	AVERAGE TIME BETWEEN COLLAPSE AND ADVANCE LIFE SUPPORT (MINUTES)	RATE OF SUCCESSFUL RESUSCITATIONS (PERCENT)
EMT	1	4	20 (AT HOSPITAL)	20 (AT HOSPITAL)	3
EMT-D	1	4	4	20 (AT HOSPITAL)	20
PARAMEDIC	1	8	8	8	10-20
EMT AND PARAMEDIC	2	4	8	8	15-20
EMT-D AND PARAMEDIC	2	4	4	8	20-30

EMERGENCY CARDIAC CARE can be provided by a regular emergency medical technician (EMT), by a specialized EMT trained in the use of defibrillators (EMT-D) or by a paramedic. An EMT typically receives 81 hours of training in basic life-support techniques (primarily CPR); an EMT-D has an additional 10 hours of instruction in the operation of defibrillators. Because paramedics are trained to be able to provide more advanced life support, including the administration of drugs and endotracheal intubation, their training requires some 400 to 1,500 hours. An emergency medical service consists of combinations of these health-care specialists, each team assigned to specially equipped vehicles. The high cost of a paramedic unit compared with an EMT unit means that a community can afford fewer paramedic units than EMT units, so that it takes longer on the average for paramedics to arrive at the scene of a cardiac arrest. Because the faster a cardiac-arrest victim receives medical treatment (both basic and advanced) the greater his chances of resuscitation are, emergency medical service programs that combine EMT's or EMT-D's with paramedics have proved to be most effective.

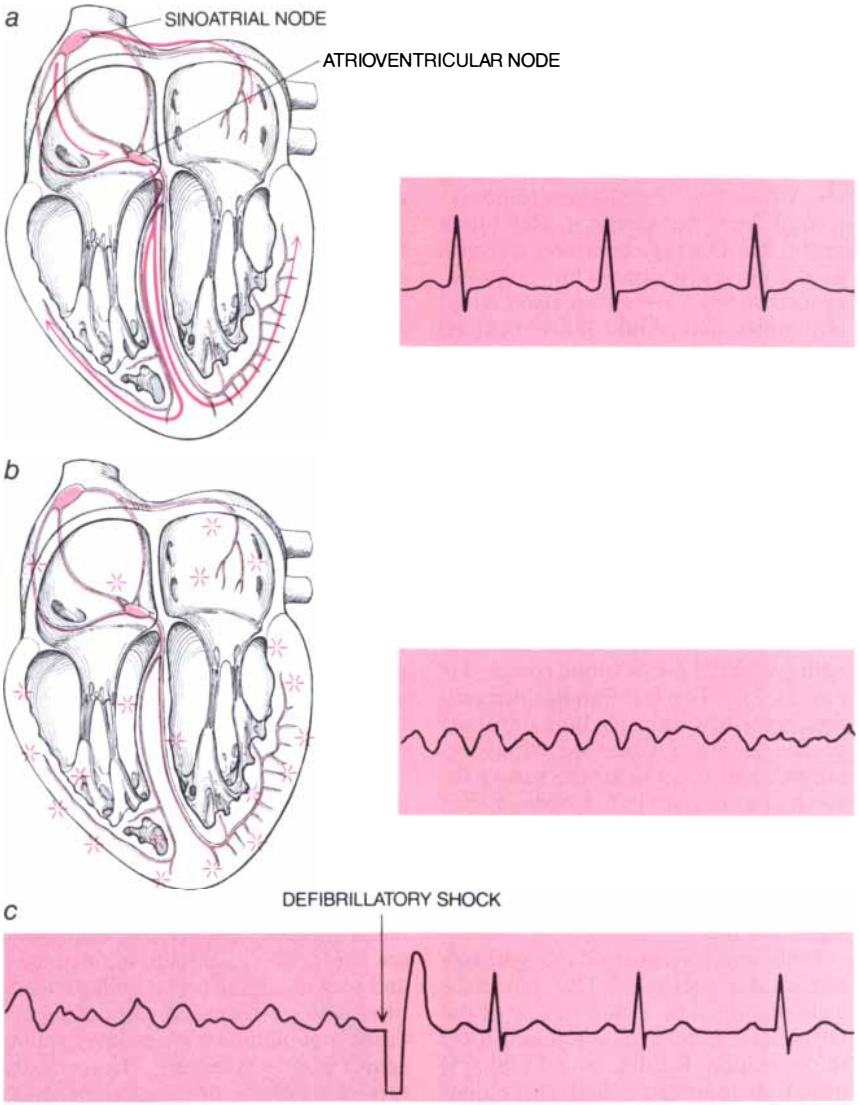
however. The devices are easy to operate: a rescuer attaches adhesive electrodes to the victim's chest and the machine does virtually all the rest. The automatic defibrillator is programmed to detect and recognize ventricular fibrillation, charge capacitors with the proper current and then deliver a defibrillatory shock. Synthesized-voice or liquid-crystal-display messages help the operator to set up the device. Spouses of potential victims can learn to attach the electrodes and to use the device. Studies suggest that the device is highly effective: in more than 95 percent of the cases ventricular fibrillation was correctly identified. In no case was a shock delivered when no fibrillation was present.

Although supplying portable, easy-to-use defibrillators to the homes of all patients with coronary heart disease could save thousands of lives each year, such an option is not realistic. It is precluded for most communities because of the expense. (The devices cost from $4,500 to $7,000, although with increased demand they may become less expensive.) Moreover, such a measure would not profit those potential sudden cardiac death victims who are unaware of their heart disease.

A more practical approach might be to put automatic defibrillators in public places and to train members of the community in their use. Defibrillators on airplanes and ships, in office and residential buildings and in public facilities might have a significant impact on the speed with which victims receive defibrillation. Many questions will have to be answered, however, before the effectiveness of public access to defibrillators can be determined. What are the estimated benefits and costs? Who will pay? How frequently should people be retrained in their use? Is a lay person likely to use the device in a real emergency?

One last alternative provides the shortest time possible to defibrillation: surgical implantation of automatic defibrillators. Mieczyslaw Mirowski and his co-workers at the Johns Hopkins University School of Medicine have developed an implantable device that continuously monitors heart rhythm, identifies ventricular fibrillation and delivers low-energy defibrillatory discharges, all within 15 to 20 seconds. Through 1983 more than 220 patients had such defibrillators implanted in their bodies, and there have been many successful conversions of ventricular fibrillation. This option, however, can really be recommended only to the small proportion of identifiable patients who are at risk of imminent death from ventricular fibrillation.

In spite of the fact that a number of emergency-medical-care options can decrease mortality from sudden cardiac death, such steps represent only a partial solution to the problem. The ultimate solution is the prevention of heart disease. An effective means of combating atherosclerosis, the leading cause of heart disease, is still elusive. Changes in lifestyle and diet, increased exercise, control of hypertension and new medications can all help, but for now efforts must be directed toward speeding the delivery of medical treatment to victims of sudden cardiac death. If resuscitative techniques already shown to be effective, namely rapidly provided CPR and defibrillation, could be more widely implemented, whether by emergency medical technicians and paramedics or through other public-health programs, thousands of additional lives per year could be saved.



VENTRICULAR FIBRILLATION is the commonest arrhythmia associated with sudden cardiac death. Normally the electrical impulse that causes the heart-muscle cells to contract (*a*) originates in the sinoatrial node, travels to the atrioventricular node and thence spreads symmetrically through the muscular walls of the two ventricles by way of embedded nerve fibers. As reflected in the even spacing of the regular electrical-signal complexes recorded on an electrocardiogram, the chambers of a healthy heart undergo coordinated and rhythmic contractions. During ventricular fibrillation (*b*) sporadic electrical impulses arise from multiple sites in the heart. The electrocardiogram of a fibrillating heart indicates that the contractions are uncoordinated, resulting in immediate loss of blood pressure and pulse. Ventricular fibrillation can be stopped by applying an electric shock across the chest wall and through the heart muscle (*c*). The shock depolarizes all heart-muscle cells and allows the natural rhythmic output of the pacemaking nodes to take over. New devices are able to recognize the electrocardiographic signature of fibrillation and automatically deliver shocks to counter it; they could enable lay people to apply this lifesaving measure.

Glycosphingolipids

The composition of these membrane molecules changes dramatically with cell differentiation and the onset of cancer. Exploiting such changes could lead to improved diagnosis and treatment of cancer

by Sen-itiroh Hakomori

In 1951 a 66-year-old woman entered a hospital in Charlottesville, Va., to have a malignant tumor removed from her stomach. Her blood group was *O,* a type commonly known as the universal donor, but laboratory tests showed her serum also carried antibodies that would kill almost all types of blood cells except her own. No donors could be found whose blood was compatible with the antibodies; the potential need for blood transfusion during her impending surgery therefore placed her at considerable risk. To assess the risk she was given a small trial transfusion, amounting to about 25 milliliters of type *O* blood.

Predictably, the woman's initial immune reaction was dramatic. The concentration of the serum antibodies reacting with the donor blood rose to one part in 512. The test findings directed the choice of surgical action: only part of her stomach could be safely removed, and so some of the tumor tissue had to remain in her body. Yet to the great surprise and relief of everyone the remaining tumor tissue disappeared after the surgery. The woman lived to the age of 88 with no further signs of cancer.

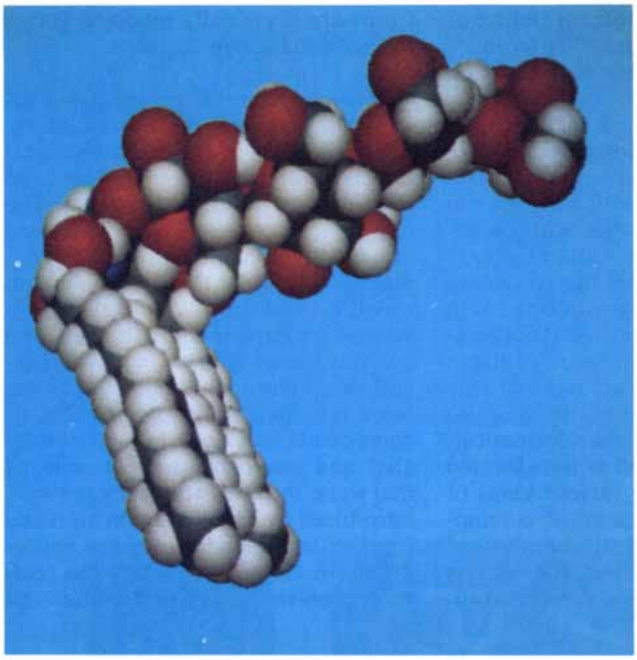
What was the cause of the woman's remarkable recovery? The first clues were identified by Philip Levine of the Ortho Research Foundation soon after the operation. Routine blood typing is done only for the so-called *ABH* blood-group system, in which blood and tissue are classified according to the clinically familiar types *O, A, B* and *AB*. There are many other blood-group systems, however, and the expression or suppression of various antigens belonging to each system determines a separate, independent blood type. Within the so-called *P* blood-group system the woman was the first person ever documented whose normal tissues do not display two kinds of specific immunological marker, the *P* and P_1 antigens. Such a lack of immunological marking is the defining characteristic of a rare blood type within the *P* system known as the *p* blood group, which has since been found in only one out of 100,000 people in the general population. When Levine tested the woman's blood serum, he made another striking observation: the serum carried immunoglobulin *G* antibodies to both *P* and P_1 antigens. The formation of such antibodies must have been induced by *P* and P_1 antigens. Where had they come from? Levine reasoned they had probably been expressed by the tumor tissue.

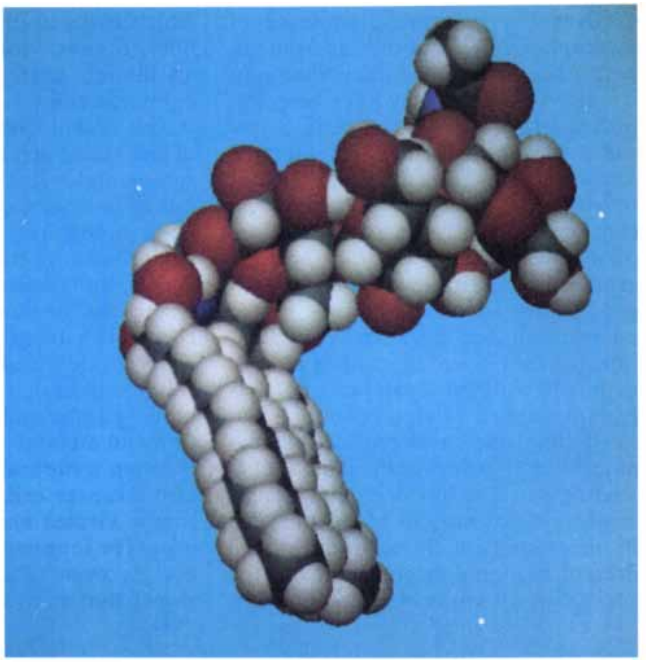
Thirty years after the surgery our laboratory analyzed a sample of the woman's tumor tissue, which had been freeze-dried and stored in a tissue archive at Levine's laboratory. Our analysis showed that the tissue did indeed contain two kinds of antigen: one that reacted with anti-*P* antibody and another that reacted with anti-P_1 antibody. We also found that both antigens are members of a class of molecules called glycosphingolipids, which are embedded in the membranes of cells throughout the body. As the name implies, the glycosphingolipids are part lipid, or fat-soluble substance, and part sugar; the prefix *sphingo-* identifies them as a special class of glycolipids that include a lipid called sphingosine in their structure. Thus we confirmed Levine's previous conclusion that the formation of antibodies in the woman's serum had been induced by incompatible antigens in the tumor.

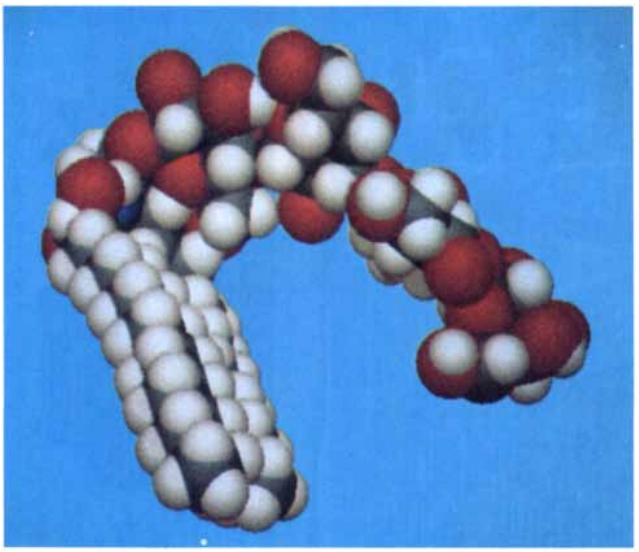
The woman's strong immune reaction to the donor blood was brought about by the presence of anti-*P* and anti-P_1 antibodies in her system, which she could tolerate only because her healthy tissues did not carry *P* and P_1 antigens. When she received additional *P* and P_1 antigens with the transfusion, the antigens stimulated the rapid production of additional anti-*P* and anti-P_1 antibodies. The high level of antibodies presumably triggered a complex reaction by the cells of her immune system, which selectively destroyed the tumor cells displaying the incompatible *P* and P_1 antigens.

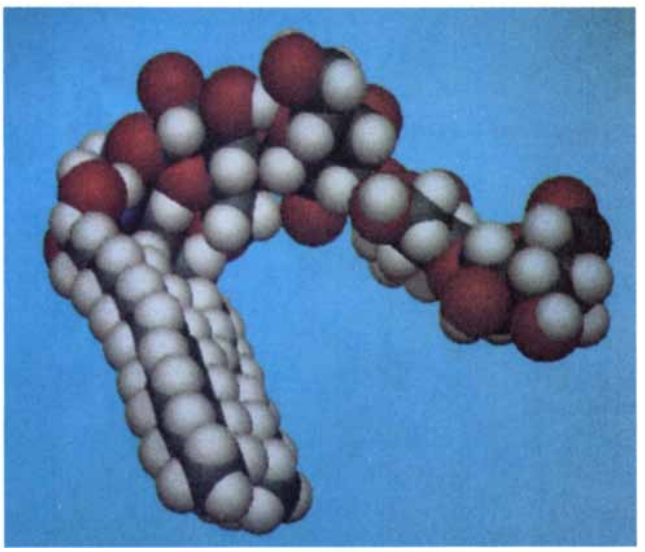
It has been known for some time that glycosphingolipids are found in all animal cells and in some plant cells. Until recently, however, investigators had paid little attention to these molecules because their biological functions had not been clearly defined. In the past 10 years a number of studies have firmly established the concept that blood-group antigens and many other antigens that are modified or inappropriately expressed in the development of malignant tumors are glycosphingolipids. Consequently the glycosphingolipids have become objects of intense scrutiny by cancer immunologists, and much has been learned about their functions in ordinary cell life as well as their role in cancer and other diseases.

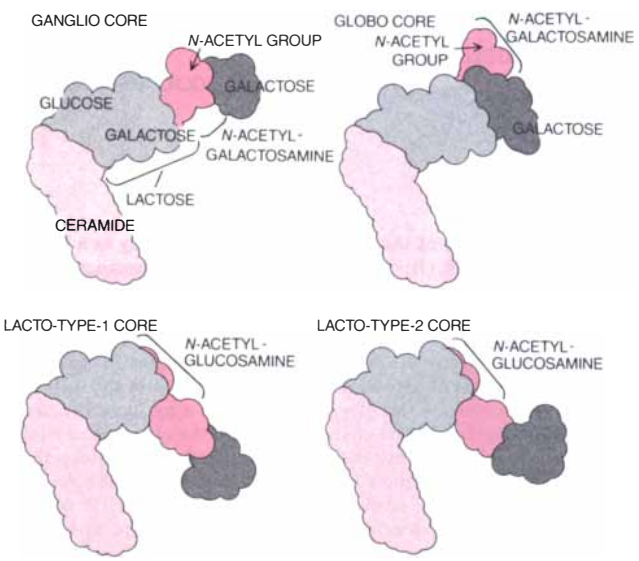
By virtue of their distribution on the membrane of the cell, glycosphingolipids act in many ways to regulate the interactions of the healthy cell with its environment. For example, they serve as distinguishing markers for cells from various organs of an animal, and so they mediate cell-to-cell recognition and communication. Moreover, the expression of glycosphingolipids on the cell surface changes as the cell divides and differentiates, which suggests they are essential for the systematic growth and development of the organism. It is known that they regulate growth in at least two ways. By sensing the density of similar cells in their vicinity they can promote or inhibit cell division. By interacting with receptor proteins on the cell surface they can also inhibit the response of the cell to growth factors that might be found in the medium surrounding the cell.









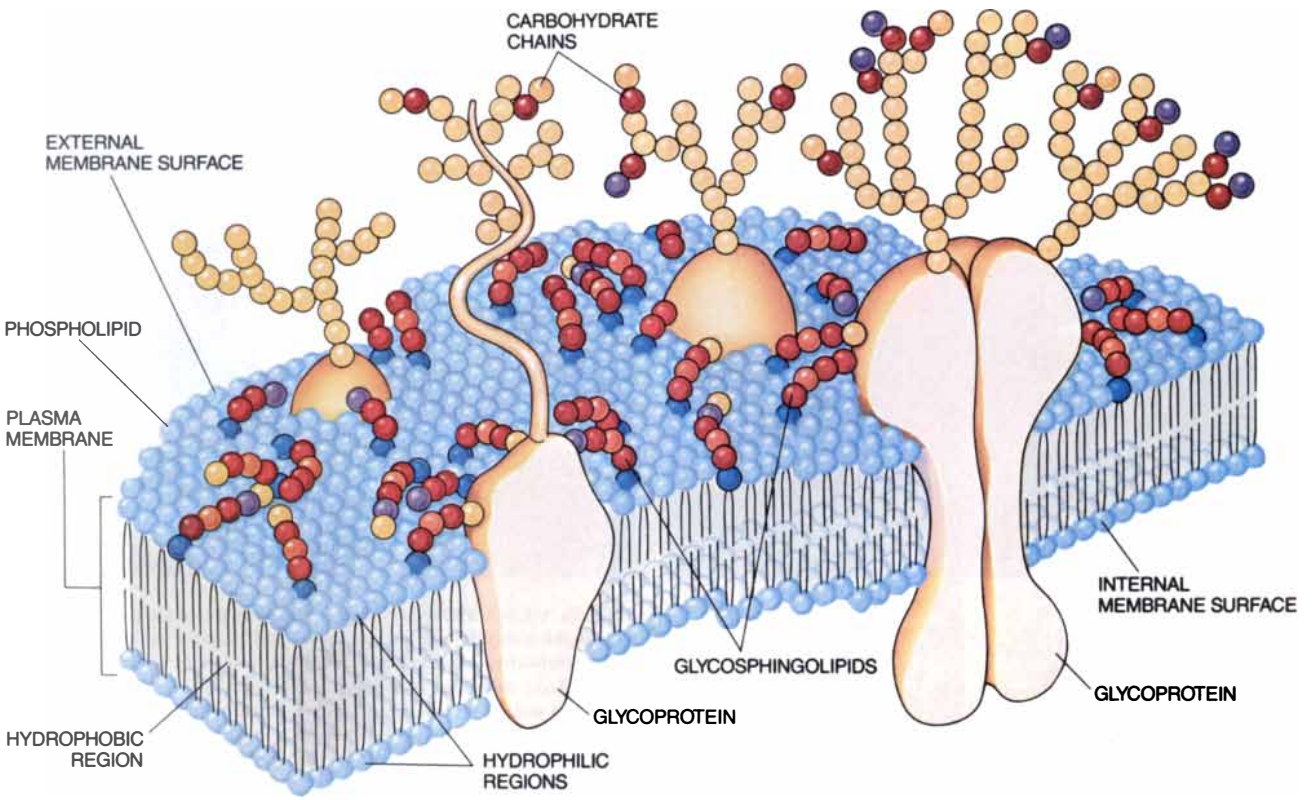


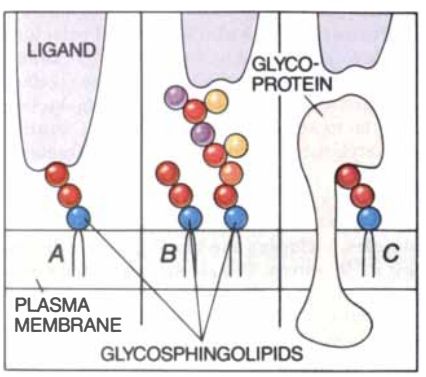
MOLECULAR MODELS depict the four basic core structures of glycosphingolipids, a class of molecules usually embedded in the cell membrane. All glycosphingolipids are made up of a carbohydrate chain attached at a right angle to a lipid called ceramide. Most but not all glycosphingolipids share one of the four basic core structures, which are defined according to the identity and chemical bonding of the sugars closest to the ceramide. In each of the four core structures ceramide is linked to glucose and the glucose is linked to galactose. In the ganglio-series glycosphingolipids (*upper left*) the chain is continued with a sugar called *N*-acetylgalactosamine, to which another galactose is attached. The internal galactose and the terminal galactose are each linked to an acidic sugar called sialic acid (not shown). In the globo series the initial glucose-galactose chain is continued with another galactose and *N*-acetylgalactosamine (*upper right*). In the lacto series the initial chain is continued with a sugar called *N*-acetylglucosamine and galactose. Depending on the position of the last bond, the lacto-series molecules are further classified as type 1 (*lower left*) and type 2 molecules (*lower right*). The lacto-series glycosphingolipids can lengthen and branch to yield a variety of molecules, including the blood-group antigens of the clinically important *ABH* system. The models show the minimum-energy conformation of the molecules; they were constructed by Steven B. Levery of the author's laboratory and Ronald E. Stenkamp and Keith D. Watenpaugh of the University of Washington.

Given the widespread presence of glycosphingolipids in cells and the importance of their regulatory functions, it is not surprising they have been implicated in a number of serious diseases. It is thought they are infection sites on the host cell for several kinds of virus and bacterium. Their role in cancer is aptly illustrated by the case of the woman with the rare *p* blood group: the kinds of blood-group antigens expressed on the surface of the normal cell and the relative populations of each kind of antigen change in the cell with the development of malignant tumors. To appreciate this basic finding one must realize that the name blood-group antigen is a misleading one. The blood-group antigens are found not only in blood but also in many kinds of tissue, and they are present in high concentrations on the surface of all kinds of epithelial cells, which make up the mucous linings of many organs. More than 90 percent of all human cancers are derived from epithelial cells.

The recent chemical identification of the blood-group antigens as glycosphingolipids links the powerful methods of immunology with a substantial body of diverse biochemical knowledge. One important result of the linkage is the realization that observed changes of the blood-group antigens in cancer can be generalized. Biochemists have identified several other glycosphingolipids that are not related to the blood-group antigens, whose composition and metabolism alter dramatically when a normal cell is transformed into a cancer cell by various kinds of tumor viruses and chemical carcinogens. The immunochemical analysis of the glycosphingolipids in tumors has shown that many tumor-associated antigens are chemically modified forms of the blood-group antigens.

Such important experimental findings have led to substantial renewed interest in the basic chemical structure of glycosphingolipids and their role in the life of the cell. The first glycosphingolipid was discovered in 1874 by Johann Ludwig W. Thudichum in brain tissue; Thudichum called it cerebroside. More recent investigators have found that brain and nervous tissues are rich in glycosphingolipids; indeed, for many years the most intensive study of the class of compounds was done by neurobiologists and neurochemists. In spite of that work, the romantic prefix *sphingo-*, introduced by Thudichum in his name for the compounds, remains appropriate. In Greek mythology the fearsome monster Sphinx, part woman and





PLASMA MEMBRANE, or outer envelope of the cell, is shown schematically as a bilayer formed primarily of phospholipid molecules. Glycosphingolipid molecules, shown as strings of colored balls that represent various sugars, are embedded in the outer layer of the membrane. The hydrophilic sugars lie along the surface of the membrane in the watery medium surrounding the cell, and they are perpendicular to the hydrophobic ceramide region of the molecule that anchors it in the bilayer. Many other molecules are embedded in the membrane, including glycoproteins, other kinds of glycolipids and cholesterol (not shown). The peripheral part of the carbohydrate chain of the glycosphingolipids may have the same structure as the peripheral part of the carbohydrate chain attached to the glycoproteins; the coloring indicates the similarities. The diagram at the left shows the effects of interactions of glycosphingolipids and other molecules on the surface of the cell. Glycosphingolipids whose carbohydrate chains are exposed at the surface can serve as receptors for ligands that bind to the cell (*a*). They can also become closely associated with other glycosphingolipids (*b*) or with membrane proteins (*c*). In the last two cases part of the carbohydrate chain can become inaccessible to potential ligands; the glycosphingolipid is said to be cryptic.

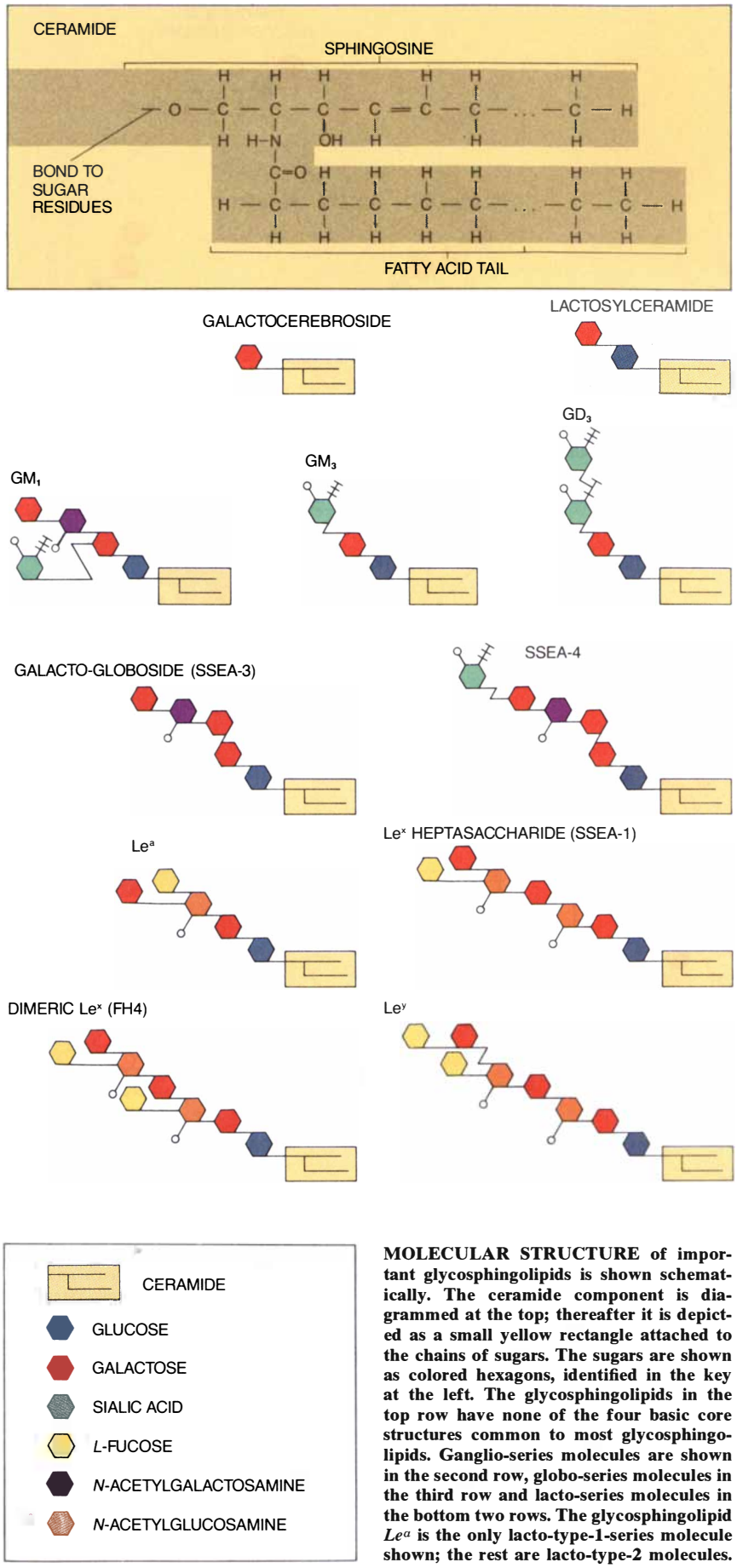
part winged lion, terrorized the city of Thebes by devouring all who could not correctly answer her riddles. The function of many glycosphingolipids in neuronal structures continues to be a riddle.

The molecular structure of the glycosphingolipids has been known for a long time. Each molecule is shaped somewhat like the letter L: one leg is formed by a carbohydrate chain, or chain of simple sugar residues, and the other leg is a lipid called ceramide [*see illustration at right*]. The two legs of the L thus form two distinct components. The carbohydrate chain is hydrophilic, and it interacts with the watery medium surrounding the cell. The second component, the fat-soluble ceramide, is hydrophobic, and so it tends to exclude water molecules nearby.

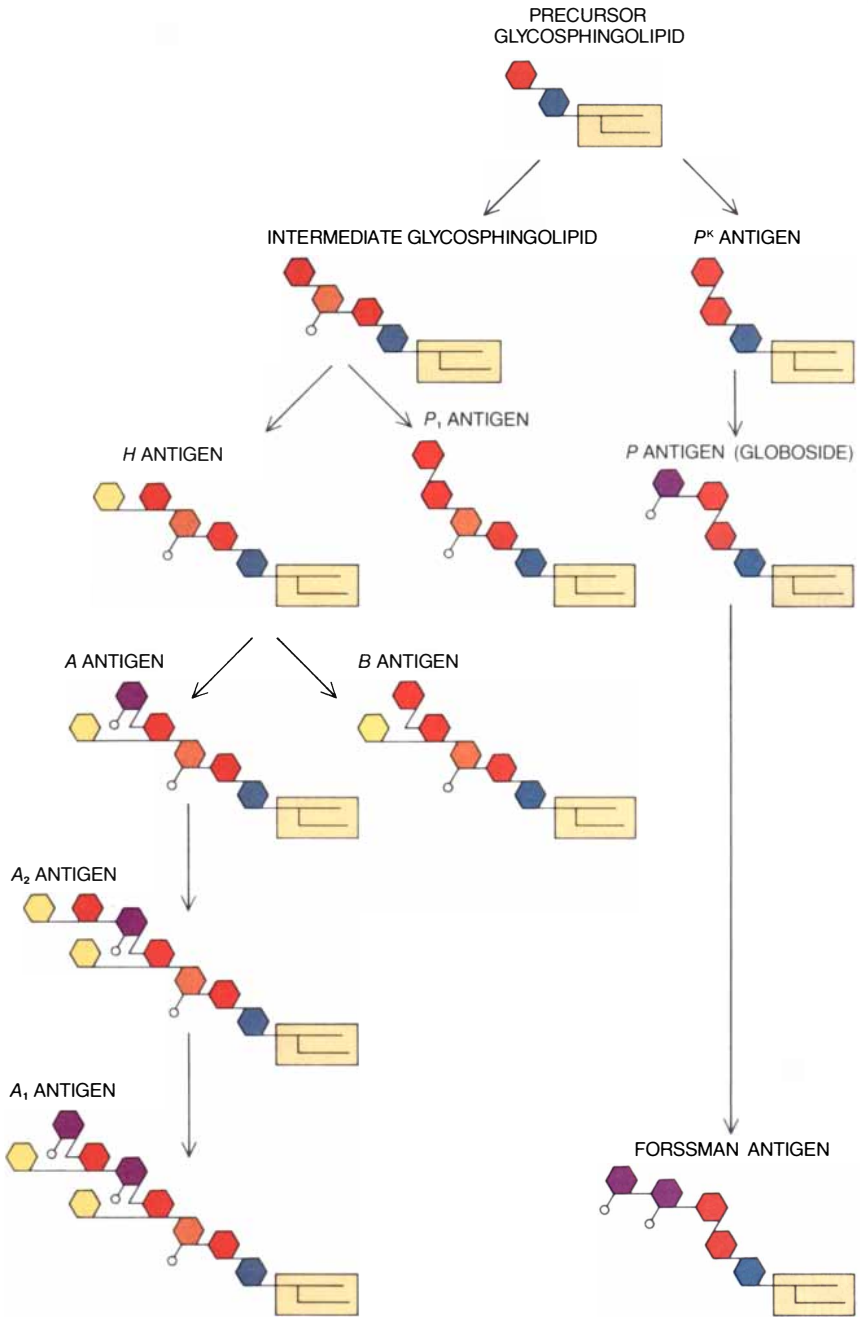
There is considerable variation of structure in both ceramide and the hydrocarbon chains, and about 130 varieties of glycosphingolipids are now known. About 40 of them are classified into the so-called ganglio structure, 10 are classified into the globo structure and 60 are classified into the lacto structure, depending on the sequence of sugars found in the core of the carbohydrate chains and the nature of the chemical linkages between the sugars [*see illustration on page 45*]. The remaining 20 glycosphingolipids are not so classified. The number of ways the components of a glycosphingolipid can be combined is limited because certain carbohydrate structures are preferentially linked to certain ceramides, depending on the cells and tissues in which they are found.

It is worth noting that the same sequence of sugars found in the lacto structure (but not in the ganglio or globo structures) is often bound to proteins. Because the sugar sequence—and not the substrate of ceramide or protein to which the sugars are attached—often determines the antigenic specificity, antibodies specific for the blood-group antigens can also recognize certain glycoproteins. The recognition of sugar chains on glycoproteins explains the presence of the blood-group antigens on mucous linings as well as in secretions.

The tissue-dependent distribution of glycosphingolipids is well illustrated by the gangliosides, an important class of glycosphingolipids discovered in 1936 by Ernst Klenk. He found gangliosides only in brain tissue, but other investigators have since found them on all cell membranes in all other tissues as well. Gangliosides are defined by the presence of an acidic sugar called sialic acid in their carbohydrate chains. Certain membranes of nerve



MOLECULAR STRUCTURE of important glycosphingolipids is shown schematically. The ceramide component is diagrammed at the top; thereafter it is depicted as a small yellow rectangle attached to the chains of sugars. The sugars are shown as colored hexagons, identified in the key at the left. The glycosphingolipids in the top row have none of the four basic core structures common to most glycosphingolipids. Ganglio-series molecules are shown in the second row, globo-series molecules in the third row and lacto-series molecules in the bottom two rows. The glycosphingolipid *Le*^{a} is the only lacto-type-1-series molecule shown; the rest are lacto-type-2 molecules.



BLOOD-GROUP ANTIGENS, which are of prime clinical importance in blood- and tissue-donor compatibility, are glycosphingolipids. The diagram shows the chemical and structural interrelations of two blood-group systems, color-coded according to the key in the illustration on the preceding page. The best-known blood-group system is the *ABH* system, but all the antigens belonging to that system share the same precursor molecule as the *P* blood-group system (*top*). All blood and tissue cells can be classified within each system. In the *ABH* system type *O* cells display an intermediate precursor molecule and the *H* antigen. In addition type *A* and type *B* cells display the *A* and the *B* antigens respectively. Type *AB* cells display all four antigens. In some type *A* people called type A_1 the *A* antigen can be elongated to form A_1 antigen. In other type *A* people called type A_2 the elongation of the chain is insufficient to form A_1 antigen, and so it forms A_2 antigen instead. The structures of the A_1 and A_2 antigens were determined by Henrik Clausen and Levery of the author's laboratory. In the *P* blood-group system the rare type P^k cells express only the P^k antigen and its glycosphingolipid precursor. Type P_2 cells display both the P^k antigen and the *P* antigen, or globoside, which is derived from the P^k antigen. Type P_1 cells display the P_1 antigen in addition to the *P*-system antigens displayed on type P_2 cells. The structure of the *P* system was established by Donald M. Marcus of the Baylor College of Medicine. About one in five people carry the Forssman antigen in tissue, which is derived from globoside.

cells are particularly rich in gangliosides whose carbohydrate chains include more than one kind of sialic acid. The patterns of their distribution vary from cell to cell and from species to species, as Tamio Yamakawa of the University of Tokyo suggested in 1952.

It is energetically favorable for the glycosphingolipid molecules to become embedded in the membranes of the cell, and most glycosphingolipids are found there. The cell membrane can be understood (in a simplified model) as a bilayer, or double layer, made up mainly of molecules called phospholipids and cholesterol. The phospholipid molecule, like the glycosphingolipid, has a hydrophilic head group and two hydrophobic hydrocarbon tails. In a laboratory container filled with water phospholipid and cholesterol molecules spontaneously form a spherical vesicle, or bubble-shaped compartment, whose thickness is two layers of molecules. The hydrophilic head groups on the inner layer face the water inside the vesicle, and the hydrophilic head groups on the outer layer face the water outside the vesicle. The hydrophobic tails of both layers of molecules forming the surface of the vesicle are thereby sequestered from the water.

The same general arrangement of phospholipid molecules is characteristic of the membranes of the cell, but many other substances, notably protein molecules and glycosphingolipids, may be embedded in the matrix of phospholipid and cholesterol. In general the plasma membrane that makes up the surface of the cell is much richer in glycosphingolipids than the membranes of intracellular organelles are, and glycosphingolipids in the plasma membrane seem to be limited to the outer part of the bilayer. The two hydrocarbon tails of the ceramide component are inserted into the hydrophobic interior of the membrane, and they add structural rigidity to the membrane. The carbohydrate chain lies along the outside of the cell surface, roughly perpendicular to the hydrocarbon tails. It is thereby exposed to extracellular substances unless it is partially obscured by proteins and other glycosphingolipids that may be embedded nearby in the membrane.

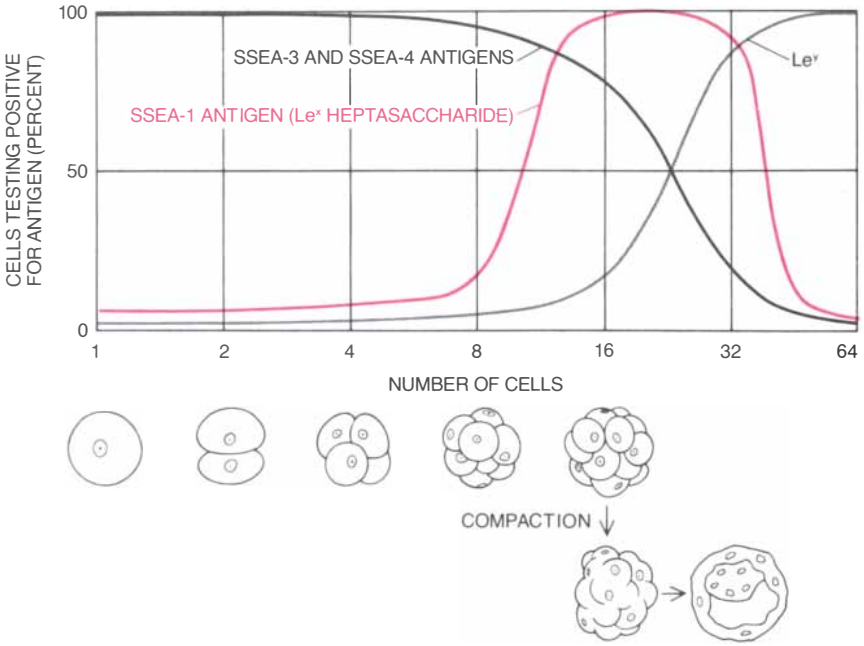
One of the most important consequences of the chemical identification of the blood-group antigens as glycosphingolipids is the realization that the antigens must be secondary gene products. In other words, their structure is not encoded in DNA the way the sequence of amino acids in a protein is. Instead a glycosphingolipid

is synthesized in a series of reactions catalyzed by multiple enzyme proteins called glycosyltransferases.

The sugar sequence is determined by glycosyltransferase enzymes in the membrane. The enzymes sequentially recognize a sugar nucleotide, or sugar donor molecule, as well as a precursor substrate to which a sugar residue from the nucleotide is to be transferred. The only molecules taking part in the process whose structures are transcribed directly from the DNA code are the glycosyltransferase enzyme proteins. The mechanism was initially proposed for the synthesis of gangliosides by Saul Roseman of Johns Hopkins University.

The synthesis of glycosphingolipids could be controlled in two ways. The first control is common to the synthesis of many other enzymes: the transcription of the sequence of DNA bases and their translation into enzyme proteins can be accelerated or inhibited by the activation or inhibition of special sequences of DNA called promoters or enhancers. Alternatively, the synthetic activity of the glycosyltransferase enzymes in the membrane could be rapidly altered by relatively minor chemical modifications of the enzyme molecules after the enzymes are synthesized. Chemical modification of the enzymes may influence their distribution within the membrane and their interactions with substrates there. Furthermore, the expression of glycosphingolipids on the surface of cell membranes depends not only on the rate of synthesis but also on the way they and other membrane molecules are assembled on the membrane. Their accessibility to antibodies or to other external substances can be affected by their proximity to other molecules [*see illustration on page 46*].

Because the expression of glycosphingolipids is controlled by more than one mechanism, the concentration of glycosphingolipids displayed on the surface of the cell may be much more sensitive to subtle shifts in the environment than, say, the concentration of a protein is. Rapid changes in the expression of glycosphingolipids could be brought about by a coordinated regulation of the genes that encode the enzymes needed for glycosphingolipid manufacture, by shifts in the competition among certain enzymes for available substrate or by a change in the orientation and exposure of glycosphingolipid antigens already on the cell surface. Investigators in my laboratory and elsewhere have now established that certain glycosphingolipids are found on specific kinds of cell for quite brief stages in embryonic development. The findings suggest glycosphingolipids are closely linked to the mechanisms of cell growth and differentiation. Cells may find it more efficient to change glycosphingolipids than to change proteins as they respond to the changing circumstances of development.



STAGE-DEPENDENT APPEARANCE and disappearance of glycosphingolipid antigens is plotted for the preimplantation mouse embryo. The *SSEA-1* antigen (the acronym stands for stage-specific embryonic antigen) appears in the eight-to-32-cell stage and declines greatly after the cells cluster tightly together in a stage called compaction (*colored curve*). When *SSEA-1* declines, the *Le^y* antigen appears (*gray curve*); it is chemically similar to *SSEA-1* in that it has an additional fucose sugar at the terminal galactose. The *SSEA-3* and *SSEA-4* antigens are highly expressed until the early stage of compaction, and they disappear almost completely by the 32-cell stage (*black curve*). Such glycosphingolipids may be closely associated with the regulation of cell-cell recognition and the growth of tissues.

It now appears that glycosphingolipids can affect the cell and its growth in two major ways. They may modulate the functions of proteins that reside in the plasma membrane, and they serve, along with the proteins, as distinctive surface markers necessary for maintaining proper intercellular communication. The first hints that glycosphingolipids may influence the functioning of membrane proteins came from evidence that glycosphingolipids and proteins are closely associated within the membrane. The proximity of glycosphingolipids and proteins in the membrane was demonstrated in studies by Clifford Lingwood of my laboratory and Tae Hwa Ji of the University of Wyoming. Ranwel Caputto and his colleagues of the Catholic University of Cordoba in Argentina then found that glycosphingolipids in the brain may activate ATPase, a protein that releases the energy needed for the transmission of messages along a nerve.

Receptor proteins embedded in the plasma membrane often serve as an essential link in the complex sequence of events that finally leads to mitosis, or somatic-cell division. Recent studies by Eric Bremer of my laboratory, in collaboration with Daniel F. Bowen-Pope, Elaine W. Raines and Russell Ross of the University of Washington, clearly indicate that certain glycosphingolipids can break the link and thereby inhibit cell growth.

Animal cells can generally undergo mitosis and proliferate only in the presence of specific, extracellular hormones called growth factors. The growth factors bind to specific receptor proteins at sites on the external side of the plasma membrane. The receptor protein is a large molecule that lodges in the membrane, partly inside and partly outside the cell. Its cytoplasmic part is an enzyme that catalyzes the attachment of phosphate groups within the cytoplasm to various proteins, including the receptor itself. The binding of the growth factor can lead to the activation of the enzyme associated with the cytoplasmic part of the receptor, and it induces the attachment of a phosphate group to the receptor. That

attachment appears to cause a conformational change in the receptor, and such changes allow receptors to aggregate on the membrane surface. The receptors and their attached growth factors are then taken into the cell.

The investigators found that when the glycosphingolipids GM_1 or GM_3 are added to cells proliferating in the presence of growth factors, the cell proliferation is inhibited. Moreover, in cells from membrane preparations whose proliferation was inhibited by GM_1 or GM_3, the receptor proteins do not readily become attached to a phosphate group. The effect seems to be specific for GM_1 and GM_3; no other glycosphingolipids led to the same results. We therefore believe the functioning of the receptor is sensitive to its glycosphingolipid environment in membranes. It is worth mentioning that when cells are transformed by tumor viruses, the resulting tumor cells show a much lower content of GM_3 in some kinds of tumors and a much lower content of GM_1 in other kinds of tumors. The reductions may be correlated with a loss in the control of the growth of the cells that is exercised by the glycosphingolipids.

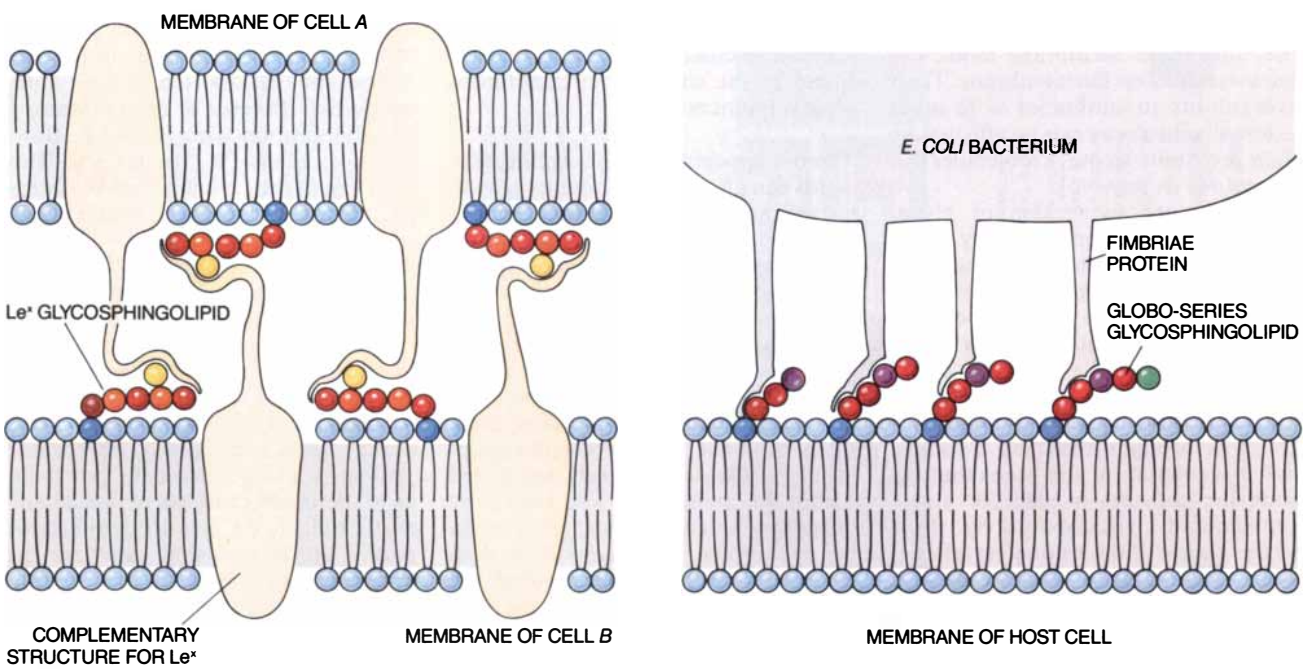
One must not conclude, however, that glycosphingolipids are always associated with the inhibition of cell growth and differentiation. Studies by Yoshitaka Nagai and his colleagues at the University of Tokyo indicate that the addition of another glycosphingolipid called GQ_{1b} to tumorous embryonic nerve cells induces the formation of mature nerve cells and many other changes. More recently Masaki Saito, Hisao Nojiri and their colleagues at Jichi University in Japan observed that when certain mouse leukemia cells are incubated with GM_3, they differentiate into macrophage cells with a higher content of GM_3.

In addition to their role in the regulation of proteins glycosphingolipids carry out certain cellular functions in their own right. Their most important function is to encode differences at the cellular level among species, among individuals from the same species and even among cells in a given individual. The blood-group antigens, for example, show how glycosphingolipids can vary among individuals in the human population.

There has been a great deal of interest in recent years in the changes cell-specific glycosphingolipid markers undergo in normal development as well as in cancerous growth. For example, Davor Solter and Barbara B. Knowles of the Wistar Institute of Anatomy and Biology in Philadelphia established the existence of an antigen called *SSEA-1*. (The acronym stands for stage-specific embryonic antigen.) *SSEA-1* does not react to its antibody in the fertilized egg, but it can be detected sometime between the third and the fifth cell division, or in other words in embryos made up of from eight to 32 cells. The cells of the embryo then undergo a process called compaction: they begin to adhere strongly to one another in order to maximize their intercellular contacts. Once the process is completed, the concentration of *SSEA-1* declines rapidly.

The chemical structure of *SSEA-1* was identified as a carbohydrate chain called Le^x on a glycosphingolipid or a glycoprotein by Reiji Kannagi, Steven B. Levery and Edward Nudelman of my laboratory and by Ten Feizi and her colleagues at the Medical Research Council's (MRC) Clinical Research Centre in Harrow, England. Bruce Fenderson and Uri Zehavi of my laboratory and Susan J. Kimber and her associates at the MRC Laboratory for Molecular Biology in Carshalton, England, have found that Le^x or its conjugate structure acts to inhibit compaction in the 16-to-32-cell stage of the embryo. Presumably the Le^x structure on a cell somehow senses the presence of other cells nearby. It then interferes with the complex process of cell adhesion, which could be based on a specific affinity between cell-surface Le^x and its receptor. Compaction can take place in a timely manner with the decline of Le^x after the 32-cell stage.

It now appears that many bacterial toxins and viral or microbial organisms can exploit the capacity of glycosphingolipids to mediate interactions



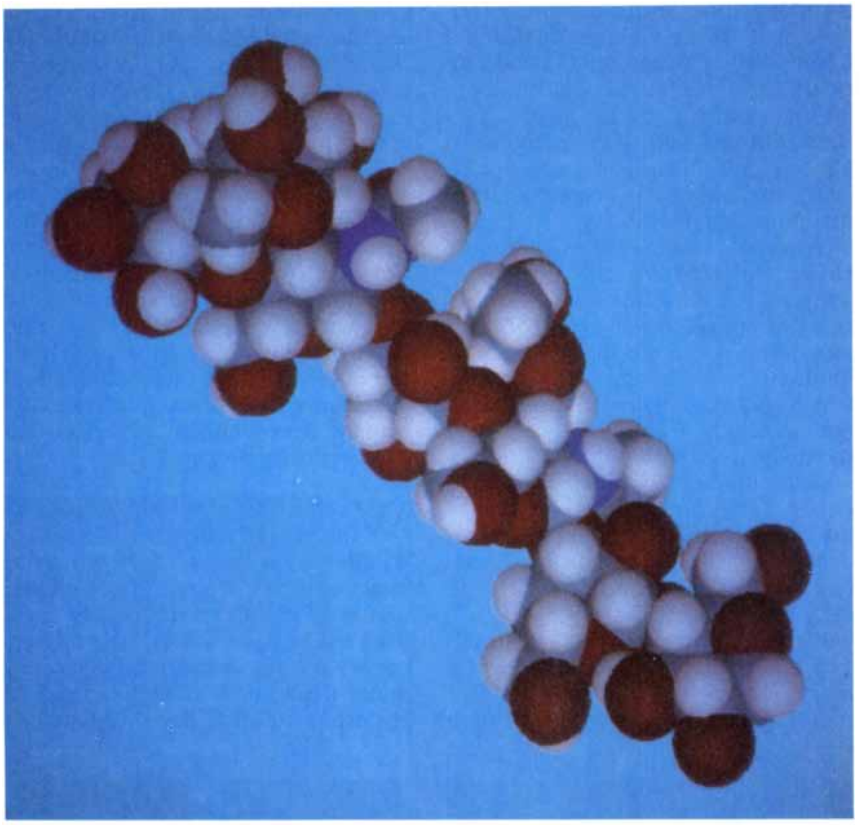
CELL-TO-CELL RECOGNITION is shown as it is thought to take place between two animal cells and between a bacterium and a host cell. Membrane proteins on each animal cell (*left*) may recognize the structure of an *SSEA-1* antigen displayed on the surface of the other cell. Such recognition is typical in embryos made up of from eight to 32 cells before implantation. Fimbriae proteins on a pathogenic *Escherichia coli* bacterium (*right*) can recognize the internal sugars of globo-series glycosphingolipids on a host cell and exploit them as infection sites. The sugars of the glycosphingolipids are color-coded according to the key in the illustration on page 47.

of the cell with its environment. Some 30 years ago William E. van Heyningen of the University of Oxford reported the first evidence that gangliosides interact with tetanus toxin. Subsequent evidence suggests the ganglioside GM_1 is the specific receptor for tetanus toxin, and it is also the receptor for cholera toxin. Cholera toxin acts not by invading the cell but by causing the membrane to lose its impermeability to water and to water-soluble electrolytes in the cell. Since the work of van Heyningen there has been a great deal of interest in determining whether or not other glycosphingolipids are receptors for other toxins and bioactive factors. Investigators have now shown that gangliosides interact—although not necessarily as receptors—with many biologically important substances, including botulinum toxin, interferon, interleukin, serotonin, hormones and Sendai virus.

Several groups of investigators have recently found evidence that specific glycosphingolipids on the surface of a host cell interact with the proteins of viral and bacterial parasites. Workers in Finland and Sweden discovered that certain strains of *Escherichia coli* bacteria, which cause an infection of the urinary tract and kidneys called pyelonephritis, display fine fibrillar structures called fimbriae on their surfaces. The fimbriae proteins interact with a glycosphingolipid called globoside and its chemical precursors, and the interaction is probably the first step in the infection of the host cell.

In a recent study Karl-Anders Karlsson and his colleagues at the University of Göteborg labeled *E. coli* bacteria causing urinary infections with a radioactive marker. The investigators then observed the bacterial interactions with glycosphingolipids separated on a thin layer of silica gel. Of 32 glycosphingolipids tested the only kinds that became attached to the bacteria were the ones having a particular chain of three sugars linked to the ceramide component of the molecule. Arbitrary chains of sugars attached to the initial three-sugar chain had no effect on the bacterial attachment. Thus the surface proteins of the bacteria have apparently learned to recognize short, internal sequences of sugars that could be displayed by many kinds of glycosphingolipid on the host cell.

Karlsson's techniques have also determined the binding sites for certain kinds of bacteria that make up the normal "intestinal flora" of the digestive tract. Such bacteria are harmless, and they benefit their host by inhibiting the growth of pathogenic bacteria. For example, the ordinary intestinal bacterium called propionibacterium has a surface protein that binds to a very common glycosphingolipid called lactosylceramide. The carbohydrate structure of lactosylceramide is a chemical precursor of most known glycosphingolipids. Indeed, the two sugars attached to the ceramide component of lactosylceramide are identical with the first two sugars in the three-sugar chain recognized by the *E. coli* bacteria that cause urinary infections.



TUMOR-ASSOCIATED ANTIGEN *FH4* is represented by this computer-generated model. The shaded area in the map at the left is a hydrophobic region of the molecule that may interact with antibody. *FH4* antigen has been identified as the glycosphingolipid dimeric Le^x, and it is strongly expressed in human cancers derived from gastrointestinal cells. The model was constructed by Levery, Stenkamp and Watenpaugh.

Whereas toxins and viral or bacterial infections may take advantage of existing glycosphingolipids on a healthy cell, cancerous growth is clearly associated with altered glycosphingolipids. Without doubt such change expresses some underlying abnormality. Whatever its cause, a disruption of the glycosphingolipid populations could interfere with an essential part of the cell's resources for maintaining an orderly social life. It would not be surprising to find that glycosphingolipid alteration is a major cause of the chaotic and undisciplined social interactions characteristic of cancer cells.

The kind of change brought about in the glycosphingolipids depends on the host cell and the specific cancer-producing agent. Some tumor cells accumulate simple glycosphingolipids because the synthesis of the more complex glycosphingolipids to which the simple ones give rise is somehow blocked. Other tumor cells, particularly human epithelial cancer cells, synthesize unusual carbohydrate chains and thereby accumulate novel glycosphingolipids called neoglycolipids.

Both precursor glycosphingolipids and neoglycolipids were originally identified as tumor-associated antigens in classical studies with rabbit antibodies. The knowledge that tumor cells generate variant glycosphingolipids could not be fully exploited, however, until the technique for grow-

ing monoclonal antibodies was developed in 1976 by George Köhler and Cesar Milstein of the MRC Laboratory in Cambridge. Monoclonal antibodies share a highly specific affinity for a single antigen, and so the presence of a given antigen can quickly be detected by attaching a labeling molecule to the monoclonal antibody.

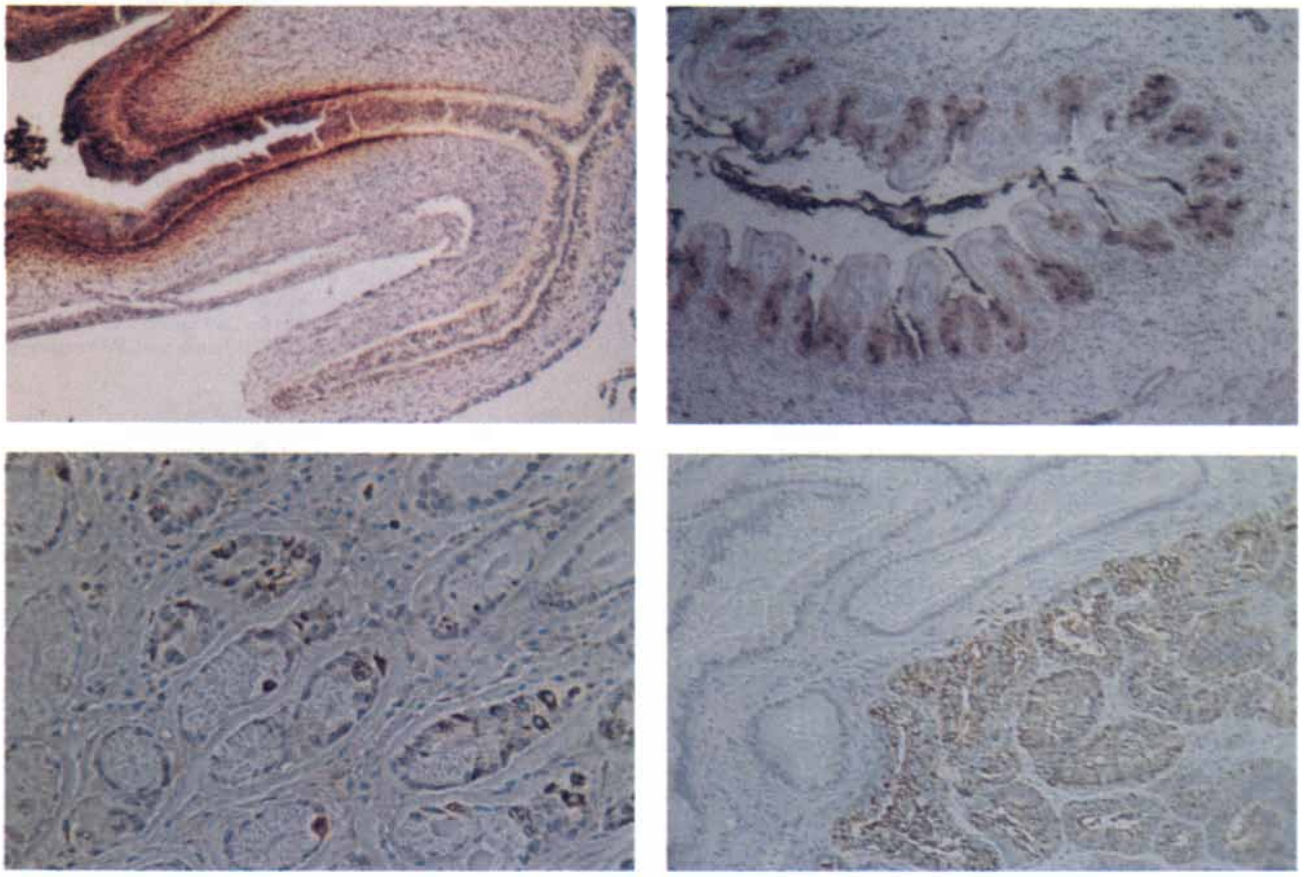
To raise large numbers of monoclonal antibodies specific for tumor antigens one must first inject the antigens into mice. Spleen cells from the immunized mice are then fused with mouse tumor cells. The resulting cells, which are called hybridomas, combine the capacity for unlimited growth of the tumor cells with the ability to generate antibodies from the spleen cells. In all more than 10 kinds of tumor-associated glycosphingolipid antigens defined by specific monoclonal antibodies have been identified and chemically characterized.

Because different hybridoma cells give rise to different antibodies, they must somehow be screened. Most tumor immunologists adopt a "shotgun," or random, approach: they inject undetermined antigens from tumor tissues into the mice and then screen the resulting hybridomas for their ability to react with the tumor-associated antigens. In our laboratory we adopt a less conventional method whereby the screening is done, in effect, at the outset. We first obtain and chemically characterize a glycosphingolipid associated with a specific kind of cancer. The glycosphingolipid is then coated on bacteria, and the bacteria are injected into the mice. In screening we can therefore seek the one kind of hybridoma that reacts preferentially with the glycosphingolipid antigen that was originally prepared.

We have applied our procedure to obtain monoclonal antibodies directed against a variety of novel tumor-associated structures. In their studies of the developmentally regulated antigens Le^x and Le^y, Nudelman, Levery and Kannagi found that many tumor tissues hold a large quantity of glycosphingolipids closely related to Le^x and Le^y. The novel antigens include the dimeric and trimeric forms of Le^x and forms to which sialic acid is attached. Yasuo Fukushi and Kazuo Abe of my laboratory then successfully raised monoclonal antibodies that distinguish among several different kinds of Le^x.

One such antibody, the so-called *FH4* antibody, reacts specifically with the tumor-associated antigen dimeric Le^x but not with the simple form of the compound. Fukushi then tested the capacity of labeled *FH4* antibody to bind to the healthy tissues of embryos and thus stain them. The staining indicated the strong expression of dimeric Le^x in the gastric epithelial cells of 35- to-45-day-old embryos, but it was limited to the deep crypt area of gastric and intestinal epithelia in 100-day-old fetuses. It practically disappears in the cells of the newborn and the adult [*see illustration below*].

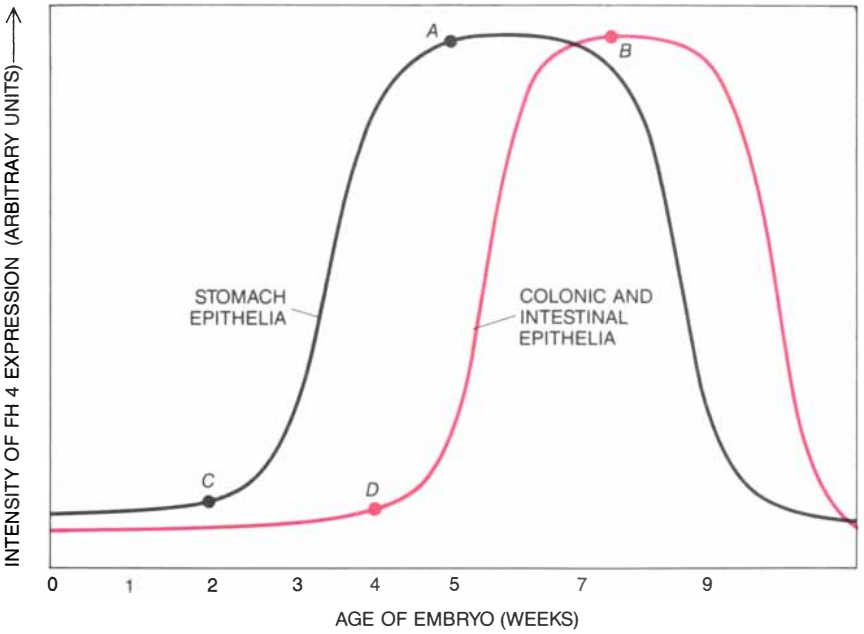
Cell-surface antigens that exhibit such behavior are called oncofetal antigens: they are found in the same relative abundances on cancer cells as



EXPRESSION OF *FH4* ANTIGEN is shown in human development and in a gastrointestinal tumor. The dark brown stains in the photomicrographs are caused by the binding of *FH4* monoclonal antibody to the antigen. The stains at the upper left indicate strong expression of the antigen in the stomach epithelial cells of an embryo slightly more than five weeks (38 days) old. The photomicrograph at the upper right shows the stomach epithelial cells of a 120-day-old fetus; only the deep folds of the stomach lining are strongly stained. In adult stomach epithelia almost none of the tissue is stained (*lower left*). At the lower right the tumor cells of a stomach cancer are stained but normal cells are not. Antigens expressed on tumor cells and in fetal development are called oncofetal antigens.

they are at brief stages in the development of certain kinds of embryonic cells. Several oncofetal antigens are now known. Another example is a sialic acid derivative of dimeric Le^x that is defined antigenically by the antibody *FH6*. It strongly stains fetal tissue, does not stain adult gastrointestinal tissue and strongly and specifically stains gastrointestinal tumors. Such findings provide particularly strong confirmation for the frequent observation that cancer cells resemble the relatively undifferentiated cells found in early embryonic development.

Another important blood-group-related oncofetal antigen is the sialic acid derivative of a glycosphingolipid called Le^a antigen. A monoclonal antibody directed against the Le^a antigen was established by Hilary Koprowski and his colleagues at the Wistar Institute, and the antigen was later isolated and chemically characterized by Victor Ginsburg, John L. Magnani and their colleagues at the National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases. Because monoclonal antibodies bind only to the carbohydrate chain of a glycosphingolipid, the antibodies can also recognize such chains on serum glycoproteins that are not attached to cells. Antibodies to the sialic acid derivatives of the glycosphingolipids Le^a and Le^x, for example, react with the serums of cancer patients. Such circulating antigens can readily be assayed, and so their presence is quite useful in the diagnosis of cancer. The study of such antigens is a major topic of clinical investigations.



INTENSITY OF EXPRESSION of the *FH4* antigen is plotted for human fetal development. In the epithelial cells of the stomach *FH4* expression reaches its maximum intensity when the embryo is between five and seven weeks old. In the epithelia of the colon and the intestine the maximum intensity is observed from ages seven to nine weeks. The *FH4* antigen is expressed in differentiated cancers of the stomach or the colon much as it is in the fetal epithelia (*A, B*). In undifferentiated cancers of the stomach or the colon *FH4* is not expressed; such cancer cells resemble the cells of the embryo in early stages of life (*C, D*).

It remains true, nonetheless, that the cell-bound glycosphingolipid antigens may hold the best promise for the treatment of cancer. For example, ordinary Le^x and its analogues dimeric and trimeric Le^x and Le^y are attached to the tumor cell and not released into serum. In principle, once the antibodies specific for tumor-associated antigens are identified, they can be grown and administered to cancer patients. Cancer cells displaying altered populations of tumor-associated antigens might then be preferentially sought out by the antibodies and destroyed through complex mechanisms triggered by the antibodies and certain kinds of killer cells. The gradual disappearance of the malignant tumor in the woman with the rare *p* blood group was inadvertently brought about in just this way.

Recent immunological studies have shown that many monoclonal antibodies directed against tumor-associated antigens are indeed directed against glycosphingolipids. The results of at least two clinical studies appear to justify some of the most optimistic expectations. In one study a group of investigators led by Lloyd J. Old and Kenneth O. Lloyd of the Sloan-Kettering Institute for Cancer Research in New York established an antibody with a high affinity for the glycosphingolipid GD_3. Alan N. Houghton and his associates at Sloan-Kettering administered the antibody to 12 patients with melanoma, a virulent form of cancer that leads to dark tumors on the skin, which eventually metastasize and invade various organs. In three of the 12 patients there was a clear regression of the melanoma, and four other cases showed a mixed response.

In the second study Ronald B. Herberman of the National Cancer Institute administered anti-GD_3 antibody to another 12 patients with melanoma. The antibody was prepared and its antigenic specificity was established by Ralph A. Reisfeld and his co-workers at the Scripps Clinic and Research Foundation. Again the antibody was effective in three out of the 12 patients, although other kinds of antimelanoma antibodies showed no effect. Furthermore, although the glycosphingolipid GD_3 is present in moderate quantities in normal tissues of the kidney and the intestine and in high quantities in the retina, no patients developed disorders of these organs from the antibody treatment. The only adverse effects were inflammatory skin reactions around the melanoma. Such findings are highly encouraging, and they suggest the glycosphingolipid antigens in tumor cells are more accessible to an attack by antibodies than the same antigens in normal tissues are.

There are several proposed strategies for removing or killing tumor cells that might exploit the specificity of monoclonal antibodies. Radioactive isotopes attached to the antibody could bind to the tumor cells and thereby irradiate them or indicate their position in the body with precision. Additional radiative treatment or the surgical removal of the tumor would follow. Drugs bound to the antibodies might also kill the tumor cells selectively.

Furthermore, lest one forget, antibodies constitute the natural and highly effective means the body employs to isolate and kill diseased cells. A monoclonal antibody, like any other antibody, not only binds to a cell that displays an antigen but also marks it for destruction by the immune system of the host. Many investigators are now working to turn the fortunate clinical accident of 35 years ago into an effective and routine treatment for many forms of cancer.

The Climate of Mars

It started out much like the earth's early climate, but it evolved differently. Once warm enough to support flowing water, Mars is now so cold that carbon dioxide freezes at the poles every winter

by Robert M. Haberle

Fall has come to the northern hemisphere of Mars. In the mid-latitudes the mean temperature is dropping past −70 degrees Celsius; at the north pole it has probably already reached −123 degrees, cold enough to freeze carbon dioxide, the principal constituent of the thin Martian atmosphere. A cap of dry ice is now forming on the pole. Before winter ends ice will reach the 50th parallel. Meanwhile in the southern hemisphere winter has ended, and carbon dioxide is evaporating from the south pole. Along the edge of the retreating polar cap a sharp temperature contrast between ice and soil warmed by the spring sun is giving rise to strong winds. During the short but hot southern summer, as Mars makes its closest approach to the sun, the winds will lift dust off the surface in great swirling storms. Ultimately dust may envelop the entire planet.

The climatic conditions on Mars today are known in such remarkable detail thanks to the spacecraft missions of the 1960's and 1970's, notably the Viking mission. Although it is harsh and inhospitable by terrestrial standards, the Martian climate is in some ways similar to that of the earth. Both planets, for example, have a global wind system marked by trade winds in the tropics and cyclonic storms in the mid-latitudes. Yet the differences are even more striking. Whereas the dominant gases on the earth are nitrogen and oxygen, the atmosphere of Mars is 95.3 percent carbon dioxide. Moreover, the Martian atmosphere is by far the less massive of the two. The pressure it exerts on the planet's surface is on the average only 6.1 millibars, compared with an average of 1,013 millibars on the earth. As investigators await the arrival of the next spacecraft at Mars—the *Mars Observer,* scheduled for launch in 1990—they continue to study the Viking data. In the process they are achieving a better understanding of how the Martian climate has evolved, becoming similar to the earth's climate in some respects and radically different in others.

Early in the 4.6-billion-year history of the solar system the climates of Mars and the earth were probably much more similar than they are now. Indeed, it seems likely the two planets started out with atmospheres whose composition was about the same. On both planets the primitive atmosphere is thought to have consisted of gases released from rocks, either by volcanic outgassing, in which gases were "cooked" out of magma inside the planet and discharged at volcanoes and other vents, or by impact degassing, in which gases were released when rocks collided with the accreting planet. The rocks themselves condensed from gas and dust in the solar nebula; their composition was determined by the temperature at which they formed, which depends on the planet's distance from the sun. Mars and the earth condensed at distances similar enough for the composition of their rocks, and hence of their primitive atmospheres, to be fundamentally similar.

There is good reason to believe the primitive atmospheres consisted mostly of water vapor, carbon dioxide and nitrogen. On theoretical grounds these gases are expected to be released from typical hot magmas, and observations of the emissions from terrestrial volcanoes support this view. The atmospheres of the earth and Mars are so different today because the initial volatile substances have had different fates on the two planets.

Most of the earth's water is now in the oceans; the amount outgassed was far more than the atmosphere can hold. The presence of oceans in turn drives an active hydrological cycle that removes carbon dioxide from the atmosphere. Water evaporates from the oceans, and when it rains out of the atmosphere, it takes carbon dioxide with it. As the water weathers continental rocks, the dissolved carbon dioxide reacts with silicate minerals to form bicarbonate ions, which are carried off by rivers and ultimately deposited on the ocean floor as calcium carbonate (limestone). Most of the outgassed carbon dioxide is now stored in this rocky reservoir; the amount remaining in the atmosphere is insignificant by comparison. In contrast, nitrogen has become the dominant constituent of the earth's atmosphere (78.1 percent of the total volume) because it is chemically inert and relatively insoluble in water. Oxygen, which makes up 20.9 percent of the atmosphere, owes its abundance entirely to the presence of photosynthesizing plants.

On Mars too more water was vented than its atmosphere could hold. It has been estimated that the total outgassed water would cover the planet with an ocean from tens to hundreds of meters deep. Today Mars is much too cold to support oceans or any form of stable liquid water. The most obvious reason is that it is half again as far from the sun as the earth is and receives less than half as much solar energy. Yet distance from the sun cannot be the only factor, for it seems Mars was once warm enough to support flowing water on its surface.

Images of the planet's surface made by *Mariner 9* and the Viking orbiters revealed many channels whose appearance suggests they were carved by

SOUTH POLE of Mars is capped with carbon dioxide ice. The cap may be permanent; the photograph was made by the *Viking 2* orbiter in 1977 during the southern summer. In contrast, all the carbon dioxide ice on the north pole dissipates every summer, exposing an underlying cap of water ice. Carbon dioxide accounts for about 95 percent of the thin Martian atmosphere, and about 20 percent of it freezes at the south pole every winter. The image was prepared by the U.S. Geological Survey in Flagstaff, Ariz.

running water. Not all the channels are evidence of a warmer climate; some could have been formed by water flowing under an insulating layer of surface ice in a climate not substantially warmer than the one today. Many of the channels, however, resemble surface runoff networks, with tributaries converging into a single valley. Assuming these channels were carved by water flowing at the surface, then the global average temperature at the time must have been above 0 degrees C. (Today it is about −53 degrees.) From the number of meteorite craters superposed on the runoff channels their age has been estimated at roughly four billion years. Hence it appears Mars was comparatively warm for the first half-billion years or so of its history.

How could the surface temperature on Mars have been kept above the frost point of water? The likeliest mechanism is the familiar greenhouse effect. The gases in the Martian atmosphere are nearly transparent to most of the sun's radiation, and so the atmosphere is not heated directly by the sun. Instead, like the earth's atmosphere, it is heated from below, by energy reradiated from the surface at longer, infrared wavelengths. Because the gases do absorb at these wavelengths, they trap the heat. Water vapor and carbon dioxide are particularly effective greenhouse gases.

The greenhouse effect can be measured by comparing the temperature at which a planet radiates the energy it absorbs from the sun with its observed surface temperature. The earth's effective radiating temperature is −18 degrees C., and its global mean surface temperature is 15 degrees. The atmospheric greenhouse thus warms the surface by 33 degrees. Mars, on the other hand, radiates at −56 degrees, just three degrees below its surface temperature. Its greenhouse effect is modest now because its atmosphere is thin.

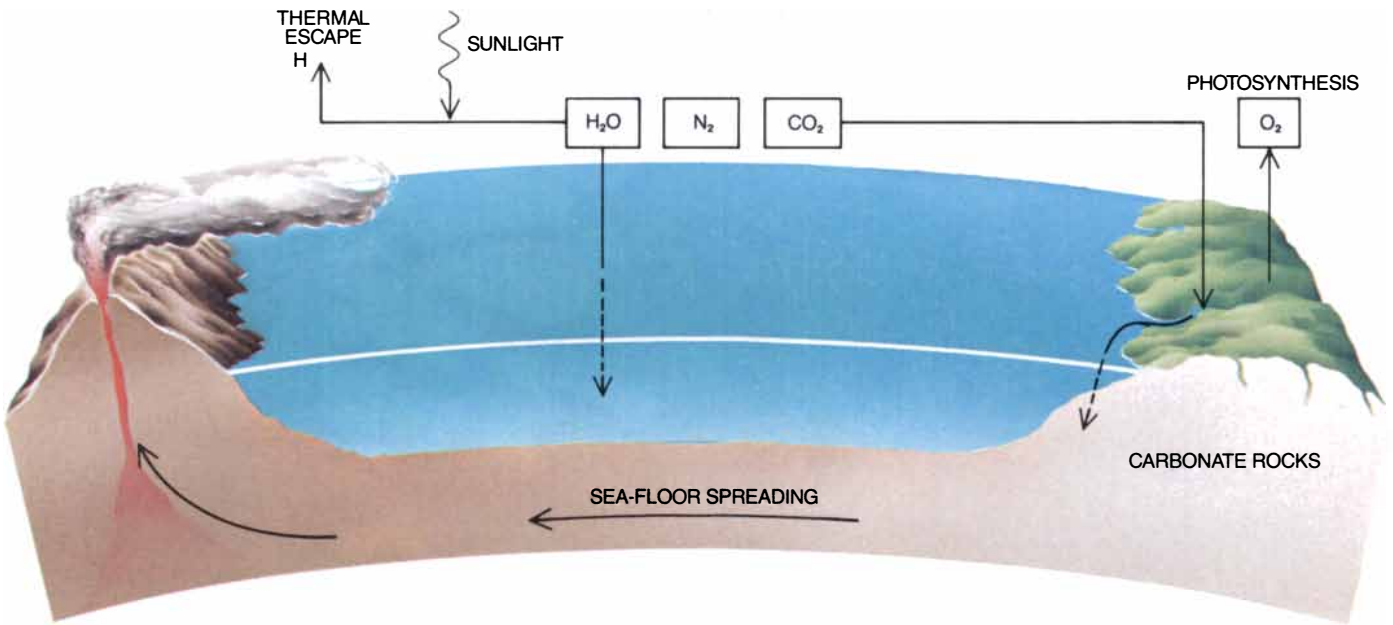
For the temperature on Mars to have been above 0 degrees four billion years ago there must have been more greenhouse gases in the atmosphere. The obvious candidates are carbon dioxide and water; theoretical models indicate in particular that at least 1,000 millibars of carbon dioxide would have been necessary. Assuming that much carbon dioxide was indeed outgassed, where has it all gone?

Large amounts may be adsorbed in the Martian regolith, a surface layer of fine-grained soil created by meteoritic bombardment. In the 1970's Fraser P. Fanale and William A. Cannon of the California Institute of Technology's Jet Propulsion Laboratory showed that carbon dioxide adheres to soil grains like those believed to make up the regolith. Carbon dioxide is also stored as ice in the polar caps, but most of this material sublimes (passes from the solid to the vapor state) every summer. Although there may be a permanent residual cap at the south pole—the Viking orbiters photographed ice there in 1977, during the southern summer—the amount stored in this reservoir appears to be small compared with the amount outgassed.

Another possibility is that much of the carbon dioxide on Mars is in the same type of reservoir that holds the bulk of the earth's carbon dioxide: carbonate rocks. As long as Mars was warm enough to support liquid water, the water drew carbon dioxide out of the atmosphere and converted it into carbonates much as the oceans do on the earth. It is not clear whether there has ever been extensive precipitation on Mars, but carbonate formation can also take place without rain, since carbon dioxide dissolves in open bodies of water as well.

There is a crucial difference between the earth and Mars, however, in the ultimate fate of carbon dioxide. On the earth some of the gas is constantly recycled into the atmosphere as a result of plate tectonics. Where two of the earth's moving crustal plates collide one of them dives into the mantle, carrying carbonate sediments with it. The intense heat and pressure in the mantle free carbon dioxide from the sediments. The gas then escapes through volcanoes into the atmosphere.

Less carbon dioxide has been recycled on Mars because Mars is less volcanically active than the earth. In particular it shows no evidence of plate



PRIMITIVE ATMOSPHERES of Mars and the earth were probably similar, but they evolved differently. On both planets the primitive atmosphere is thought to have been predominantly carbon dioxide, water vapor and nitrogen released by volcanic outgassing and impact degassing. On the earth (*left*) the nitrogen has remained in the atmosphere. Most of the water vapor has condensed into oceans. Most of the carbon dioxide has been removed from the atmosphere through precipitation and weathering and then stored in carbonate rocks. Some carbon dioxide, however, is constantly being recycled into the atmosphere through volcanoes. On Mars (*right*) the rate of

tectonics; the entire crust appears to consist of a single plate. The reason is that Mars is only a little more than half the size of the earth. In general a planet's size determines the strength of the internal heat sources—mainly radioactive decay processes and gravitational energy released during accretion—that drive tectonic and volcanic activity.

If there were no volcanic recycling on the earth, weathering processes would convert the small amount of carbon dioxide in the atmosphere into carbonate within several thousand years. A 1,000-millibar carbon dioxide atmosphere on Mars could have survived longer: for 10 to 100 million years without recycling, according to calculations done by James B. Pollack of the National Aeronautics and Space Administration's Ames Research Center. Yet even 100 million years is not long enough to explain the formation of the valley systems by liquid water. Consequently Pollack has suggested that early in its history Mars may have had enough energy left over from accretion to drive global-scale volcanism. The carbon dioxide thus recycled, he believes, could have sustained a strong greenhouse effect for the half-billion or so years during which the valleys were carved.

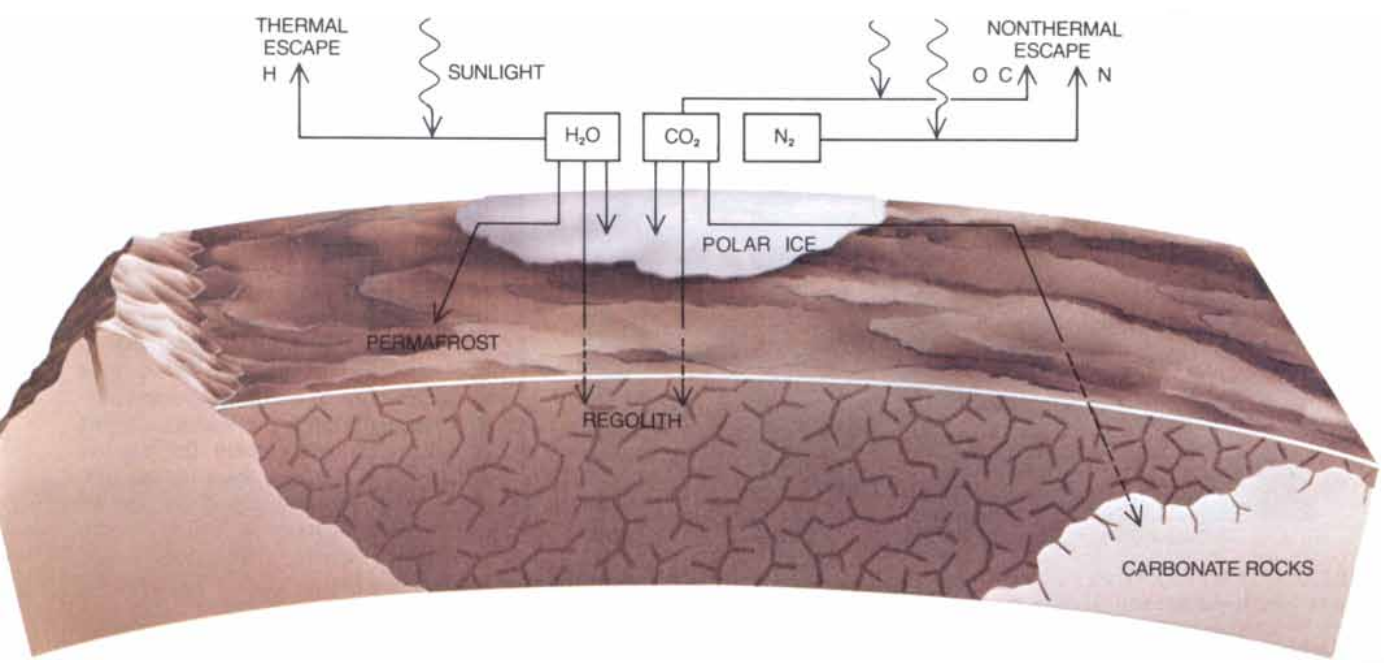
As volcanic activity abated, the atmospheric carbon dioxide level and hence the greenhouse effect declined. Eventually the planet cooled to the point where liquid water was no longer stable. At that point the atmospheric surface pressure was probably well above the level observed today. The surface pressure on Mars is so low now, according to Ralph A. Kahn of Washington University in St. Louis, because carbon dioxide continued to be removed from the atmosphere and stored as carbonates by transitory pockets of liquid water. Such pockets, he argues, could have existed long after the global mean temperature had dropped below the freezing point; specifically, they could form as long as the surface pressure was high enough to limit evaporation. Through the action of transitory water pockets the pressure was gradually reduced to its current value of 6.1 millibars.

Ironically, then, the presence of liquid water may ultimately have led to its own destruction. Some of the water may be stored in polar ice deposits or in widely dispersed permafrost. The regolith may also bind water chemically or physically, as it is thought to bind carbon dioxide.

If this scenario is correct, the small size of Mars has had at least as much influence on its climate as has its distance from the sun. Moreover, the size of Mars has determined the fate not only of water and carbon dioxide but also of nitrogen, which is relatively scarce in the Martian atmosphere. The lower level of volcanic activity on Mars meant that less nitrogen was outgassed than on the earth. In addition, because Mars is smaller, its gravity is weaker, and so nitrogen can escape from the planet. Although the nitrogen does not have enough thermal energy to escape, it can acquire the necessary energy by a process known as dissociative recombination. Electrons are knocked out of nitrogen molecules by ultraviolet light in the upper atmosphere; when the electrons and the ionized nitrogen recombine, there is more than enough energy to split the molecules into their constituent atoms. The excess gives the atoms enough kinetic energy to escape.

An estimate of the amount of nitrogen outgassed on Mars can be obtained from the relative abundances of nitrogen 15 and nitrogen 14. Because nitrogen 15 is heavier and thus less likely to escape, its concentration has gradually increased in relation to that of nitrogen 14. Models of this evolution begin with the assumption that Mars and the earth initially had the same isotope ratio, namely the ratio that is still observed on the earth, from which even nitrogen 14 cannot escape. The models indicate that as much as 30 millibars of nitrogen was vented from Mars, nearly 200 times the amount now in the atmosphere. The present concentration is in turn just a minute fraction of the concentration in the earth's atmosphere.

Although the earth and Mars have evolved atmospheres of vastly differ-



volcanism is now so low that virtually no carbon dioxide is recycled. As a result weathering processes driven by liquid water have destroyed the thick carbon dioxide atmosphere that kept the planet warm enough (through the greenhouse effect) for liquid water to exist. Most of the carbon dioxide is probably stored in carbonates or adsorbed in the fine soil of the regolith; a small amount may be in a permanent ice cap at the south pole. The north pole has a permanent water-ice cap, but most of the water on Mars must be in the regolith or in dispersed permafrost. Nitrogen has escaped Mars's weak gravity with the help of energy from photochemical reactions.

ent composition, the two planets have some climatic features in common. For example, they have similar global circulation patterns. There are several reasons for this.

First, since both the Martian and the terrestrial atmospheres are nearly transparent to sunlight, they are both heated from below. Second, on both planets the tropics receive significantly more solar energy than the poles; the resulting temperature gradient produces a pressure gradient that drives a meridional (north-south) circulation. Third, the two planets rotate rapidly and at about the same rate. (A Martian day is about 40 minutes longer than an earth day.) The rotation rate determines the magnitude of the Coriolis force, an apparent force that deflects moving air parcels.

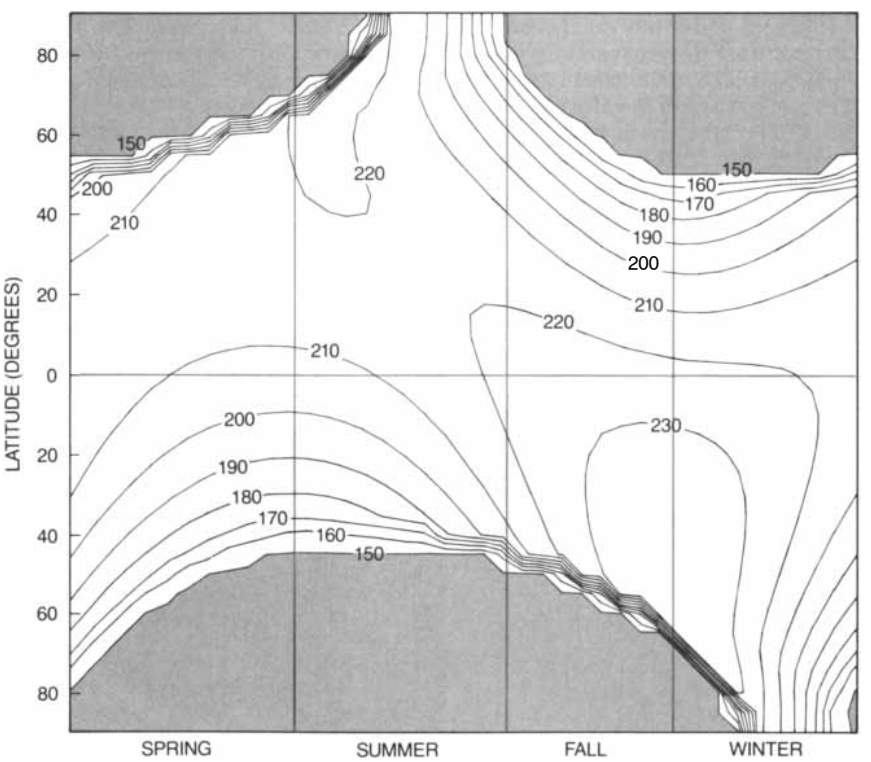
Theoretical modeling studies carried out by Conway B. Leovy of the University of Washington, Pollack and me have revealed the broad features of the Martian global circulation. As on the earth, the meridional circulation takes the form of two Hadley cells, one in each hemisphere. The cells share a common rising branch near the equator, where heating from the sun is strongest. Air rises there, moves toward the poles at upper levels and then cools and sinks in the subtropics (at latitudes of about 25 degrees). Finally it returns to the tropics near the surface. Because the Coriolis force deflects moving air to the right in the northern hemisphere and to the left in the southern hemisphere, the deflection of poleward-moving air in the Hadley cells gives rise to westerly winds aloft; at the surface the deflection of equatorward-moving air gives rise to easterly trade winds like those on the earth.

A different circulation regime, again much like the corresponding one on the earth, prevails at mid- and high latitudes. There winds are predominantly from the west both at the surface and aloft. A narrow band of intense winds called the jet stream blows in the upper atmosphere, at an altitude of perhaps between 30 and 40 kilometers. Near the surface storm systems consisting of high- and low-pressure centers migrate eastward, following the path of the overlying jet stream and drawing energy from it. The anticyclonic and cyclonic winds swirling around the pressure centers perform the function carried out at lower latitudes by the Hadley cells: they transport energy toward the poles.

Unlike storms on the earth, whose movements are notoriously difficult to forecast, the Martian weather systems are almost monotonously regular. At the Viking lander sites storms typically moved past every two to four days. The cause of the regularity is uncertain, but it may be related to the fact that carbon dioxide, the principal constituent of the Martian atmosphere, rèadily absorbs and emits thermal radiation. An extensively radiating atmosphere may quickly damp out the temperature perturbations caused by storm systems. In contrast, storms on the earth may be more irregular because nitrogen, the principal constituent of the earth's atmosphere, is radiatively inert.

The circulation on Mars changes with the seasons as it does on the earth, and for the same reason. A planet has seasons if its rotation axis is not perpendicular to its orbital plane; as the planet moves around the sun, the inclination of the axis leads to a change in the latitudinal distribution of incoming solar radiation and a corresponding change in the temperature distribution. The greater the angle of inclination (the obliquity) is, the more pronounced the seasonal changes are. At present this angle is roughly the same for the earth (23.5 degrees) and Mars (25.2 degrees), and so seasonal changes on the two planets are similar. Two circumstances, however, lead to a somewhat different pattern on Mars: it has no oceans, and its orbit is more elliptical than that of the earth.

One way the absence of oceans manifests itself on Mars is through seasonal changes in the configuration of the Hadley cells. The rising branch of the Hadley cells tends to be above the warmest point on the planet. Since the Martian soil, unlike the terrestrial oceans, responds quickly to changes in the amount of incoming solar energy (the insolation), the warmest point probably follows closely the annual path of the sun. At the solstices the rising branch may therefore be as much as 25 degrees of latitude away from the equator. The theoretical models suggest that the two Hadley cells are then replaced by a single cross-equatorial cell. This never happens on the earth; the joint rising branch of the Hadley cells follows the sun only partway on its spring and fall excursions away from the Equator, because the oceans respond only sluggishly to changes in temperature.



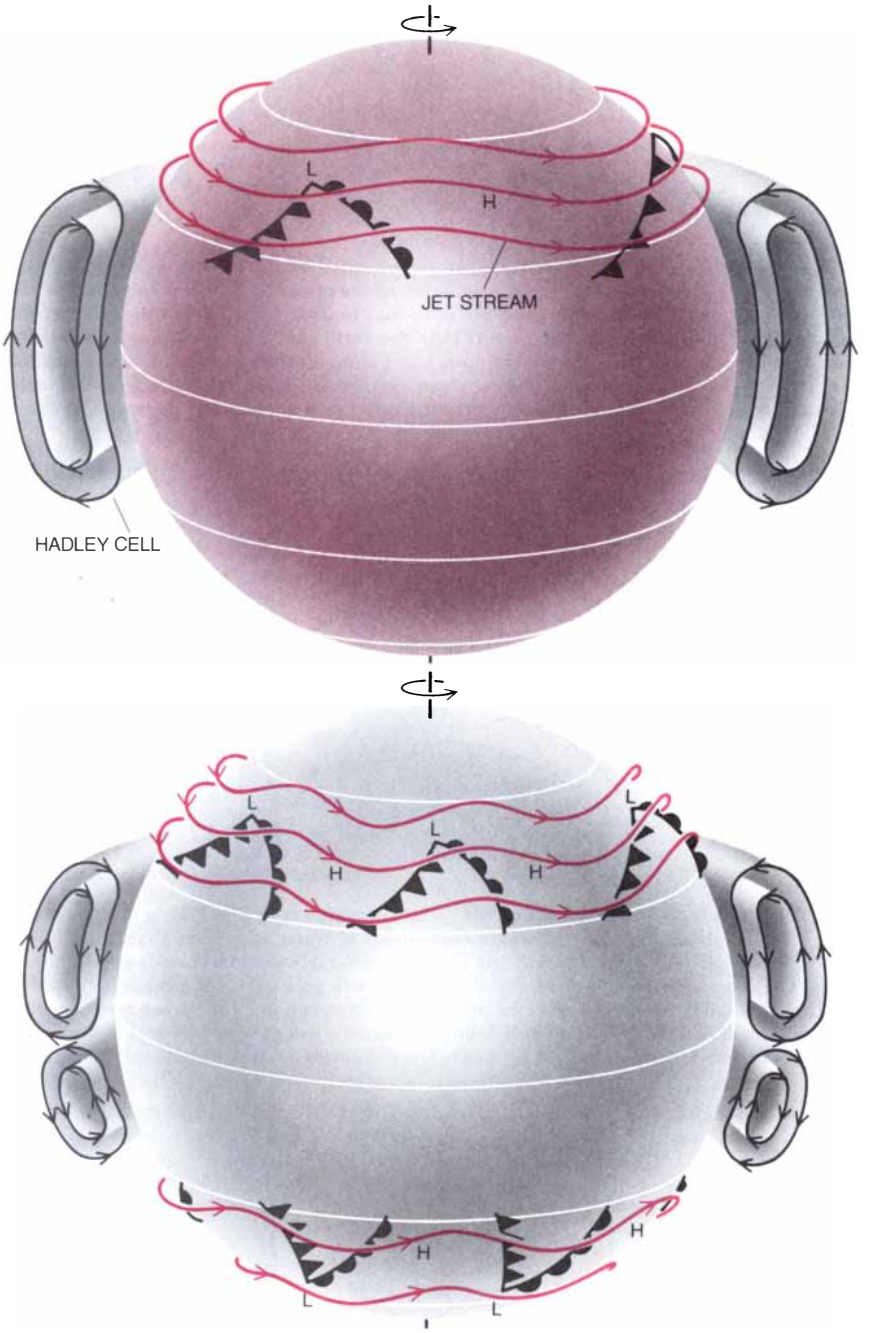
SURFACE TEMPERATURES on Mars follow a seasonal cycle, but the seasons are of unequal length and intensity because the Martian orbit is distinctly elliptical. (The seasons shown here and in subsequent illustrations are the northern ones.) The contours represent lines of equal temperature in degrees Kelvin (degrees Celsius above absolute zero, with 273 degrees Kelvin being equal to 0 degrees C.). Carbon dioxide freezes at about 150 degrees, and so the 150-degree contours mark the approximate boundaries of the polar ice caps (*gray*). The south cap grows larger than the north cap because fall and winter are longer and colder in the south. The highest temperatures occur in the southern subtropics in late spring (northern fall). As the ice caps retreat the latitudinal temperature gradient along their edges becomes steep. The map is based on a model developed by Hugh H. Kieffer of the U.S.G.S.

The absence of oceans also affects the circulation in the mid-latitudes. The strength of a jet stream depends on how much the temperature below it changes with latitude; the greater the change, the greater the pressure gradient aloft and the stronger the jet. On the earth the oceans keep the latitudinal temperature gradient from changing much with the seasons. Hence the jet streams blow all year round, although their intensity is greatest in the winter. Mars is different. In winter and spring the Martian jet stream is nearly four times stronger than the jet streams on the earth, because in those seasons the polar ice cap extends to the mid-latitudes. The temperature contrast between the carbon dioxide ice and the neighboring soil is strong, particularly in the spring, when the soil next to the ice quickly warms. In summer, however, the jet stream disappears, because by then the ice cap has receded and the soil throughout the hemisphere has roughly the same temperature. The same pattern is followed by the storm systems that draw their energy from the jet stream.

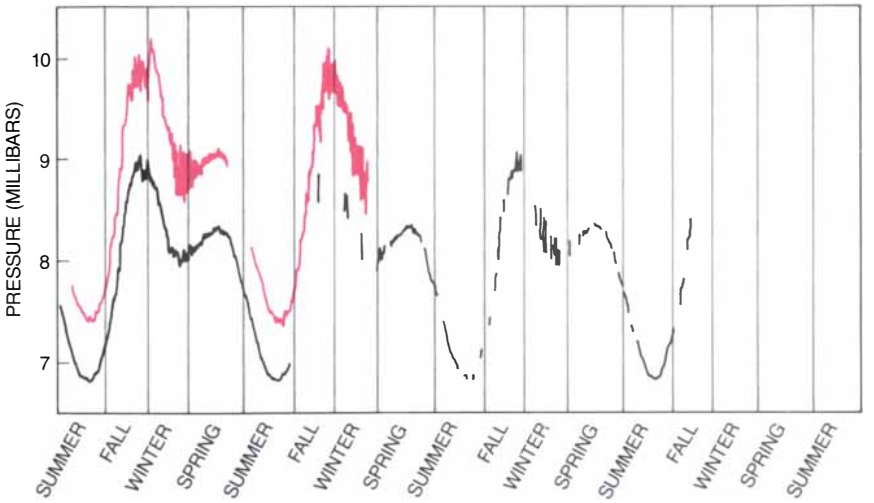
The second key difference between the earth and Mars as far as seasons are concerned is that the orbit of Mars is more eccentric, or elliptical, than the earth's orbit. As a result the Martian seasons are of unequal duration and intensity. (They are all longer than terrestrial seasons because the Martian year is 687 earth days long.) The planet makes its closest approach to the sun (perihelion) late in the southern spring, which is 52 earth days shorter than the northern spring, when the planet is at its greatest distance from the sun (aphelion). At perihelion Mars receives about 40 percent more solar radiation than it does at aphelion; the corresponding difference for the earth is only 3 percent. Compared with the northern seasons the southern spring and summer are short and hot, and the fall and winter are long and cold. The asymmetry has an important influence on three interacting seasonal cycles: the cycles of carbon dioxide, of water and of dust.

At high latitudes during winter the temperature on Mars becomes cold enough to freeze carbon dioxide. This condensation of a major atmospheric constituent is a uniquely Martian phenomenon. About 20 percent of the atmosphere is cycled in and out of the polar caps every year, causing a planetwide variation in surface pressure. In effect the atmosphere oscillates on a seasonal basis, piling up at the winter pole. Because fall and winter are longer in the southern hemisphere, the south polar cap grows larger than the north polar cap. At its greatest extent it reaches a latitude of 45 degrees or so, whereas the north cap reaches a latitude of 50 degrees.

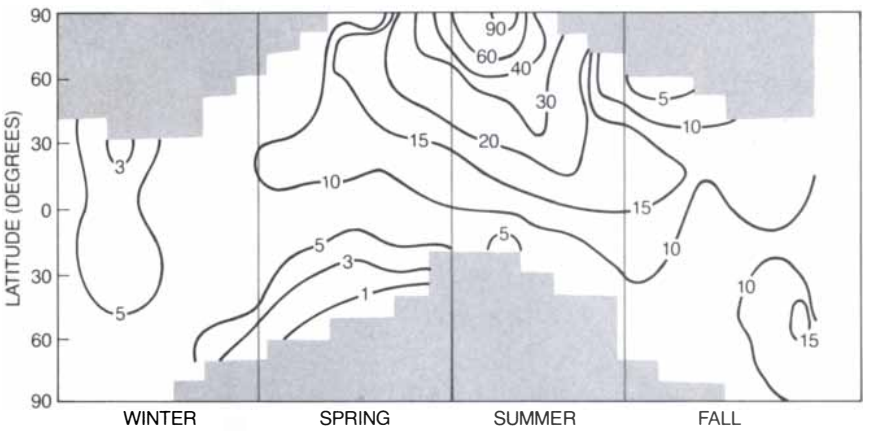
Not only does the carbon dioxide cycle determine the seasonal variations in surface pressure but also it influences the seasonal behavior of water in the Martian atmosphere. As spring gives way to summer in each hemisphere the polar caps retreat, but they never completely disappear. The composition of the residual caps was uncertain until the Viking orbiters flew over them in 1976 and 1977. The north residual cap was observed to be water ice, whereas the south residual cap was mostly carbon dioxide. From meas-



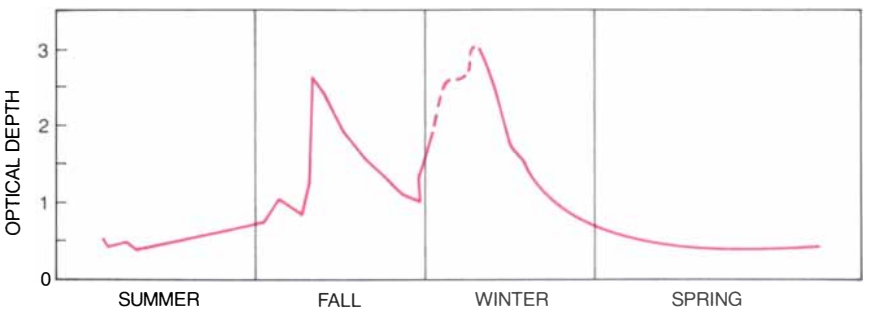
WIND SYSTEMS on Mars (*top*) are similar to those on the earth (*bottom*). On Mars, however, seasonal changes in the winds are more pronounced because the Martian soil responds more rapidly to changes in incoming solar energy than the earth's oceans do. The illustration depicts the global circulation during the northern winter. On the earth the rising branch of the tropical Hadley cells is displaced slightly south of the Equator. On Mars the rising branch moves some 25 degrees south of the equator; the two Hadley cells that exist during the fall give way to a single cross-equatorial cell. On both planets the circulation in the mid-latitudes is dominated by high-altitude, westerly jet streams that guide storm systems at the surface. On Mars, however, the latitudinal temperature gradient that drives the jets is strong only in the winter hemisphere; jets and storms vanish in the summer.



CARBON DIOXIDE CYCLE on Mars was documented by the Viking landers, which recorded seasonal fluctuations in the atmospheric surface pressure. The pressure tends to rise when carbon dioxide evaporates from the poles and to drop when carbon dioxide freezes, but since the south polar cap grows larger than the north cap, its fluctuations predominate. The pressure is highest in the southern spring and summer (northern fall and winter) and lowest in the southern winter. *Viking 2* (*color*) recorded higher pressures than *Viking 1* because it was at a lower altitude. The pressure data were analyzed by James E. Tillman.



WATER CYCLE on Mars depends on the carbon dioxide cycle. The Viking orbiters measured seasonal changes in the latitudinal distribution of water vapor; its abundance in the atmosphere is given in terms of the depth (in micrometers) it would have if it were in liquid form. The peak abundance was observed over the north pole during the summer, when all the carbon dioxide ice had evaporated off the pole, exposing the underlying water ice. No peak was observed over the south pole, which remained covered with carbon dioxide. In the gray areas no measurements could be made. The data were analyzed by Bruce M. Jakosky.



DUST CYCLE on Mars was charted by the *Viking 1* lander by measuring the degree to which sunlight was attenuated in passing through the atmosphere. The more dust in the atmosphere, the greater the atmosphere's opacity, or optical depth. The Martian atmosphere never clears completely. Opacity peaks during the fall and winter mark the onset of global dust storms. The broken segment represents a time when only a lower limit on the opacity was obtained. The data were analyzed by James B. Pollack and his colleagues.

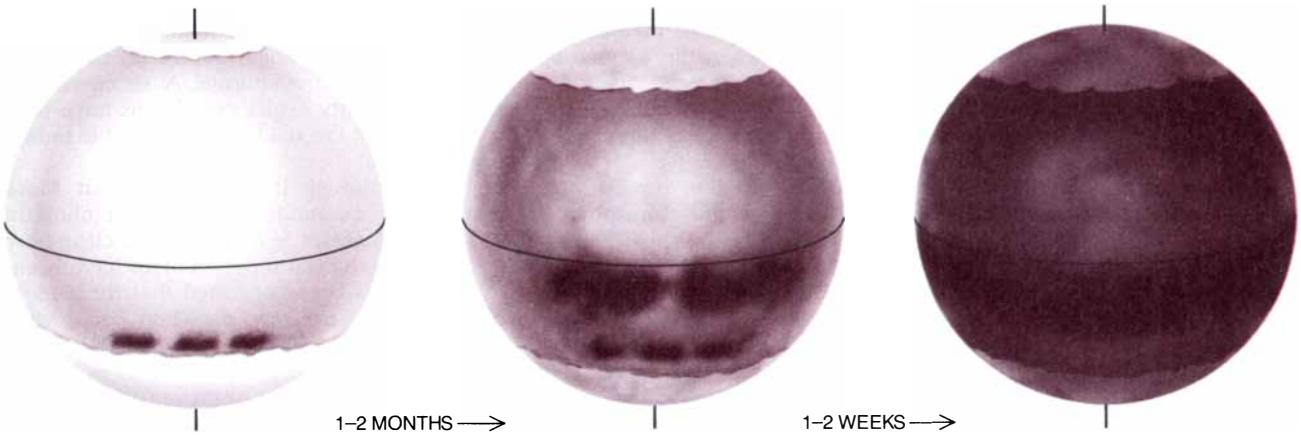
urements of reflected sunlight the Viking orbiters were able to map the areal distribution of water vapor in the atmosphere. The maximum abundance occurred during summer over the north polar region; no similar peak was observed over the south pole during summer.

The pattern has a simple interpretation. The permanent water-ice cap at the north pole becomes a source of atmospheric water vapor every summer when it is exposed by the evaporation of the carbon dioxide ice above it. Conversely, if carbon dioxide ice always survives summer at the south pole, any water vapor brought in contact with it would be frozen and removed from the atmosphere. In other words, the south polar cap would be a water-vapor sink that is growing at the expense of the north cap. In the long run the latter would not be permanent at all but would disappear.

It is not certain, however, that carbon dioxide ice survives summer after summer at the south pole. Telescopic observations of Mars in 1969, when it was close to the earth, found much more water vapor in the atmosphere during the southern summer than the Viking orbiters found during the same season in 1977. In 1969 the carbon dioxide cap may have sublimed off the south pole, exposing an underlying water cap like the one at the north pole. Whether or not the southern carbon dioxide cap is permanent, though, the fact that it seems to survive at least some summers, whereas the north cap does not, requires an explanation. The explanation probably has something to do with the third seasonal cycle on Mars: the dust cycle.

By measuring the attenuation of sunlight in the Martian atmosphere, the Viking landers were able to monitor the behavior of atmospheric dust over a full set of seasons. Evidently the atmosphere never clears completely. It is clearest during the northern spring and summer, but even then the sky photographed from the landers had the pinkish-yellow color of sunlight scattered off suspended particles. In the northern fall and winter sharp rises in opacity mark the arrival of global dust storms. Such storms are so intense that dust particles are lifted to heights in excess of 40 kilometers and surface features are totally obscured. Since there is little or no precipitation on Mars to clean out the atmosphere, the dust may remain suspended for weeks or even months.

Global dust storms arise particularly during the northern fall and winter (southern spring and summer), when Mars is near perihelion, because that is



GLOBAL DUST STORMS occur during the northern fall and winter (southern spring and summer) because that is when Mars is heated most intensely by the sun. As the planet approaches perihelion during the southern spring (*left*), local storms develop along the edge of the retreating polar cap. The dust raised by these storms spreads throughout the atmosphere. Because dust absorbs solar radiation, the atmosphere is heated; tidal winds and the Hadley circulation are thereby intensified. The effect is particularly strong in the southern subtropics, where larger local storms develop (*middle*), raising more dust and further heating the atmosphere. A few days later tidal winds become strong enough to raise dust throughout the southern subtropics. The dust fills the rising branch of the cross-equatorial Hadley cell and is thereby spread to the northern hemisphere (*right*). The storm subsides when the atmosphere becomes so laden with dust that its lower levels are no longer heated by sunlight. At that point the surface winds slacken and dust-raising ends.

when the heating of its atmosphere and the vigor of its circulation are greatest. According to a theory proposed in 1973 by Leovy, Richard W. Zurek, now of the Jet Propulsion Laboratory, and Pollack, global storms are attributable to a positive-feedback effect between surface winds and dust in the atmosphere. Dust raised by the winds tends to intensify them because it absorbs solar radiation and thereby heats the atmosphere directly.

According to this theory, winds driven by the strong temperature gradient along the edge of the retreating south polar cap generate numerous small storms that gradually increase the general level of atmospheric dustiness. The dust intensifies the Hadley circulation as well as the so-called tidal winds. (These winds are a result of the daily heating cycle: air rises on the warm daytime side of the planet and flows toward the nighttime side.) The combined strength of the winds is greatest in the southern subtropics, where insolation is greatest. The topography there favors the formation of additional storms that raise more dust into the atmosphere.

Once a critical level of dustiness is reached, the positive feedback accelerates rapidly. The Hadley circulation expands and spreads dust to the equator and beyond. A computer simulation I developed showed that after six days of dust-raising the dust cloud can spread to the northern hemisphere. The feedback continues until the atmosphere is so laden with dust that its lower layers receive little heat, whereupon surface winds slacken.

The dust cycle may feed back into the carbon dioxide and water cycles in a way that would help to explain the differences between the north and south polar caps. Given that global dust storms occur in the northern winter but not in the southern winter, it is natural to suppose the north carbon dioxide cap incorporates more dust as it is forming. Since a dirty frost cap reflects less solar radiation and absorbs more than a clean one, all the carbon dioxide would sublime from the north cap by the beginning of summer, whereas the cleaner and brighter south cap would survive. An analysis of Viking data by David Paige of the J.P.L. shows that the south cap is indeed brighter than the north cap.

Unfortunately other Viking data undermine this straightforward hypothesis. Global dust storms do not occur on Mars every year; they were observed in only two of the four years the *Viking 1* lander was operating. If the hypothesis were correct, the seasonal variation in the atmospheric carbon dioxide level, and hence in surface pressure, should be different in a year with global storms from the variation in a year without them. In a year without storms the north cap would presumably be clean, and it might even survive summer as the south cap does. Thus the surface pressure in the northern spring would be lower than if all the cap's carbon dioxide had sublimed into the atmosphere. Yet the Viking data show no significant difference from year to year in the seasonal pressure variation. Evidently the carbon dioxide cycle is not governed directly by the presence or absence of global dust storms.

Paige has suggested that dust may nonetheless explain the annual dissipation of the north cap and the survival of the south cap. Through a close analysis of the Viking data he has found that the caps grow brighter with the approach of summer, as the amount of sunlight falling on their surface increases. The reason, he argues, is that dust particles on the surface of the ice become warm enough to sublime the carbon dioxide grains that support them. The dust particles sink into the ice, leaving behind a cleaner and brighter surface.

The reflectivity of the cap thus depends not on the amount of dust deposited but on the depth to which the particles sink, and that depth depends on the amount of solar energy that falls on the cap in spring and summer. Because Mars is closer to the sun during the southern summer than it is during the northern summer, the south cap is always brighter (more reflective) than the north cap, regardless of yearly variations in the dustiness of the atmosphere. In other words, the south cap survives the summer precisely because the southern summer is intense. A polar cap that protects itself from the sun in this way would have an important influence on the long-term stability of the Martian climate.

There is good reason to believe the Martian climate changes cyclically, because like the earth Mars undergoes periodic variations in three orbital parameters: the eccentricity, the obliquity (tilt of the rotational axis) and the direction of the axis, which slowly precesses. Each of these variations can affect the climate by changing the seasonal and latitudinal distribution of incoming solar energy. An increase in the eccentricity increas-

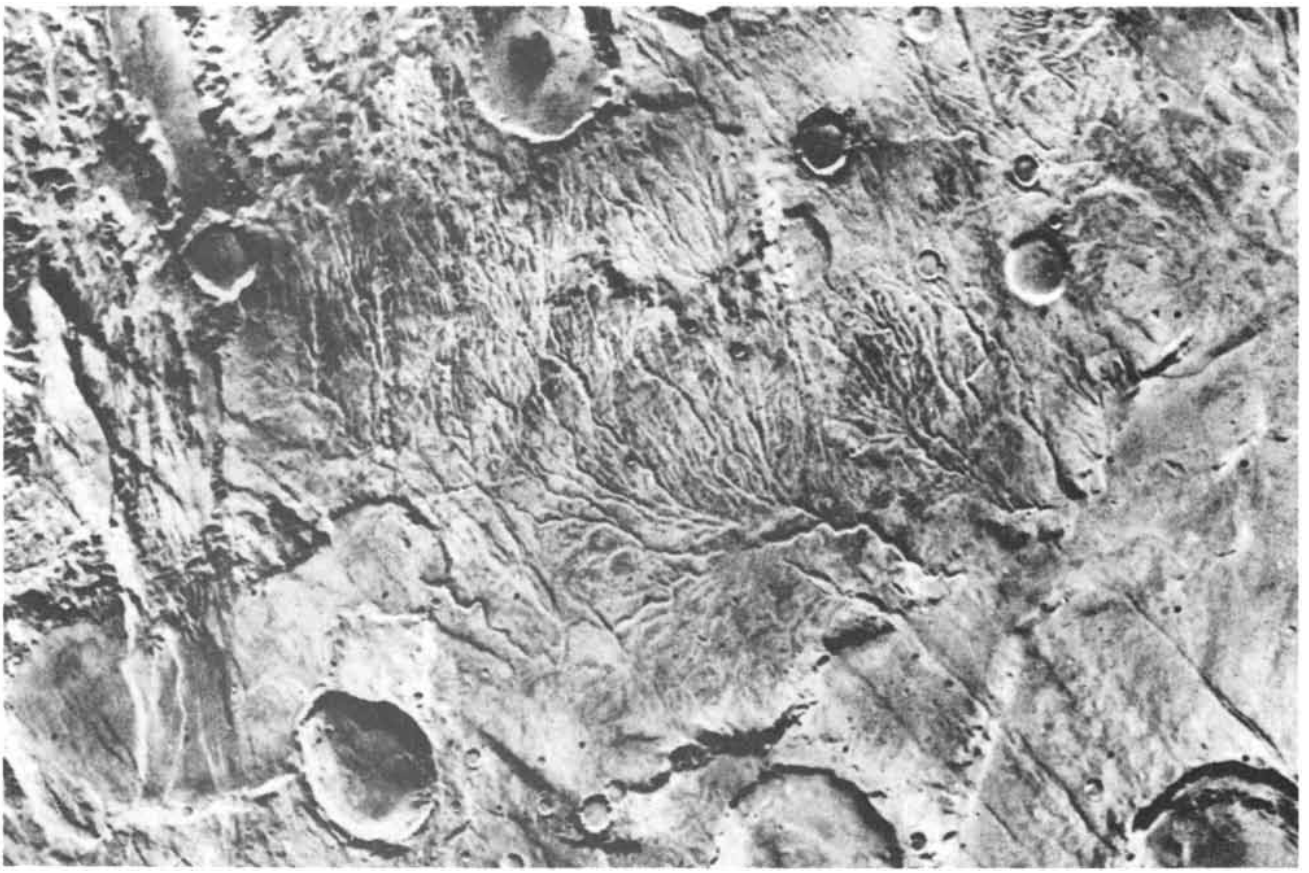
es the difference between the hemispheres in the intensity of the seasons; an increase in the obliquity exposes the poles to more sunlight; the precession of the axis determines at which season the planet is at perihelion.

The earth's orbital variations are widely thought to account for the oscillation of its climate over the past million years between ice ages and warm periods. The orbital variations of Mars are much larger than those of the earth. (They also have longer periods, on the order of 100,000 to one million years.) The obliquity of Mars, for example, which is currently 25.2 degrees, ranges about 13 degrees in either direction from the current value. In contrast, the earth's obliquity varies by only about one degree.

The primary climatic effect of variations in the Martian orbit would be to change the atmospheric surface pressure on the planet by changing the amount of carbon dioxide in the atmosphere. It is likely, for instance, that a summer-surviving carbon dioxide ice cap forms on at least one of the poles during periods of low obliquity, when the poles receive little sunlight. The surface pressure during such periods would be comparatively low. Dust storms would be rare or nonexistent, because when the surface pressure is low, much stronger winds are required to lift dust off the surface. Conversely, during periods of high obliquity the carbon dioxide ice cap would evaporate, and carbon dioxide might also be driven into the atmosphere from the regolith. The surface pressure would increase and dust storms would become frequent.

Mariner 9 and the Viking orbiters may have found evidence of such changes. At both poles the orbiters photographed layered terrain whose age has been estimated at less than a few million years. The nearly uniform spacing and thickness of the layers suggests they were formed by atmospheric sedimentation processes (dust settling out of the atmosphere) modulated by periodic climatic changes. Owen B. Toon of the Ames Research Center and his co-workers have proposed that the layers are buried sheets of water ice. According to their hypothesis, each sheet was laid down during a period of low obliquity, when the carbon dioxide ice caps froze water vapor that had been driven into the atmosphere from permafrost. During the ensuing period of high obliquity, dust raised by frequent global storms buried the sheet before all the water ice could evaporate. A succession of obliquity cycles could thus have produced the stacked layers visible today.

Although it seems clear that Mars has undergone cyclical climatic changes, it is by no means clear how pronounced those changes have been. It has been suggested that the atmospheric pressure on Mars could vary from less than a millibar to several times its current value of 6.1 millibars. The magnitude of the pressure fluctuations, however, depends on how much carbon dioxide can be stored in a summer-surviving ice cap and on how much is exchanged between the atmosphere and the regolith. Neither of these quantities is known. Furthermore, if Paige's hypothesis is correct and the Martian ice caps protect themselves from increases in solar radiation, then the pressure swings caused by evaporation of the ice would be damped. When the *Mars Observer* enters polar orbit, it will eliminate some of these uncertainties. In the process it should help investigators to determine the nature and intensity of changes in the Martian climate.



VALLEY SYSTEM in the southern hemisphere of Mars is evidence that the planet's climate was once warm enough to support liquid water. The branched pattern of the channels, which are thought to be four billion years old, suggests they were formed by converging streams; their small size indicates the water-discharge rate was low. The area in the photograph is 250 kilometers wide.

Graphic proof of high-tech leadership.

The Casio fx-7000G $79.95 It's the world's first programmable scientific calculator with a graphic display. The only calculator that can give form to your figures, instantly drawing graphs that represent your formulas and calculations on its big LCD screen.

And with a screen measuring 16 characters by 8 lines, the fx-7000G can display enough information for some of your biggest ideas. The formula or graph can then be stored and, when needed, brought back for an instant replay.

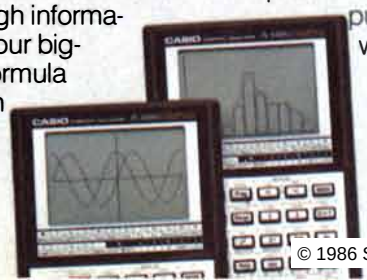
The fx-7000G has 82 scientific functions—including linear regression, standard deviation, as well as binary, octal and hexadecimal calculations—making it useful in virtually any scientific or technical endeavor.

In fact, without requiring computer knowledge—or costing computer dollars—the fx-7000G, with its 422 step memory, performs many functions you'd otherwise find only on a computer.

There's ample evidence of Casio's high-tech leadership throughout our entire line of scientific calculators. Our smaller-screen fx-6000G, for example, offers many of the features of our fx-7000G at even a smaller price.

If you have any doubt as to who's the leader in high-tech calculators, simply pick up an fx-7000G and draw your own conclusions.

For more information call 1 (800) 553-3338 Ext. 20



CASIO®
Where miracles never cease

Casio, Inc. Consumer Products Division. 15 Gardner Road, Fairfield, N.J. 07006

INTRODUCING THE NEW

Take the Eldorado hallmarks: sporty elegance and legendary luxury. Add aerodynamic design, exhilarating performance and microchip technology. The result is an Eldorado that sets a new standard for personal luxury cars.

AERODYNAMIC DESIGN.

From smoothly curved fenders to tailored door-into-roof design, this is the most aerodynamic Eldorado ever. That means wind noise is reduced. Road stability is remarkable. And the styling is timeless.

GM

LET'S GET IT TOGETHER...BUCKLE UP.

ELDORADO. REDESIGNED AND REDEFINED FOR 1986.

EXHILARATING PERFORMANCE.

Eldorado pulls away with a front-drive, transverse-mounted V8 engine, a Cadillac exclusive. The luxurious ride is refined with a sports car-inspired independent rear suspension that helps master the corners and curves without excessive lean or sway. This is a driver's car.

MICROCHIP TECHNOLOGY.

Microchip technology and digital readouts abound. The new, totally integrated Driver Information Center keeps you continually updated on the status of your Eldorado. Think of it as the monitor of the powerful, "personal" computers that keep Eldorado continually operating at optimum levels.

ONLY AT YOUR CADILLAC DEALER.

You're invited to personally test drive the new Eldorado.

CADILLAC'S 4-YEAR/50,000-MILE LIMITED WARRANTY

The quality of Eldorado is backed by a 4-year/50,000-mile limited warranty.

In some cases, a deductible applies. See your dealer for details.

BEST OF ALL...IT'S A CADILLAC.

SCIENCE AND THE CITIZEN

Genes, Not Jeans

Advances in biotechnology have brought U.S. agriculture to the brink of its third 20th-century revolution. According to a two-year study by the congressional Office of Technology Assessment (OTA), the tools of molecular biology may change the economics of agriculture as profoundly as they can alter the DNA of crops or livestock.

The first major change was the mechanical revolution (1920–50); it materialized as farmers switched from the power of animal muscle to machine power, thereby boosting U.S. agricultural production. The dramatic rise in productivity was accelerated by the second revolution (1950–80), in which farmers were able to exploit chemicals to control pests and disease and to fertilize their crops. The effects of the third revolution, the biotechnology revolution, may be even more profound, according to the OTA.

By means of biotechnology, microorganisms can be converted into manufacturers of hormones, enzymes and feed supplements that have been too expensive or too difficult to produce in large quantities. For example, genetic instructions for the synthesis of a hormone that can increase milk production in cows by as much as 30 percent can be inserted into the DNA of the bacterium *Escherichia coli*. The bacteria then produce economically significant quantities of the hormone.

Another possibility is the insertion of genes for new traits into the reproductive cells of livestock and poultry. The descendants of the modified parents might display such traits as accelerated growth, improved muscle mass or immunity to a particular disease.

Biotechnology can modify crops so that they make more nutritious proteins, resist insects and disease, grow in harsh environments or produce their own nitrogen fertilizer. Another technique, cloning, generates intact plants from single cells, yielding genetically identical individuals that can be superior in various ways to those grown from seed. It is also possible to transfer genes controlling desirable traits from one species to another.

All these changes, the OTA notes, can be expected to improve the competitiveness of U.S. agriculture in world markets. On the other hand, they may alter radically the structure of the domestic agricultural sector. The number of moderate-size farms, the report explains, might be sharply reduced. The high yields expected from the new technology, together with the economic advantages of large farms, will create "substantial incentives for farms to grow in size," placing owners of moderate-size farms under pressure to sell. (A farm of moderate size is currently defined as selling from $100,000 to $199,000 per year in farm products.) The result could be that by the year 2000 the agricultural system will consist of 50,000 farms that are large (sales of $200,000 to $499,000) or very large (sales of $500,000 or more). Together such farms would account for 75 percent of all agricultural output; 1.2 million small or part-time farms would produce the balance. In 1982 there were 2.2 million farms, of which 86.5 percent were small or part-time, 8.1 percent of moderate size and 5.4 percent large or very large.

Large and very large farms, the report speculates, should be able to make the transition to biotechnological methods without the aid of any Government farm-support programs. Farms of moderate size will need help. Income supports in particular, the study found, provide significant benefits to farms of moderate size, which otherwise have trouble keeping afloat in the face of the trend to bigger farms; Congress should therefore consider excluding large farms from support programs.

Diagnosis and Prescription

The Medicare system celebrates its 20th birthday this year in a health-care environment very different from the one in which it was devised. How well has it adapted to the new technologies, treatment facilities and payment schemes characterizing contemporary U.S. medicine, and how must it change to adapt to them better? These questions are addressed in recent reports from the Harvard Medicare Project and from the Office of Technology Assessment.

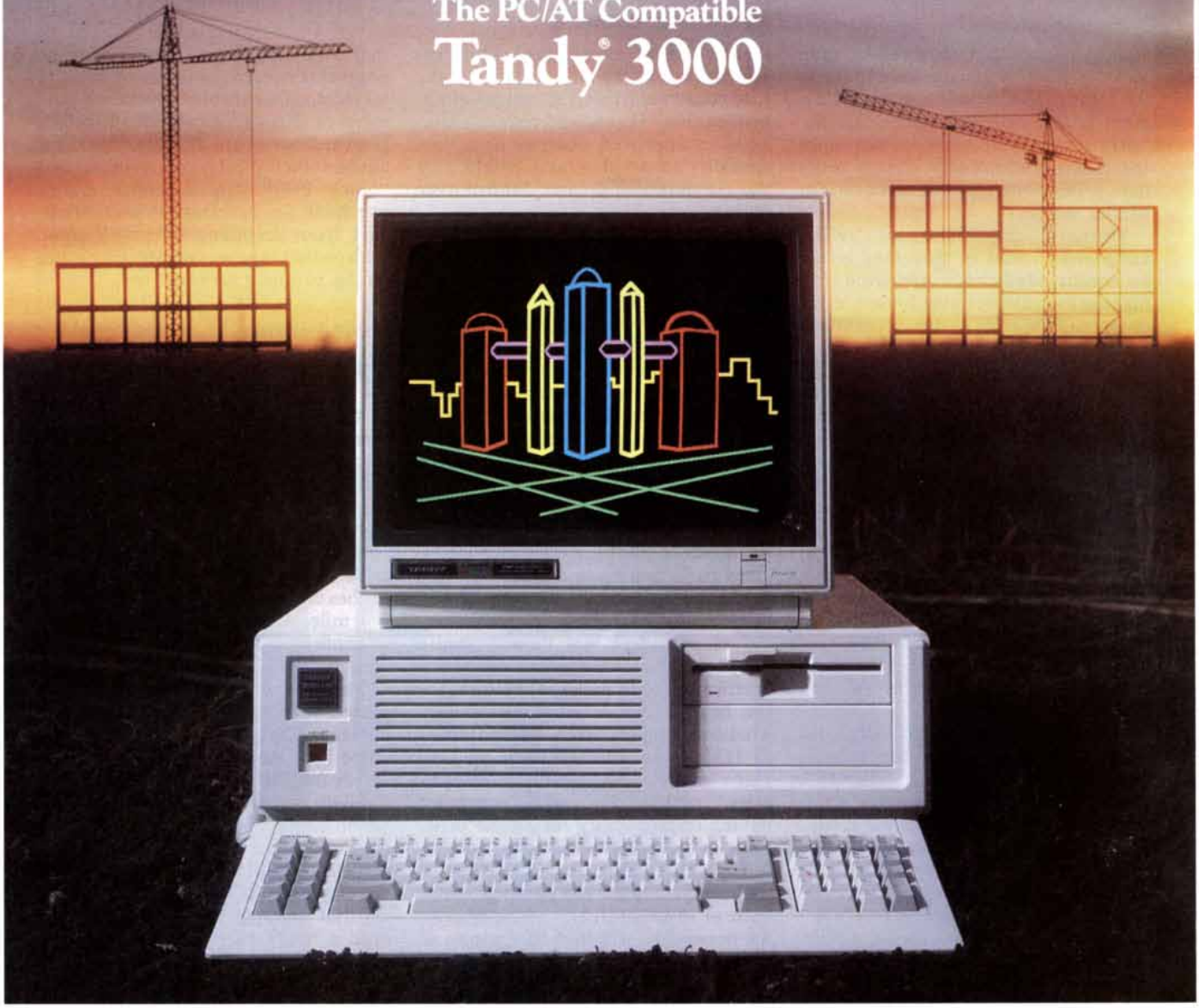
The Harvard report suggests that the system no longer addresses the health-care needs of the elderly as well as it did when it was first implemented. In particular, the report says, "older persons now pay a larger portion of their income for health care than they did before Medicare was enacted." The program "has become so complicated that few elderly people understand how it works, what it entitles them to, or even where to go to find answers to their questions." Current methods of payment tend to discriminate against the poorest and sickest patients and make it difficult for a beneficiary to predict how much he or she will have to spend out of pocket for health care in any given year. The Harvard Medicare Project outlined 40 proposals addressing these problems. An article in the *New England Journal of Medicine* by David Blumenthal, Mark Schlesinger, Pamela Brown Drumheller and other members of the project summarizes the most important recommendations.

One aspect of Medicare the report criticizes is the system of deductibles and copayments. (In general, Medicare pays 80 percent of the approved charges for visits to physicians after the patient has paid a certain annual "deductible"; the remaining 20 percent, which must be paid by the beneficiary, is called a copayment.) The report states that copayments place an added burden on the sick, who require more services, and on the poor (because the amount of a copayment is not related to the beneficiary's income). Because copayments are a set percentage of medical charges, they have become increasingly burdensome as the cost of medical care has risen. In the past 10 years "the total amount that Medicare beneficiaries paid in copayments has risen 50 percent faster than their income."

The report advocates reducing or eliminating Medicare copayments and deductibles. The physician copayment requirement should be reduced from 20 to 10 percent of approved fees, the authors find. The lost revenue could be made up by increasing the total take from premiums, but the premiums should be scaled to a beneficiary's income.

A major proposal is that Medicare should cover nursing-home care and the outpatient care of chronically ill elderly people. These costs are now covered partly by Medicaid, but in many states the elderly are not eligible for Medicaid unless they first impoverish themselves and their households. "Our society cannot in good conscience continue to deal with chronic illness in this manner," the report maintains, because "it robs the elderly of their dignity and of any real hope of economic and emotional security."

A number of reforms are recommended in Medicare's reimbursement of doctors and hospitals. Physicians are currently allowed to charge the "customary, prevailing and reasonable" fee for each service performed. Under this "CPR" system the physi-



Making your ideas for today take shape for tomorrow.

To turn your ideas into reality, you need a tool for productive thinking: the Tandy® 3000.

The Tandy 3000 is compatible with programs designed for both the IBM® PC/AT and PC/XT. So whether your application involves statistical analysis, program development, database management or computer-aided design, you can get the job done on a Tandy 3000.

We went beyond PC/AT compatibility

Because the Tandy 3000 operates at 8 megahertz, you get the job done 70% faster than most pc's, and 33% faster than the PC/AT. The Tandy 3000 handles complex operations with incredible efficiency.

Power to share

The Tandy 3000 goes beyond conventional thinking, too. Since it is designed to use XENIX® as well as MS-DOS®, you can add low-cost display terminals so several associates can use the 3000 simultaneously.

Tandy . . . Clearly Superior™

Take command of your future with the Tandy 3000—only $2599. Or choose the Tandy 3000 HD, with a 20-megabyte hard disk drive for just $3599.

Available at over 1200 Radio Shack Computer Centers and at participating Radio Shack stores and dealers.

Radio Shack®
COMPUTER CENTERS
A DIVISION OF TANDY CORPORATION

Send me a Tandy 3000 brochure.
Radio Shack, Dept. 86-A-572
300 One Tandy Center, Fort Worth, TX 76102

Name _____
Company _____
Address _____
City _____
State _____ ZIP _____
Phone _____

Prices apply at Radio Shack Computer Centers and participating stores and dealers. Monitor and MS-DOS not included. IBM/Registered TM International Business Machines Corp.
MS-DOS and XENIX//Registered TM Microsoft Corp.

cian is encouraged to provide more services than necessary, according to the report. The present system of payment is also said to discriminate unfairly among types of physicians, since specialists traditionally charge more than primary-care physicians even for similar services.

Moreover, physicians may "refuse assignment"; that is, they may refuse to accept Medicare's approved fee, making the patient liable for any extra charges. This adds to the general unpredictability of out-of-pocket expenses and to the confusion of beneficiaries. The Harvard report recommends replacing the CPR system of reimbursement with a sliding scale, which would list the relative value of various services regardless of the provider (although certain services could be provided by only a few kinds of physicians or facilities). It also recommends that physicians be required to accept assignment. Perhaps most important, the project members propose a step-up in Medicare's turn toward encouraging beneficiaries to join health-maintenance organizations (HMO's) and other prepaid plans.

The OTA report deals almost exclusively with payment plans and focuses closely on controlling costs. The OTA offers four basic strategies for Congress to choose from. The first strategy is to make minor reforms in the current system of payments, perhaps including a reduction of approved charges. While these reforms were being effected, Congress might undertake research on more drastic refinements. The second strategy, which resembles one of the Harvard proposals, is to construct overall schedules of approved fees. This would reduce variations in approved charges from one specialty to another and would give the Government greater control over its expenses.

A more radical reform would be to move from the fee-for-service system to payment for packages of services. The provider would bear more of the financial risk of caring for the patient and would be discouraged from providing unnecessary services. In this case it would be important for Medicare to oversee physicians in order to ensure that the quality of the care provided did not decrease. Payment for packages of services would require the primary physician to act as a kind of contractor, distributing a fixed amount of payment among the various physicians participating in a case.

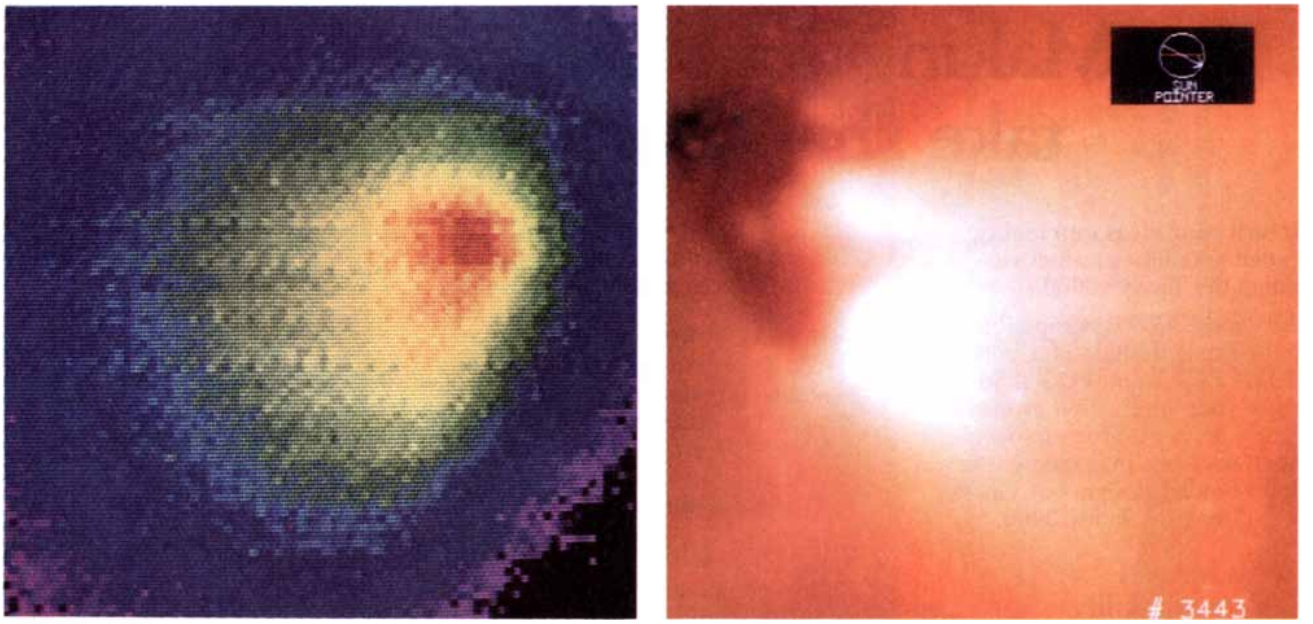
The most radical strategy would drop the system of payments to individual providers and instead would adopt a capitation system, in which Medicare would pay a set amount for the health care of each beneficiary. Medicare might make this payment to HMO's and other group medical plans or to "geographic intermediaries," such as insurance companies having responsibility for all the beneficiaries in a particular state or county.

The OTA report lists a number of options that might be pursued under each of these strategies and evaluates the possibility, within each strategy, of controlling costs. The major thrust is that Medicare should concentrate on payments covering a larger number of services, either through packaged payment or through capitation. Like the Harvard Medicare Project, the OTA advises that piecemeal changes and "quick fixes" cannot suffice; major overhauls are necessary to keep Medicare from becoming a heavy burden both on the taxpayers who finance it and on the people it serves.

Joint Venture

Early this spring more than 100 scientists from 12 countries gathered at the Soviet Institute for Space Research in Moscow as the unmanned Soviet spacecraft *Vega-1* approached Halley's comet and beamed computer-enhanced television images of the encounter onto two viewing screens. The stream of images and data from that craft and others in a flotilla of five probes sent over a distance of 100 million miles to rendezvous with the comet confirmed the prevailing model of comets.

Vega-1 and its sister craft *Vega-2* passed within 6,000 miles of Halley's nucleus. Cameras on *Vega-1,* whose name is a contraction of the Russian words for Venus (Venera) and Halley (Gallei), are the product of a collaboration between scientists from Hungary, France, the U.S.S.R. and Czechoslovakia. The devices provided the first direct glimpse of the nucleus, confirming that the nucleus of a comet is indeed a "dirty snowball." That model, initially proposed 35 years ago by Fred



HALLEY'S COMET was the target of an international armada of spacecraft in early spring. In the false-color image from the Soviet craft *Vega-1* (*left*) the brightest parts are red and the fainter regions are orange, yellow, green and blue. The image from the European probe *Giotto* (*right*) shows two jets of dust (*bright regions*) emanating from the comet's nucleus, part of which is seen at the left (*dark area*).

L. Whipple of the Smithsonian Astrophysical Observatory, holds that the nucleus of a comet consists of a mixture of ice, silicates (a class of minerals) and perhaps metals. Whipple's theory has been corroborated over the past several years by measurements made with infrared telescopes indicating that although comets contain a considerable amount of dust, they are still predominantly composed of ice. Astronomers announced that Halley's "dirty snowball" is a potato-shaped object about 10 miles long and five miles wide.

As the two Soviet craft closed in on the nucleus, two Japanese satellites hurtled past the tail. *Sakigake,* meaning pioneer, flew to within 4.3 million miles of the comet's head; *Suisei,* meaning comet, came within 94,000 miles. Halley's tail, like that of any other comet, forms when the comet approaches a star such as the sun. The nucleus absorbs radiation and heats up until the ice sublimates, or goes directly from the solid to the gaseous phase. The result is an outflow of gas and dust particles. The Japanese craft confirmed earlier reports of outbursts, or sudden increases in intensity, in the flow. Apparently much of the outflow originates from a crack and two point sources on the surface of the nucleus. When one of those sources faces the sun, the outflow increases.

An ultraviolet camera aboard *Suisei* also confirmed a phenomenon known as bow shock, which arises from interactions between the gases of a comet and the solar wind (a stream of particles emanating from the sun). Astronomers have theorized that just as the bow of a moving ship creates a wave, so the passage of a comet through the solar wind should create a sharply defined magnetic boundary. The Japanese craft as well as others in the armada found that Halley's bow-shock region is, as expected, about 600,000 miles from the nucleus.

Information relayed from the four satellites helped the fifth member of the group, the European spacecraft *Giotto,* to come to within 335 miles of the nucleus. *Giotto* was named for the Florentine painter Giotto di Bondone, who apparently saw the comet in 1301 and portrayed it as the star of Bethlehem in his fresco "Adoration of the Magi" (see "Giotto's Portrait of Halley's Comet," by Roberta J. M. Olson; SCIENTIFIC AMERICAN, May, 1979). In spite of damage from dust particles traveling at speeds as high as 155,000 miles per hour, *Giotto* was able to determine that the nucleus of Halley's comet is extremely dark; apparently much of the dust of the nucleus resides on its surface. Data from the satellite also indicate that the gas and dust from the comet's core are emitted in focused streams. The reason is not yet known; clarification will undoubtedly call for more of the international collaborative effort that has been the hallmark of the Halley missions.

Balloons over Venus

On their way to Halley's comet the two Vega spacecraft flew by Venus last June and for the first time dropped weather balloons—helium-filled Teflon spheres that carried instrumented gondolas—into the atmosphere of another planet. Floating at an altitude of about 54 kilometers, within the thick sulfuric acid clouds that hide Venus, the two balloons drifted westward nearly a third of the way around the planet, one on each side of the equator. Each balloon radioed back pressure, temperature and wind data for 46 hours, until its 18 flashlight-size batteries ran out. A preliminary analysis of the data has been presented in *Science* by the Russian, French and American workers who organized the experiment.

The Vega measurements are the first to document how atmospheric conditions on Venus vary horizontally and with time. As such they will eventually add substance to models of the Venusian atmospheric circulation, which is dominated by the powerful east-to-west winds that carried the balloons around the planet at about 250 kilometers per hour. From earlier probes it is known that the wind speed at the cloud tops (an altitude of 70 kilometers) is even greater, about 360 kilometers per hour. This means the cloud tops rotate 60 times faster than Venus itself, which completes an east-west rotation every 243 days. Whereas the earth's atmosphere is more or less locked to the planet, the Venusian atmosphere "superrotates" every four days.

Somehow the surface of the planet must be "pushing" the atmosphere, transferring angular momentum upward, and the central task of Venusian circulation models is to explain how. One of the most popular models, devised by Peter J. Gierasch of Cornell University, holds that the transfer occurs when warm air rises at the equator and then moves poleward. In the process its distance from the axis of the planet decreases, which decreases its moment of inertia. Like the proverbial figure skater, the air thereby picks up rotational velocity, because its angular momentum (the moment of inertia times the rotational velocity) must remain the same.

The Gierasch model makes two predictions about the Venusian atmosphere. First, it predicts a significant north-south circulation. Second, since the superrotating east-west winds are observed not just at the poles but all over the planet, the model requires the presence of large atmospheric eddies that spread angular velocity, through a type of frictional drag, from the poles back to the equator. (An eddy is a wavelike disturbance that leads to deviations from perfect latitudinal symmetry, a condition in which the temperature, pressure and winds are constant along a band of latitude.)

It is too early to say whether the Vega balloons discovered north-south winds. The balloons were tracked on the earth by 20 radio antennas, and these convoluted tracking data, which will be the chief source of information on horizontal winds, have yet to be fully analyzed. Evidence for large eddies, however, has already emerged. The temperature recorded by the *Vega-1* balloon seven degrees north of the equator was consistently 6.5 degrees Celsius higher than that recorded by *Vega-2* at seven degrees south. The best explanation is that *Vega-1* was at the peak and *Vega-2* at the trough of an eddy that, like the balloons, moved with the westward winds.

The turbulence encountered by the balloons is further evidence for eddies on Venus. Although the balloons were deployed at an altitude where convective air currents were expected, the vertical winds (which unlike horizontal winds could be measured directly by onboard sensors) turned out to be far more powerful than had been expected. Both balloons consistently recorded updrafts and downdrafts that were stronger than typical winds on the earth; the peak downdraft of more than 12 kilometers per hour was comparable to the vertical winds in a terrestrial hurricane.

The convective layer in which the balloons drifted is only about five kilometers thick. It is bounded on both sides by stable layers that are not subject to turbulent mixing. Strong updrafts and downdrafts in the convective layer, however, might create bulges where they plow into the stable layers. The bulges would spread out in all directions as eddies, and the eddies, according to the Gierasch model, might help to spread the superrotating winds.

4.2-Meter Dash

Two physicists have timed what appears to be the fastest recorded chemical reaction. Tien Tzou Tsong and Yung Liou of Pennsylvania State University measured the time it takes the rhodium helide ($RhHe^{2+}$) ion to dissociate into its constituent elements,

rhodium and helium, when it is subjected to an electric field. They found the reaction takes place in about eight ten-trillionths (10^{-13}) of a second.

That kind of interval cannot be resolved by current electronic timers. As Tsong and Liou report in *Physical Review Letters,* they had to measure the dissociation time indirectly by modifying the device in which the rhodium helide ion was first detected: a field-ion microscope. Such a microscope images individual surface atoms at the tip of an extremely fine needle. The needle is put in a chamber filled with helium gas and is positively charged to between 5,000 and 10,000 volts. The resulting electric field at the tip of the needle is high enough to ionize helium atoms near the atoms of the tip's surface. The ions are then accelerated radially from the tip toward a fluorescent screen, where they impinge to form a visible pattern of light spots. Each spot corresponds directly to an atom at the tip's surface.

If a needle in such a microscope is subjected to higher voltage pulses, the result is field evaporation: the metal atoms at the surface of the needle tip are themselves emitted as positively charged ions. Some of the metal atoms first combine with helium atoms adsorbed on the surface, however, and then fly off as metal-helide ions. A small fraction of these compound ions dissociate quickly into a metal ion and a helium atom.

In order to measure such a dissociation reaction for rhodium helide, Tsong and Liou had to stretch out the reaction time by orders of magnitude. They did so by making a small opening opposite the microscope's needle, through which individual accelerated ions could leave the chamber and enter a 4.2-meter evacuated "flight tube" leading to an ion detector.

The time required for the breakdown of rhodium helide in this modified field-ion microscope could be determined from the slight difference in the kinetic energy, or velocity, of the rhodium ions detected at the end of the flight tube. The difference arises from the fact that some of the rhodium ions come directly from the needle surface whereas others are produced when the compound rhodium helide ion dissociates. Although this energy difference is as minuscule as the reaction time, the relatively great distance both types of ions must travel in the flight tube not only expands the timing period from trillionths of a second to tractable millionths but also effectively segregates the two groups of ions according to arrival time. The clear difference in the average flight time is the critical datum: it provides the accurate measure from which the reaction time can be calculated.

One problem remained: how to start the electronic stopwatch at the precise instant the ions are emitted from the needle surface. This was accomplished by charging the needle to a voltage just below what is required for field evaporation. Then a short laser pulse was fired at the needle tip to serve in effect as a starting gun. It supplied a burst of energy sufficient to induce the surface atoms to leave the tip and at the same time triggered the timing mechanism.

Piecework

The 19th-century prospectors who found gold in Alaska trekked through a rugged landscape that turns out to have made a far longer journey than their own.

George Gehrels, now at the University of Arizona and earlier in the laboratory of Jason B. Saleeby at the California Institute of Technology, has deduced the history of the Alaskan fragment called the Alexander terrane. It underlies much of the Alaska panhandle, but it was not always there. Once it was part of Australia. Beginning some 375 million years ago, it seems to have journeyed across the Pacific from Australia to South America. Then it turned north, scraping its way up the Americas. Gehrels hypothesizes that in the course of its wanderings and collisions it carried a goldfield from California to Alaska.

Gehrels gathered evidence for the Alexander terrane's odyssey during five successive summers he spent exploring the geology of the terrane. The oldest rocks, from 530 to 600 million years old, were volcanic; they suggest the terrane was born as an island arc, like Java or Japan. Then came an eventful history: a period of orogeny, or mountain building, from 530 to 490 million years ago; a second period of volcanic-arc evolution, from 490 to 425 million years ago, and a second period of orogeny, from 425 to 400 million years ago. For about 25 million years the mountains shed debris. After that came tectonic quiescence, from 375 to 250 million years ago.

To locate the origin of the Alexander terrane Gehrels compared its geology with what is known about the geology elsewhere along the rim of the Pacific. An excellent match emerged: the rocks of the part of eastern Australia known as the Lachlan fold belt showed periods of volcanic activity interleaved with two periods of orogeny at the same times as the ones in the Alexander terrane. Gehrels then found a corroborating match. The paleomagnetic record locked into the rocks of the Alexander terrane gave a history of the latitude of the terrane. The latitude values had been interpreted as applying to the Northern Hemisphere; Gehrels hypothesized that they might really be southern. On the new assumption the latitude of the Alexander terrane proved to match that of eastern Australia for the period from 450 to 375 million years ago.

Then something remarkable happened: the magnetic record indicates that the Alexander terrane moved somewhat to the north while Australia moved south. At precisely the same time the geologic histories diverge: the Alexander terrane entered its long tectonic quiescence, whereas eastern Australia continued to be tectonically active. Apparently Australia and the Alexander terrane had rifted apart. From 375 to 225 million years ago the Alexander terrane maintained a roughly equatorial latitude, but it was not stationary. Paleontology suggests what may have been going on. A stratum of the terrane from 225 million years ago has yielded fossils such as a species of the Triassic bivalve *Monotis* hitherto found only in Peru. Evidently the terrane had moved eastward across the equatorial Pacific basin and had come up against South America.

The magnetic record suggests that by 200 million years ago the terrane was moving northward. The rocks deposited on it 160 million years ago resemble rocks in California, and Gehrels hypothesizes that the terrane then formed the western side of a system of faults much like the San Andreas. By about 100 million years ago the terrane had arrived at Alaska and had become welded to the western edge of North America.

Gehrels was intrigued by sedimentary rock to the east of the Alexander terrane, sandwiched between it and the adjacent Stikine terrane. It too resembles California rock. He proposes that as the Alexander terrane moved north alongside California, a depression of the type termed a pull-apart basin formed between it and North America. The eastern side of the basin stayed in place; the western side went north. Both contained gold. Thus the travels of the Alexander terrane could conceivably account for two 19th-century gold rushes, one in California, the other in Alaska.

Marginal Existence

Along the edges of the plates into which the outer shell of the earth is divided, tectonics gives rise to singular life forms. At the midocean ridges, where plates move apart and volcanic activity builds new crust, seawater

seeps into the fractured ocean floor and emerges in hot springs, laden with minerals and dissolved hydrogen sulfide gas. Near some springs communities of exotic animals flourish, including giant clams and meter-long worms in tubelike casings. The organisms are sustained by bacteria that live symbiotically in their tissues and metabolize the hydrogen sulfide.

LaVerne D. Kulm, Erwin Suess and their colleagues at Oregon State University and at other institutions have found faunal counterparts to the rift communities at another kind of margin: a subduction zone, where plates converge and one plate dives under the other. The workers hypothesize in *Science* that the subduction-zone communities also derive metabolic energy from dissolved gas venting from the sea floor—but in this case the gas is methane.

The investigators spotted the clams, tube worms and other creatures at a depth of about 2,000 meters while exploring the Oregon subduction zone in the submersible *Alvin* during the summer of 1984. The subduction zone lies off the coast of Washington and Oregon where the Juan de Fuca plate, which forms part of the Pacific ocean floor, plunges under the North American plate. In the process thick layers of sediment are scraped off the descending plate and heaped up into ridges at the leading edge of the upper plate.

The workers measured small increases in the water temperature near the colonies of animals, an indication that warm fluids are seeping from the ocean floor. Water sampled just above the colonies contained anomalously high levels of methane and smaller amounts of radon, which results from radioactive decay in the earth's crust. Presumably fluids bearing methane produced by the breakdown of organic matter buried deep in the sediments are expelled as the plates converge.

Slab- and chimneylike formations of calcium and magnesium carbonates, observed at the sites, also hinted at the presence of methane. Methane itself could not give rise to such deposits; their source must be an abundance of carbon dioxide, seeping from the sea floor and combining with calcium and magnesium in seawater. The oxidation of organic matter by microbes just below the sea floor could account for some of the carbon dioxide, but unusually low levels of chemical tracers such as ammonia, which is released when organic matter is broken down, indicated that microbial oxidation of methane may supply much of the carbon dioxide that builds the deposits.

Analysis of soft tissue from the clams and tube worms suggested that the organisms congregate at the methane sources because they take nourishment there. The carbon in their tissues was unusually high in carbon 12, the lightest carbon isotope. The organisms' source of food must also be high in carbon 12, a characteristic of methane and other light hydrocarbons. As at the midocean ridges, symbiotic bacteria may enable the organisms to exploit their unusual food source.

Like the rift fauna, the subduction-zone communities may be widespread; since the authors' discovery other investigators have found similar organisms at subduction zones in the western Pacific and off Peru.

Athlete's Fate

Aside from fun, an attraction of regular exercise is the belief that jogging, swimming and other aerobic sports add years to one's life. A study of 16,936 men who entered Harvard College between 1916 and 1950 indicates that exercise does indeed yield additional years of life—two of them, in fact. This result is "a bit disappointing" to Ralph S. Paffenbarger, Jr., of the Stanford University School of Medicine, who headed the study and who exercises regularly himself. "I would have guessed that it would have been longer," he said.

The study, reported in the *New England Journal of Medicine,* was conducted by Paffenbarger, Robert T. Hyde of Stanford and Alvin L. Wing and Chung-cheng Hsieh of the Harvard University School of Public Health. Questionnaires sent to the participants in the study in 1962 and 1966 provided data on exercise: how much they walk, climb stairs and engage in sports. The investigators calculated the energy expended in these activities in kilocalories: 56 for walking seven city blocks (.94 kilometer); 28 for climbing 70 stairs; five per minute for light sports such as golf, bowling and table tennis, and 10 per minute for vigorous sports such as running, swimming and court games. They found that death rates were from a fourth to a third lower among men who expended 2,000 or more kilocalories per week in exercise than they were among less active men.

The study group then extrapolated the data to develop estimates of the years of life gained by being active rather than inactive. The absolute gain was 2.33 years; adjusted for differences in blood pressure, smoking habits, weight gain and age of death of parents it was 2.15 years. Ten percent more of the active men than of the less active alumni lived to age 80. "Our computations suggest," Paffenbarger said, "that

for every hour the active graduates spent exercising, they lived that hour over, plus an extra one or two hours."

Fatal Affinity

Every year in the U.S. the tremor and rigidity of Parkinson's disease overtake some 50,000 people. Underlying the symptoms is widespread cell death in a region deep in the brain known as the substantia nigra. The neurons of the substantia nigra make the neurotransmitter dopamine, one of the brain's chemical messengers, and their loss depletes dopamine in other areas of the brain controlling movement. The symptoms can often be alleviated by supplementing the patient's dopamine reserves with L-dopa, a dopamine precursor, but the neuronal degeneration continues. What attacks the substantia nigra?

Recent findings by Robert J. D'Amato, Zoe P. Lipman and Solomon H. Snyder of the Johns Hopkins University School of Medicine lend plausibility to suggestions that an unknown environmental toxin may be to blame. The speculations were touched off in 1981 when J. William Langston of the Stanford University School of Medicine and his colleagues identified a contaminant that was causing users of a synthetic "heroin" to develop symptoms almost identical with those of Parkinson's disease; one user even displayed the progressive micrographia, or gradual shrinking of handwriting, characteristic of Parkinson's sufferers. The contaminant, a by-product of the drug's synthesis called MPTP, presumably does not cause ordinary Parkinson's disease, but the biochemical pathways by which it destroys the substantia nigra may mimic those of a more widespread toxin.

Later work showed that in the body MPTP is rapidly oxidized by an enzyme that is abundant in the brain, forming a new product called MPP^+. MPP^+ binds to the same cell-surface receptors that enable neurons to take up dopamine and other neurotransmitters of the class called catecholamines from the fluid around the cells. The MPP^+ is thereby pumped into the neurons, where it exerts its toxic effects. A number of common substances, including chemicals found in tea, peppermint and herbicides, have molecular structures resembling MPTP's and might be metabolized similarly. The effects of such a compound, combined with the normal loss of neurons in the substantia nigra with age, could bring on Parkinson's disease.

Yet puzzles remained. Minute doses of MPTP suffice to kill dopamine neurons in monkeys and human beings, but mice and rats are far less susceptible. Even in the primate brain the chemical's effect is remarkably selective: although MPP^+ is transported throughout the brain, damage is confined to the dopamine neurons of the substantia nigra. Now D'Amato and his colleagues report in *Science* that certain features of the human substantia nigra make it uniquely vulnerable to MPTP and perhaps to other toxins.

One of those features is the dark pigment neuromelanin, to which the substantia nigra ("dark substance") owes its name. In vitro, the authors found, MPP^+ readily binds to synthetic neuromelanin and to neuromelanin isolated from the monkey substantia nigra. The workers suggest that MPP^+ transported into the dopamine neurons of the substantia nigra combines with the pigment there and is stored. Gradual release of the stored toxin maintains a high level of the substance within the cells, eventually killing them. The participation of neuromelanin would explain why rodents, whose substantia nigra is unpigmented, are largely immune to MPTP-induced Parkinson's disease. It would also account for the observation that older monkeys are the most sensitive to MPTP: the pigment builds up with age.

The locus ceruleus, a region of the brain stem, also contains neuromelanin in the primate brain and yet is not damaged by MPTP. D'Amato and his colleagues point out that the locus ceruleus is extensively infiltrated by catecholamine nerve fibers originating elsewhere in the brain. The terminals of those fibers, the workers believe, absorb most of the MPP^+ that forms in the area, leaving the nerve-cell bodies within the locus ceruleus unharmed. In the brains of mice that had been given large doses of MPTP the authors observed extensive destruction of nerve terminals in the locus ceruleus, reflecting the terminals' ability to take up MPP^+ and thereby protect adjacent cell bodies. In contrast, very few terminals infiltrate the substantia nigra, and so there the neuronal cell bodies absorb most of the MPP^+.

The authors found that several other substances known to be toxic to neurons also bind to neuromelanin. They propose that the substantia nigra has a generalized affinity for toxins, which might account not only for its destruction in Parkinson's disease but also for its normal deterioration in aging.

Son Killer

A bacterium has been isolated that prevents the wasp *Nasonia vitripennis* from producing male offspring. Discovery of the microorganism may help investigators to uncover the trigger for a curious method of reproduction found in several insect groups: parthenogenesis, or the development of an egg in the absence of fertilization by sperm. The discovery could also point the way to a new method for the control of some insect pests.

In order to reproduce, the female wasp mates with a male and stores sperm in a specialized internal sac. When she is ready to release her eggs, she deposits sperm into some of the eggs and then injects fertilized and unfertilized eggs through her stinger into a host, specifically a fly pupa. Normally eggs that are fertilized become females, and eggs that are unfertilized develop by parthenogenesis to become males. In some instances, though, a factor known as a son killer (*sk*) prevents parthenogenetic development from taking place and thus destroys the male eggs.

Until recently the nature of the *sk* factor was a mystery, but then John H. Werren of the University of Maryland at College Park and his colleagues discovered that the factor is associated with a chronic bacterial infection in the mother. They now report in *Science* that they have isolated the bacterium that leads to wastage of male eggs.

According to Werren, the wasp apparently passes the *sk* bacterium to the host pupa. There is no evidence that the bacterium enters the eggs; Werren suspects that a chemical produced by the bacterium enters the eggs from the pupa and somehow prevents activation of the genes that control parthenogenesis in the unfertilized egg (while allowing the fertilized eggs to develop normally). Now that the bacterium has been isolated, it may eventually be possible to identify and isolate the putative bacterial chemical, tag it with a radioactive label and thus determine whether it binds to the DNA in the unfertilized egg, and if it does, just where. By this technique or another one, investigators should be able to pinpoint the gene controlling parthenogenesis in the wasp and perhaps to pinpoint a protein or other substance normally activating that gene. In addition to furthering investigations into parthenogenesis, the wasp son-killer bacterium may someday prove to be an effective tool for pest control by virtue of its transmission perorally, or by mouth.

The son killer is actually one of three factors that distort the maternally determined sex ratio of offspring in the wasp (the origins of the other two factors are not known) and one of several sex-ratio distorters that appear to have evolved independently among insects. The widespread existence of such fac-

tors, which are thought to be associated with a variety of microorganisms, raises several intriguing questions that so far remain unanswered: What are the sources of the other two sex-ratio distorters in the *N. vitripennis* wasp? What can be the selective advantage for a microorganism of destroying another organism's male offspring? How is it that such microorganisms have persisted without decimating the populations they infect?

Breaking Out

A general practitioner in England writes in the *Lancet* that a patient presented him with a most unusual problem. The 16-year-old girl complained that an epidemic of black speckles periodically afflicted the antique-doll reproductions she makes. The author of the paper, Conrad M. Harris, formerly associated with St. Mary's Hospital, reports that wherever his patient touched a doll's face while she was painting it, the unsightly speckles appeared after the china was fired.

Harris and his colleagues, Rosemary H. Waring, Graham L. Hendry of the University of Birmingham and Stephen C. Mitchell of St. Mary's Hospital Medical School, suspected that the problem had its origin in the patient's sweat. When they sampled it, they found large amounts of several types of organic sulfide that are present in garlic. They then analyzed some of the pinpoint speckles and found both sulfur and iron; neither element was present in the unspeckled areas of the doll faces. Clearly, the workers concluded, the speckles were produced by a reaction (as yet unidentified) between garlic-derived sulfides and iron that was distributed inhomogeneously in the flesh-colored clay.

Further tests revealed that the girl belongs to a small group of people who are completely unable to excrete sulfoxides and other garlic metabolites in their urine and therefore excrete them solely in their sweat. The patient compounded the problem by eating "a great deal of garlic." The complexion of her dolls cleared up immediately when she adopted a simple therapeutic program: abstinence.

For some people such a regimen is harsh, however, and that may explain why the incidence of black speckles among England's china-doll population has been rising rapidly. Until recently, according to a representative of the school that trained the unfortunate girl, speckled dolls were rare; then the school began accepting not only English students but also students from the Continent, notably from France.

Aqueous Foams

They now have a wealth of uses, of which some are surprising. Yet the ability of a water-based liquid to become a frothy mass of bubbles requires a balance of forces not entirely understood

by James H. Aubert, Andrew M. Kraynik and Peter B. Rand

The suds of a shampoo, the lather of a shaving cream, the head on a glass of beer—each of these commonplace sudsy objects exemplifies a material whose properties are far from common. Each is an aqueous foam, an impermanent form of matter in which a gas, often air, is dispersed in an agglomeration of bubbles that are separated from one another by films of a liquid that is almost but not entirely water. Although aqueous foams have been a subject for scientific investigation since the 17th century, much remains to be learned about the complex chemical and physical phenomena that interact in even a single foamy bubble.

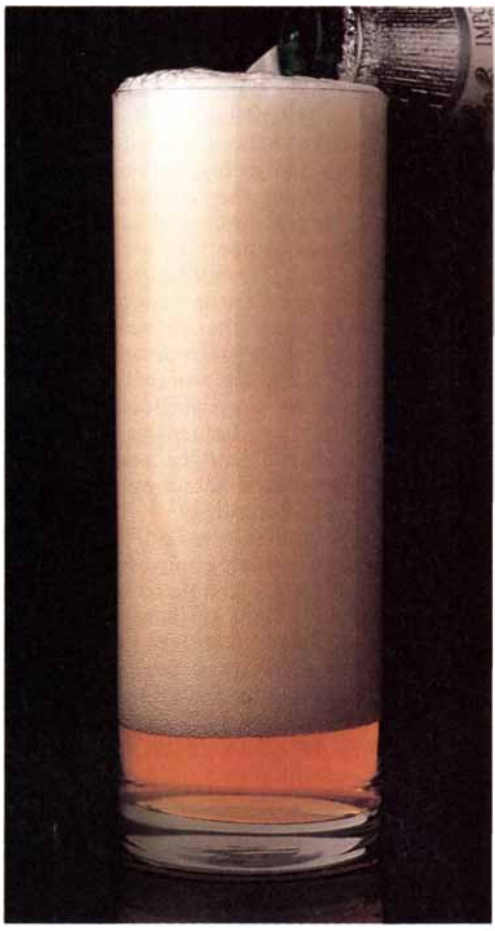
Like any other form of matter, an aqueous foam maintains its configuration only when it cannot readily transform itself into an arrangement embodying less energy. In the case of a foam the energy includes the energy of the gas contained in the bubbles of the foam, the chemical energy of substances in the liquid films that form the walls of the bubbles and most notably surface tension: the energy of the surfaces of the films. The molecules at each surface represent more energy than molecules deep in the film; thus each surface is thermodynamically unfavorable. To put it another way, a foam always "seeks" to adjust the details of its intricate structure in such a way as to minimize the total expanse of its films.

Each film, however, encounters an energy barrier. As liquid drains from the film, causing the film to thin, surface forces arise that are opposed to further thinning. The forces can be electrical, or they can be steric: they can result from the order that surfaces impose on molecules in the film. They can result from the presence of the very molecules, called surfactants, or surface-active agents, that enabled the liquid to foam in the first place. When destructive and preserving forces are in equilibrium, the films in a foam are rendered metastable, that is, stable with respect to small disturbances. This intricate balance has remarkable consequences. For example, all beer drinkers must have observed that in the moments after beer is poured into a glass the volume of the head shrinks while that of the liquid grows. Yet surely few beer drinkers know that as the foam drains, increasing the volume of beer, the bubbles remaining in the foam respond to a changing balance of forces by changing from spheres into polyhedrons.

Aqueous foams are typically 95 percent air and only 5 percent liquid, and remarkably the liquid is often 99 percent water. The remaining 1 percent consists of surfactants and other additives, such as alcohols and polymers. By means of these additives many familiar manufactured products (including shampoo, shaving cream and beer) provide a controlled amount of foaming, primarily for aesthetic reasons. In contrast, many aqueous foams are carefully engineered to perform a specific task. These include fire fighting, dust suppression, crop treatment and some tasks that may be more surprising. For example, aqueous foams are now being tested for their ability to lessen the violence of explosions. Our own involvement in the science of aqueous foams began five years ago, when we initiated at the Sandia National Laboratories an investigation of the uses of aqueous foams in geothermal wells.

Surface Tension and Surfactants

Two factors govern the ability of a liquid (under the influence, for example, of mechanical agitation) to froth into a foam. The first factor is surface tension. There are attractive forces between the molecules in any quantity of a liquid; these forces in fact cause the liquid state to exist. The principal forces are electrical and originate from individual charges or dipoles (permanent distributions of charges) on the molecules of the liquid. Even in nonpolar and uncharged molecules, electrical forces known as van der Waals forces arise because the elec-

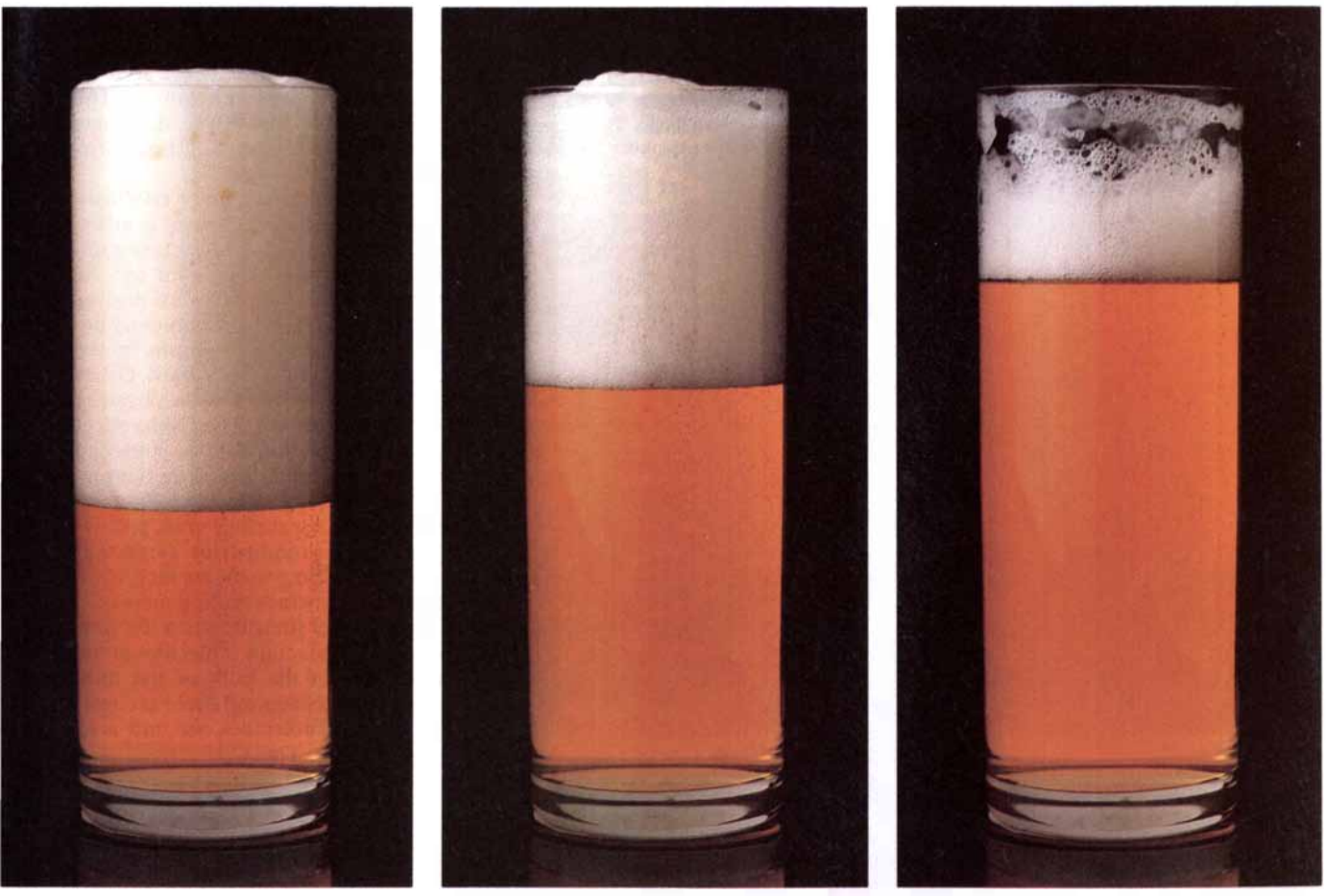


LIFE HISTORY of an aqueous foam is completed in less than an hour for the foamy head on a glass of beer. The foam is formed when the carbon dioxide dissolved in the beer returns to the gaseous state, yielding bubbles separated from one another by films of the watery liquid. At first the

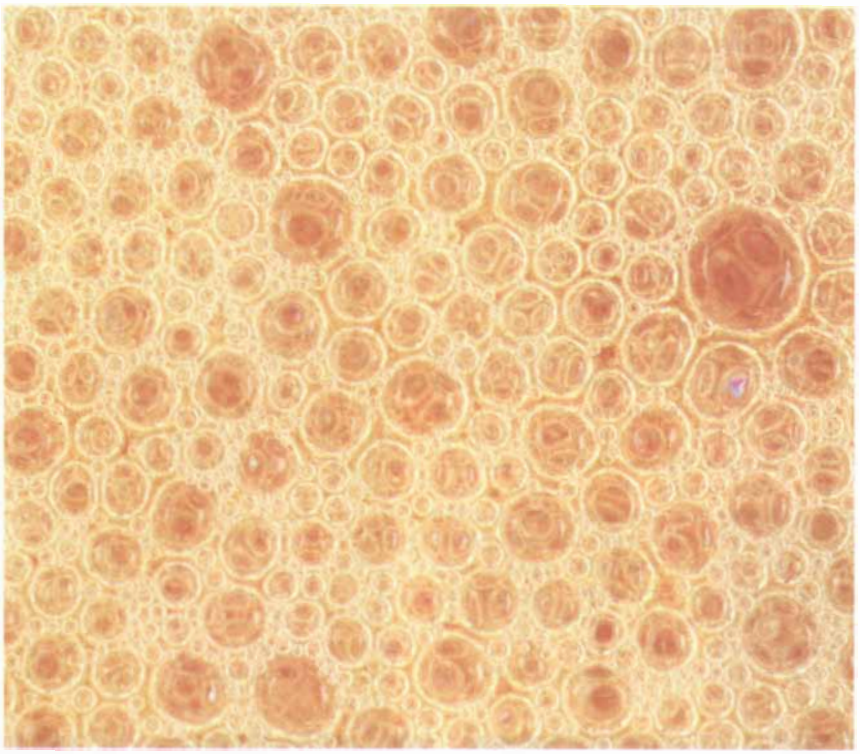
tronic "shell" of each molecule is constantly undergoing perturbations that give it temporary concentrations of negative or positive charge. The net result is an electrical attraction among molecules. Molecules deep in the liquid have identical surroundings in all directions; thus they are attracted isotropically by their neighbors, that is, equally in all directions. In contrast, molecules near the surface of the liquid have few if any neighbors closer to the surface; thus they are attracted preferentially toward the bulk fluid. Moving a molecule from the bulk fluid to the surface therefore requires work against the forces in the liquid. Hence the surface is in a higher energy state than the bulk of the fluid. The unbalanced molecular forces near the surface manifest themselves as surface tension or surface energy. The terms, used interchangeably, signify that any increase in surface area (any bringing of molecules to the surface) requires an expenditure of energy.

The second factor paramount in creating a foam is the presence of an impurity. A pure liquid cannot form a metastable foam; the foamability depends on a surfactant. In the case of an aqueous foam a surfactant molecule includes two chemical groups that differ greatly in their solubility in water. One group is hydrophilic, or highly soluble in water; the other is hydrophobic, or highly insoluble in water. Some typical hydrophilic groups are OH, CO_2Na and SO_3K; the typical hydrophobic group is the hydrocarbon chain C_nH_{2n+1}. A surfactant's surface activity depends on the sizes of these groups and on their relative solubilities. Because of the differing solubilities, surfactants are surface active, that is, they accumulate at surfaces. In particular, the hydrophobic groups on the surfactant molecules accumulate at an air-water interface. For surfactant molecules the surface is not as energetically unfavorable as it is for water molecules. Consequently the surface tension is reduced. Without this reduction the liquid could not be transformed from its bulk state, which has a minimal surface area, into a foam, whose myriad films give the liquid a far larger total surface.

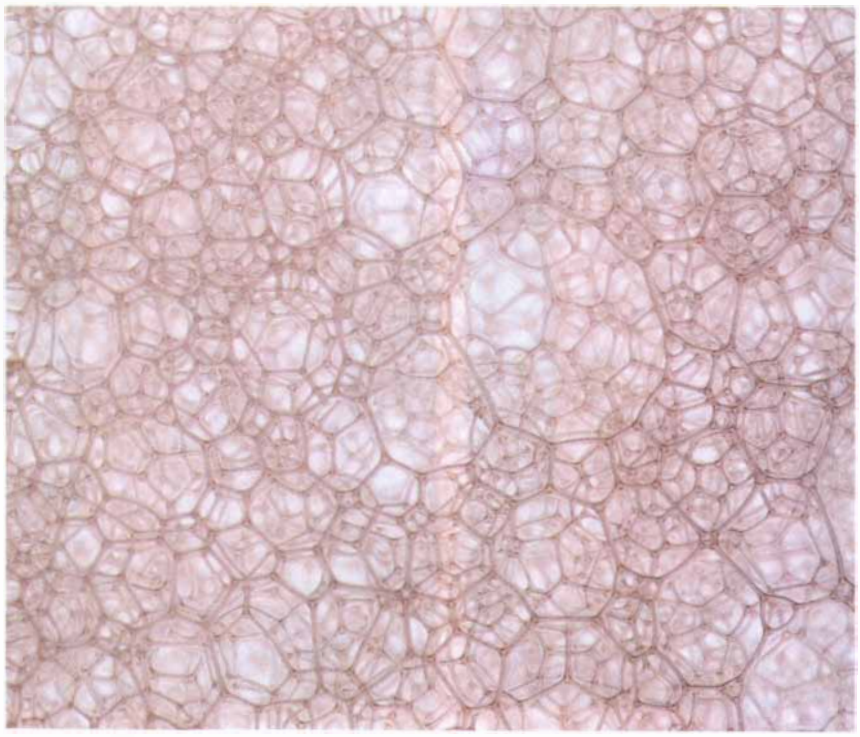
There is, however, a limit to the solubility of surfactant molecules, and once this limit is reached (in a bulk liquid or in an individual film in a foam) the surface tension can no longer be decreased because surfactant molecules added to the liquid can no longer simply go into solution. Instead the surfactant molecules begin to associate in curious ways. For one thing, dynamic micelles can form. These are subsurface associations of surfactant molecules in which the hydrophobic groups of a number of surfactant molecules cluster together, thereby minimizing their contact with the surrounding water. Surfactant molecules may also array themselves in a liquid-crystalline phase, in which the molecules are ordered rather rigidly in one



films are thick, the foam's content of liquid is great and the level of beer is correspondingly low. Like all other foams, however, the head is only metastable, or temporarily persistent: its films give it great surface area, which is thermodynamically unfavorable. Moreover, gravity draws liquid from the foam. If no other destabilizing influences came into play, the films might drain to a thickness at which they were stabilized against further thinning by electrical and chemical forces. In addition the molecules in very thin films can orient themselves in arrays that cannot be disassembled without an expenditure of energy. Yet other forces are acting. For example, the pressure in small bubbles exceeds the pressure in large ones, and so gas tends to diffuse from the small bubbles into the large ones. Some bubbles dwindle away, others coalesce. Meanwhile water is evaporating. The foam coarsens, and eventually it collapses.



SPHERICAL FOAM marks the early part of the life history of an aqueous foam. At this stage the foam is "wet": the liquid films that form the walls of each bubble are thick enough so that the bubbles are not distorted. Hence each bubble, independently of the others, minimizes its surface area for a given volume by taking a spherical shape. The foam was photographed in the head of a glass of beer less than a minute after the beer was poured.



POLYHEDRAL FOAM marks the middle age of an aqueous foam. At this stage the foam is "dry": the liquid films are thin and the bubbles interact, becoming polyhedral. The foam now manifests the geometric properties described by the Belgian physicist Joseph A. F. Plateau a century ago. For example, each film is a smooth surface, flat only if the pressures in two adjacent bubbles are equal. Moreover, the films meet only in sets of three and at angles of 120 degrees. The foam was photographed 10 minutes after the beer was poured.

direction but are free to flow or deform in other directions.

Foam Stability

Once a liquid has foamed, the question is whether the foam can persist for any appreciable length of time. For a physicist the capability of withstanding perturbations is proof that a system is stable. A foam is no exception. An example is afforded by a simple conceptual "foam" consisting of two bubbles connected by a tube. The surface tension in each bubble amounts to a force directed inward. (One can think of the force as arising from molecules at the surface of the bubble, attracted toward the interior because the surface is in a state of higher energy.) This tension is balanced by the pressure of the air trapped in the bubble. Since the bubbles are connected, their internal pressures are equal and their diameters are identical. Suppose the diameter of one of the bubbles fluctuates to an infinitesimally larger size while the other bubble shrinks by the corresponding amount. The smaller bubble now has a greater internal pressure, so that air flows from it to the larger bubble. The smaller bubble therefore decreases further in size. Its internal pressure rises again.

This destructive sequence establishes that the two-bubble "foam" is inherently unstable. Yet if surfactant molecules are present, the foam can be rendered stable. The stability derives from a pair of related phenomena described in the 19th century, one by the American physicist J. Willard Gibbs and the other by the Italian physicist C. G. M. Marangoni. The Gibbs effect occurs when a thin film is stretched and the film holds surfactant molecules in solution. The stretching increases the film's surface area, providing new opportunities for surfactant molecules to come to the surface. Nevertheless, the surface tension increases. At equilibrium (that is, when the proportions of surfactant molecules at the surface and in the bulk of the fluid are no longer changing) there are fewer surfactant molecules per unit area of the surface. The Gibbs effect dictates that a stretched film will "try" to contract, like an elastic skin; in fact, the Gibbs effect is often called Gibbs elasticity.

The second phenomenon, the Marangoni effect, is temporary. It arises because a certain amount of time is required for surfactant molecules to diffuse to the surface of a newly stretched film. Initially, therefore, the surface has a very low concentration of surfactants and the surface tension

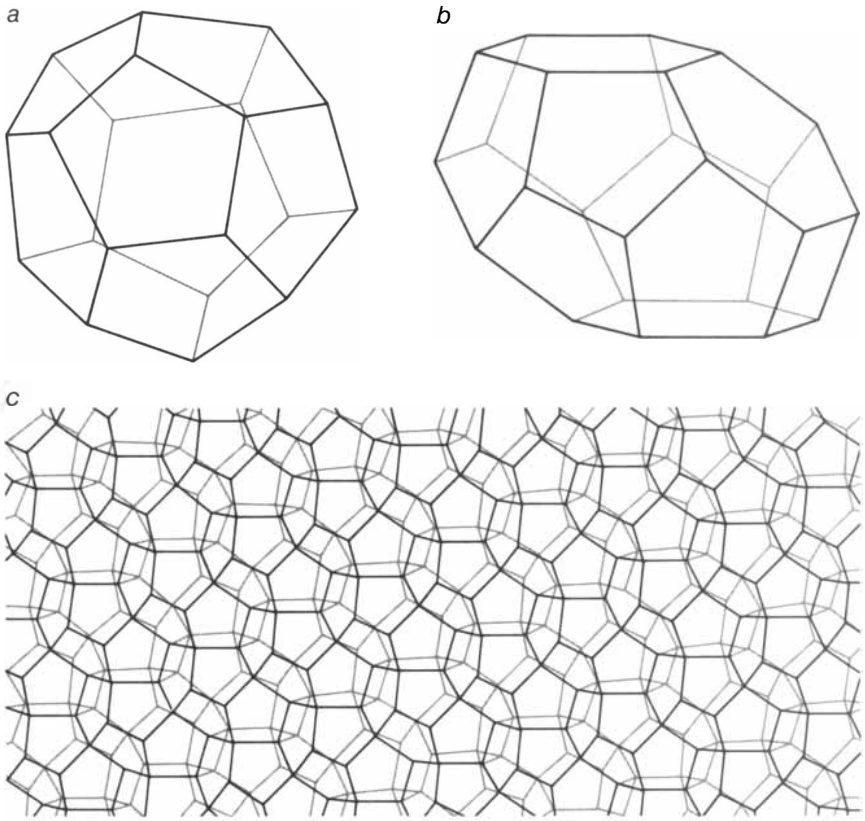
is even greater than a calculation of the magnitude of the Gibbs effect would lead one to predict. The tension slowly decreases to the Gibbs value as surfactants diffuse to the surface and the film equilibrates.

Acting together, the Gibbs effect and the Marangoni effect tend to stabilize infinitesimal fluctuations in foams. In our example of a two-bubble "foam" the combined effects increase the surface tension of the bubble that expanded, raising its internal pressure and forcing air back into the bubble that shrank. In general the Gibbs effect and the Marangoni effect stabilize the films in foams against infinitesimal fluctuations in thickness. If a fluctuation locally increases the area of a film, the surface tension will also increase, and this will stiffen the surface against growth of the fluctuation. The absence of Gibbs elasticity in a pure liquid is the main reason a pure liquid cannot foam: the films in the foam of a pure liquid are unstable to infinitesimal fluctuations.

Foam Morphology

A two-bubble "foam" is not, of course, an adequate representation of the structure of a real aqueous foam. In real foams the basic distinction is between "wet" foams and "dry" ones. In a wet foam the liquid content is high, and so the walls of the cells are thick. As a result the cells are far enough apart so that none of them is distorted by the others. Hence the cells are spherical, for the same reason that water drops are spherical: surface tension causes isolated bubbles (or drops) to take a spherical shape, which minimizes their surface area. As liquid leaves the foam, so that the foam "dries," its geometry becomes more complex. The cells, now separated by thinner walls, begin to influence one another. Thus the spherical bubbles become polyhedrons. In fact, the now traditional term for dry foams, introduced in 1953 by E. Manegold of the Technische Hochschule in Dresden, is *Polyederschaum:* polyhedral foam. (The traditional term for a wet foam is *Kugelschaum,* or spherical foam.)

The physical principles that determine the geometry of a "dry," polyhedral foam emerge from observations made by the Belgian physicist Joseph A. F. Plateau a century ago. First, the individual films forming the walls of the cells in the foam are smooth surfaces, each with a uniform curvature. A film is flat only when the pressure in two adjacent cells is equal; a curved film signifies a pressure difference between two adjacent cells. The pressure is greater in the cell on the concave side of the curve. (The application of this rule to a single spherical bubble dictates that the pressure must be greater inside the bubble.)

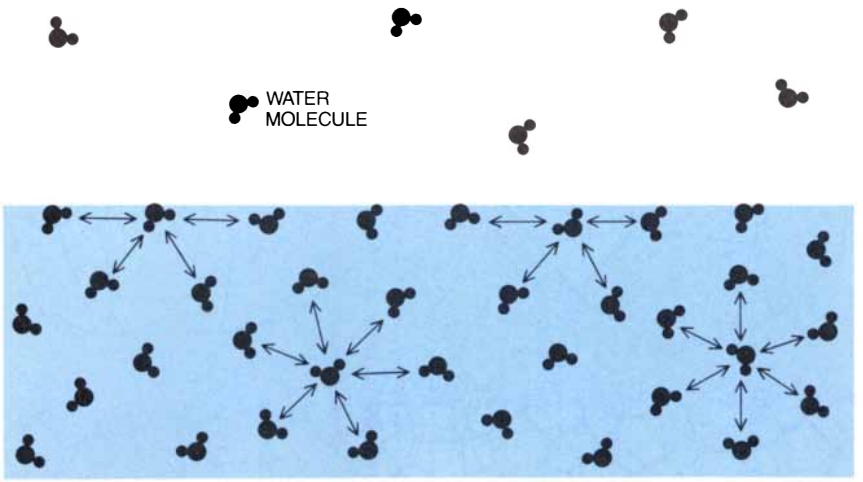


MATHEMATICAL MODEL of a polyhedral aqueous foam requires the packing of three-dimensional space with repetitions of a polyhedron. Among regular polyhedrons (ones whose faces are all the same regular polygon) the greatest success is the pentagonal dodecahedron (*a*). Its chief defects are that its packing leaves interstices and that the individual films in a real polyhedral foam are not all five-sided. A more realistic choice is an irregular polyhedron, the beta-tetrakaidecahedron (*b, c*). Although no model has yet succeeded in matching all the geometric properties of any real aqueous foam, the models do facilitate calculations of relations among liquid content, bubble sizes, film thicknesses and foam geometries.

Second, the liquid in a polyhedral foam is distributed between films and Plateau borders, the channels that form where films meet. The curvature of these borders indicates that the pressure inside a Plateau border is lower than the pressure in the cells and in the films. The resulting "border suction" is destabilizing: it tends to draw liquid out of the films. The eventual balance of forces is such that in many cases most of the liquid in a foam is in Plateau borders. When three films produce a Plateau border, the angles between them are equal, at 120 degrees, owing to a balance of surface-tension forces. A similar balance would require that four films meet at angles of 90 degrees, and so on. Yet the meeting of more than three films at a Plateau border is never observed in stable foams.

Complex mathematical arguments demonstrate that all Plateau borders must be three-film junctions. The arguments proceed from two starting points: the stability of a foam requires minimization of the energy of the system, and the various forces acting on any part of a stable foam must be in balance. Energy minimization and force balance also dictate a third principle governing the geometry of a polyhedral foam: the meeting of four Plateau borders must always have the form of a tetrahedral vertex, so that the angle between any two of the borders is the tetrahedral angle of about 109 degrees.

Instances of the principles that determine the geometry of a polyhedral foam are easy to find in a foam, yet a mathematical description of a three-dimensional foam conforming to all three conditions has not been formulated. The filling of three-dimensional space with repetitions of a regular polyhedron is not a success. The 12-sided regular polyhedron called the pentagonal dodecahedron comes the closest; in fact, a packing of pentagonal dodecahedrons has served as a model in which the relations among liquid content, cell sizes, film thicknesses and



SURFACE TENSION, the property that makes a liquid resist conversion into a foam, results from forces of electrical attraction among the molecules in the liquid (in this case water). Molecules deep in the liquid are attracted equally in all directions by fluctuations in the distribution of charge in each neighbor molecule. In contrast, molecules close to the surface are attracted most strongly toward deeper parts of the liquid. Hence moving a molecule to the surface requires the expenditure of energy. The result is surface tension, which signifies the tendency of a quantity of liquid to minimize the extent of its surface.

the curvature of Plateau borders have been studied.

The model's chief shortcoming is that a real foam does not have only five-sided films resembling the faces of a pentagonal dodecahedron. A nonregular polyhedron called the beta-tetrakaidecahedron more closely approximates the range of film shapes actually found in a foam; even so, its faces would have to be distorted to satisfy all three principles. Fortunately the theoretical calculations of many properties of a foam are not very sensitive to the precise geometry of the cells in a mathematical model of a foam. Hence the mathematical relations based on simple geometries can serve as useful approximations of the elusive reality.

Forces in Thin Films

Up to this point we have discussed aqueous foams as static structures. Yet a foam has a life history, consisting of the events that drain fluid from the foam, converting it from a wet foam into a dry one, and the events that sustain the structure of the dry foam, if only for a time. The main actors in this history are the films making up the foam.

When a film forms (either individually or in a foam), it is relatively thick. It loses liquid by drainage due to border suction. At some point, if the film survives long enough, the film drains to a thinness at which its two surfaces begin to interact. Until that point the film has been much thicker than the size of the molecules composing it and could be modeled as a bulk fluid with two surfaces independent of each other. Now, however, each surface might be said to "realize" that another surface is nearby. Hence the model must take account of new phenomena.

Several forces that act on a thin film have been identified. They include (in addition to border suction) van der Waals forces, which tend to destabilize the film, and electrostatic and steric forces, which tend to stabilize it. The short-range van der Waals forces between uncharged molecules are the cause of surface tension. The overall van der Waals force between two bodies composed of many molecules can also be significant. The overall force causes the two surfaces of a thin film to attract each other; thus the forces are destabilizing because they squeeze the film, causing it to drain.

Electrostatic forces can occur in a thin film if the liquid holds charged molecules in solution and one type of charge, positive or negative, is preferentially situated at the surfaces. Each charged surface then has a layer of countercharges subjacent to it. If the countercharge layers approach each other, as they would in a draining film, they would repel each other, inhibiting further drainage. Hence the electrostatic forces are stabilizing.

Finally, steric forces originate from the order the surfaces impose on molecules or structures within the film. They are not well understood; still, it seems clear that proximity to a surface can order molecules. For example, it is known that liquid crystals sometimes can form in a thin film even if they cannot form in a large volume of the fluid. Layered liquid crystals are often important in thin films in their final stages of drainage. The very last stage can be what is called a black film: a film thinner than any wavelength of visible light. Black films are thought to consist of surfactant molecules arranged by the surfaces of the film into a bilayered liquid-crystalline structure. The advent of order decreases the energy of a system, and so steric forces in general are stabilizing.

If the forces opposed to continued thinning are sufficiently strong, the film may drain to a thickness at which it is metastable. In this state the thermodynamic forces acting on the film sum to zero and a further decrease in the thickness of the film would raise the system's energy. The film may then persist until mechanisms other than drainage come into play and cause the film to collapse. Three such mechanisms are the evaporation of liquid from the foam, the diffusion of gas from small bubbles to larger ones and mechanical vibration. Isolated films protected from evaporation and vibration have been preserved for years.

The Yield Stress of a Foam

The life history of an aqueous foam neglects one important aspect of aqueous foams: their mechanical properties, which are remarkable, considering that the constituents of the foam (a gas and a dilute surfactant solution) behave individually as unexceptional and quite predictable fluids. A foam can flow like a liquid or remain motionless under stress like a solid. For example, shaving-cream lather is easy to spread; it offers little resistance to flow and gives the impression of having a low viscosity. Yet the stress due to gravity is not enough to cause shaving-cream lather to flow in the hand or when it has been spread; the manifestation of what seems to be low viscosity depends on whether a parameter called yield stress has been exceeded.

The physical basis for the yield stress of a foam can be explained by a simple model in which a foam is treated as a two-dimensional honeycomb of hexagonal cells. When the cells are all the same size, the model satisfies the Plateau criteria for a real foam. Suppose the foam is subjected to a shearing force; its deformation under that force is straightforward as long as the volume of each cell can stay constant and films can continue to meet in threes at angles of 120 degrees. Then a critical strain is reached, when the cells are so deformed that some of the Plateau borders have coalesced and the films between them have disappeared.

Four films now meet at each junction—a violation of the rules. In this unstable configuration the foam has been forced to generate an energetically unfavorable amount of surface area.

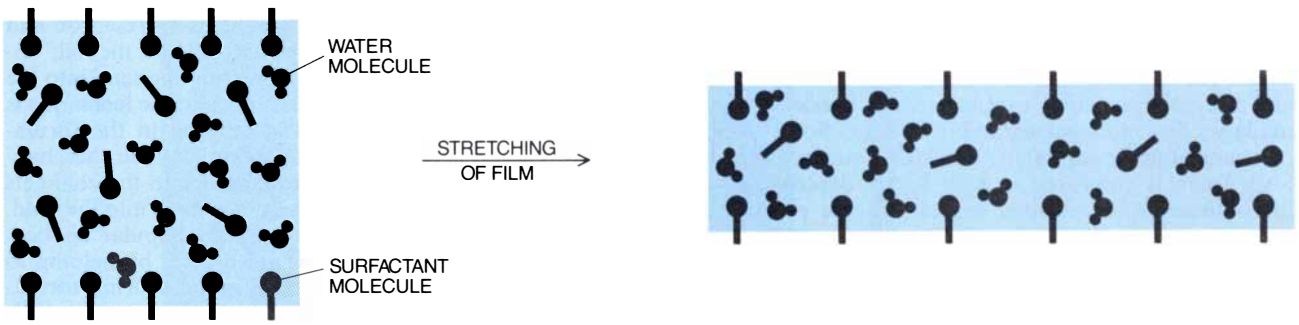
The structure of the foam thus makes a sudden transition to a new structure resembling the original honeycomb except that the cells are rearranged: some cells that originally were in contact have separated. This rearrangement constitutes the flow of the foam. In addition it defines the yield stress. The yield stress is the amount of force per unit area that triggers a deformation in which cells jump past one another to take up new positions. For dry films the yield stress turns out to be proportional to the surface tension in the films and inversely proportional to the size of the cells. The latter finding should be no surprise: fine-celled shaving-cream lather is much stiffer than coarse-celled laundry soapsuds.

The Making of Foams

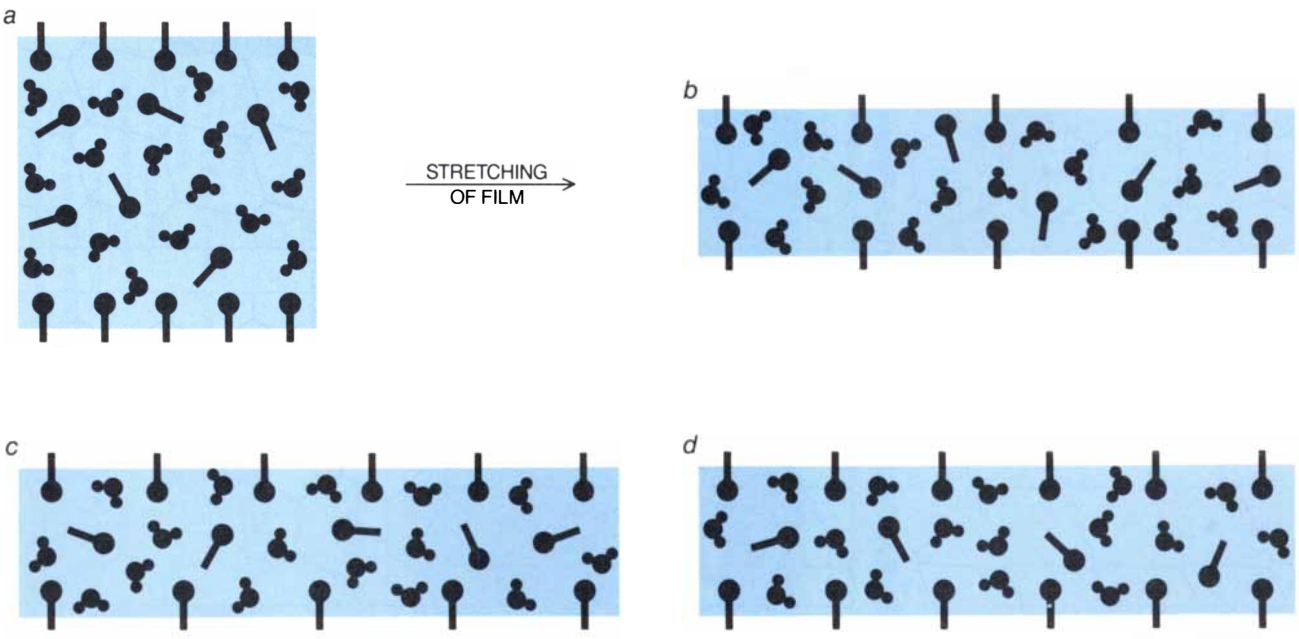
Perhaps the commonest, but also the most expensive, means of generating a foam is the method mimicked by opening a bottle of beer and pouring it into a glass, namely the aerosol method. It begins with a foamable liquid (water and a surfactant) and a foaming agent: a gas such as isobutane or a fluorocarbon compound, dissolved in the liquid at high pressure. (In beer and carbonated soft drinks the foaming agent is carbon dioxide.) The foam is produced by simply releasing the solution from its pressurized container. At atmospheric pressure the foaming agent returns to the gaseous state, thus forming the cells in the foam.

In another widely used method the foamable liquid and a gas are forced to flow through a packed column: a tube filled with obstacles that can be as simple as glass beads or a mass of stainless-steel wool. Small films form on the obstacles and the flow of gas distends the films into bubbles. The packed-column technique offers excellent control of the proportions of liquid and gas in the foam. A simplified version of the technique requires only that the liquid and gas be pumped into a pipe. If the pipe is long enough and the velocities are high enough, foam will emerge.

Low-density foams can be made by "mechanical" or by air-aspirated generators. In the mechanical generator the foamable liquid is sprayed on a perforated metal screen while a fan blows on the screen. The process is much like the familiar blowing of bub-



GIBBS EFFECT arises when a film is stretched and the film holds surfactant, or surface-active, molecules in solution. The stretching provides new surface area, which surfactant molecules can occupy, displacing water and reducing the surface tension; and yet when their migration is complete, the number of surfactant molecules per unit area of the surface is less than it was. Thus the surface tension actually increases, and the film, like a stretched elastic sheet, "tries" to shrink. The Gibbs effect is also called Gibbs elasticity.



MARANGONI EFFECT arises because surfactant molecules take time to diffuse to the surface of a newly stretched film. Hence the surface tension of a newly stretched film is even greater, temporarily, than the value due to the Gibbs effect. Together the Gibbs effect and the Marangoni effect ensure that if a fluctuation slightly increases the area of a film in a foam, its surface tension will rise, stiffening the film against any further increase. Pure liquids cannot foam primarily because they are not stable to such fluctuations.

bles through a hoop. Foams with a composition of one part liquid (by volume) to 1,000 parts gas can readily be produced. The air-aspirated generator is similar except that it has no fan; the flow of air required to make the foam is induced by the spray of the liquid.

Low density is perhaps the most important property of a foam. In fire fighting, for example, two types of low-density aqueous foam are in common use. Low-expansion foams, which have a density of between .1 and .2 gram per cubic centimeter, are made by fitting the fire hose with an air-aspirating nozzle (a nozzle that draws air into the emerging stream of water) and adding surfactants to the water. The resulting foam markedly increases the effective volume of the water. Moreover, by reducing surface tension the surfactant increases the water's ability to wet other surfaces (instead of, say, beading up).

The other type of fire-fighting foam, a medium- or high-expansion foam, which has a density of between .002 and .05 gram per cubic centimeter, is made not by a modified nozzle but by a mechanical or air-aspirated generator. Such foam is employed to fight fires in confined or inaccessible spaces. It is particularly effective in ships and mines because a person can easily breathe even when covered with the foam. Since the foam contains little water, it can put out fires in buildings without causing the water damage characteristic of traditional fire fighting. Buildings that store paper or house computers are therefore among the best candidates for sophisticated fire-detection systems that release high-expansion foam. At present the most important use of fire-fighting foams is probably for fires in flammable liquids. Indeed, only foams have been proved to be successful for dealing with fires in large flammable-liquid storage tanks. The foam is often pumped into the bottom of the tank; it floats to the surface and spreads.

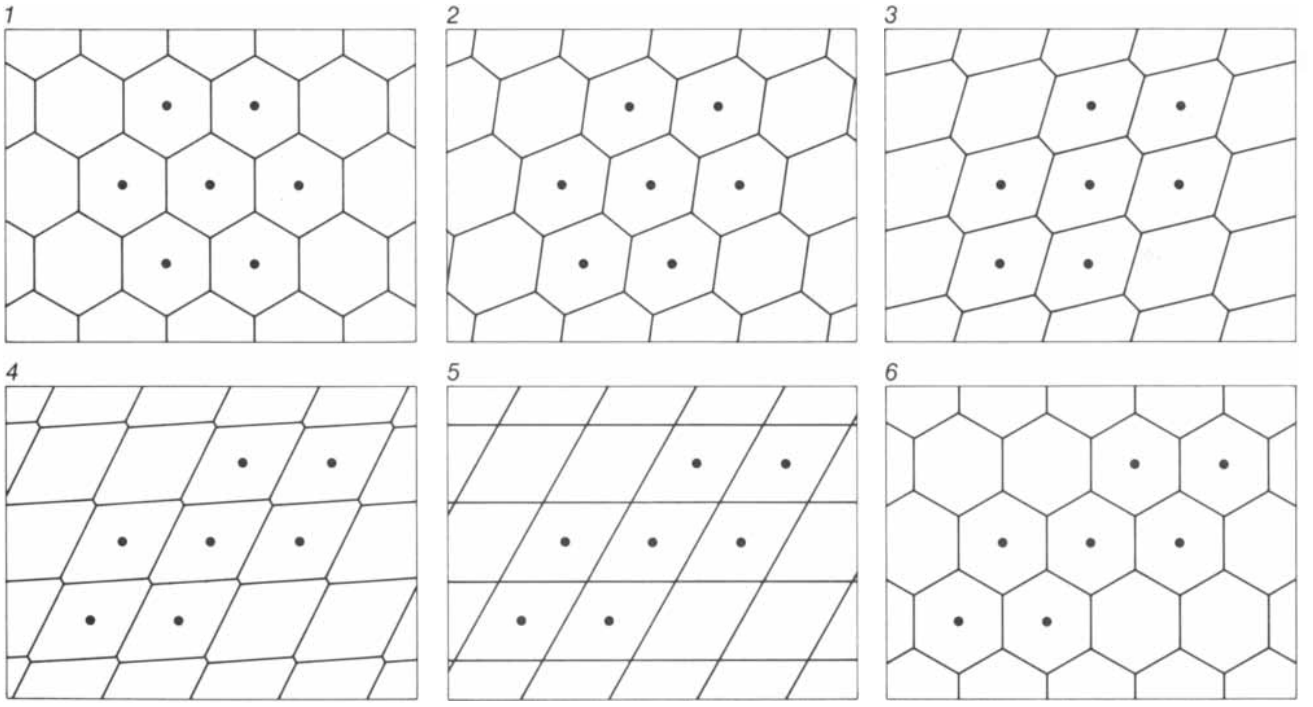
Foams for Resource Recovery

Drillers of oil and gas wells find foams useful in several ways. For one thing, deep wells often contain water, which must be removed before the well can begin to yield energy resources. If a suitable surfactant is mixed with the water and a gas is released in bubbles at the bottom of the well, the water can be converted into foam and carried out of the well.

Some wells penetrate "underpressurized" geologic formations, such as depleted gas reservoirs. In these wells the pressure in the formation is less than a reference pressure called the hydrostatic head: the pressure of a column of water whose height equals the depth of the well. The general problem in drilling is to achieve a balance between the pressure in the formation and the weight of the drilling fluid (typically a mud). An imbalance one way allows the fluid to contaminate the formation; an imbalance the other way precipitates a blowout. In the case of an underpressurized formation the drilling fluid must be lighter than water, and so a foam is appropriate. By controlling the density of the foam and the injection pressure (which is independent of the density) the pressure at the bottom of the well can be carefully governed. The foam serves all the functions of a drilling fluid. It removes drilling debris from the well; it also can bring to the surface uncontaminated samples of the formation at the bottom of very deep depleted-gas wells.

The high viscosity of foams makes them useful to drillers as a so-called diverting agent. In tar sands (a type of oil-bearing reservoir) oil is often recovered by injecting steam into the well. The steam heats the reservoir and reduces the viscosity of the oil, enabling it to flow from the sand into the well. The drawback to the technique is that channels develop in the formation, so that steam stays in the channels. Foam injected into the channels can force the steam back into the sand.

In an increasingly popular application, foams aid drillers by helping to fracture a gas- or oil-bearing formation. It is common practice in the drilling industry to pressurize a well and



FLOW OF A FOAM is actually a rearrangement in which bubbles jump past one another to take up new positions. Here a foam is represented by a two-dimensional honeycomb array of hexagonal cells (*1*). Under the influence of a shearing force the array is deformed (*2–4*); nevertheless, the cells retain their net size and films continue to meet in threes at angles of 120 degrees. Then, at a force corresponding to what is called the yield stress, films begin to meet in fours (*5*). The situation is unstable: the foam immediately transforms itself into a new honeycomb (*6*). Dots in a constellation of cells show how the relative positions of those cells have changed.

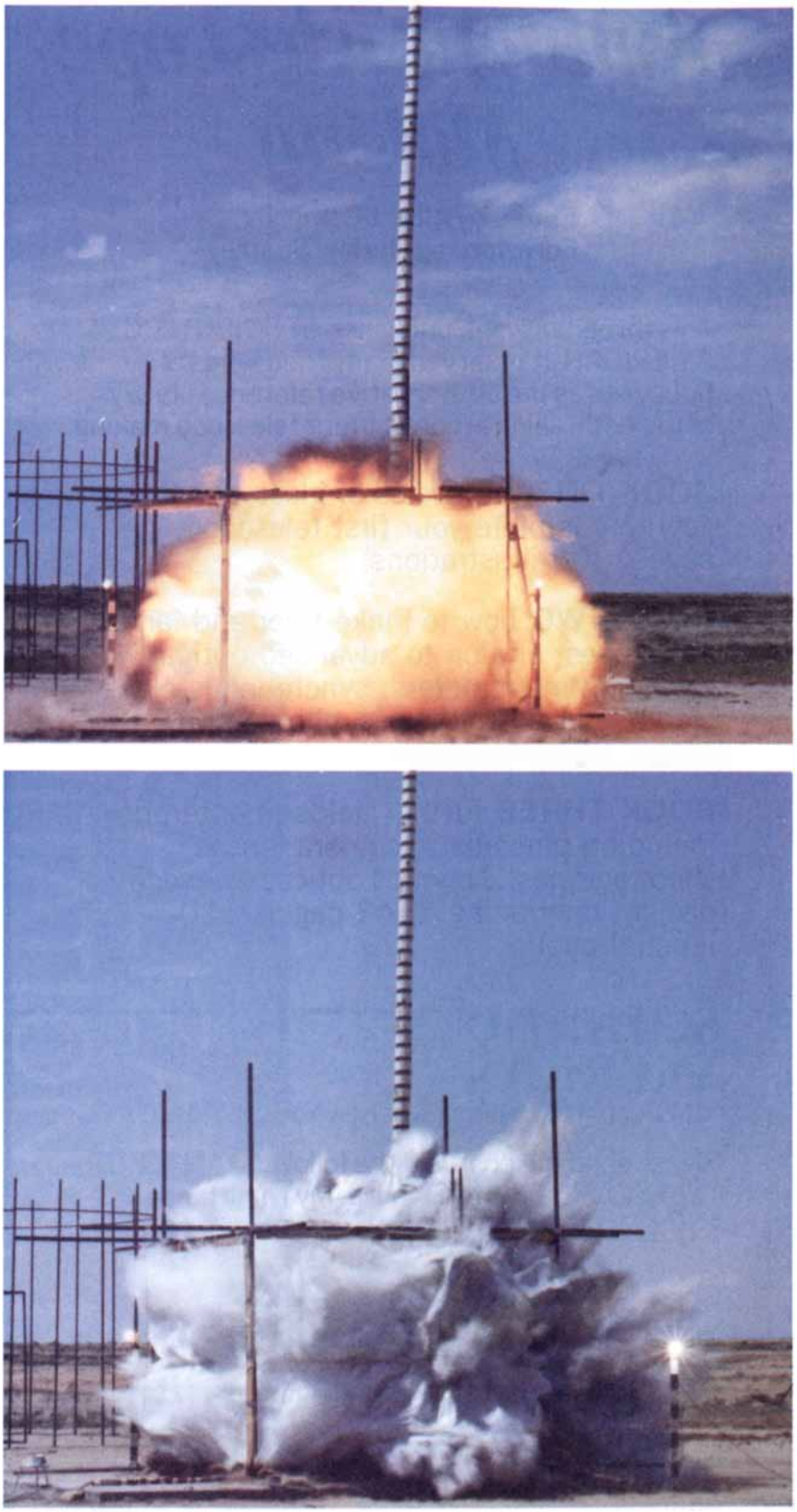
thereby make cracks in a formation, enhancing (the driller hopes) the flow of hydrocarbons. Often sand is added to the pressurizing fluid as a "proppant" to fill the fractures and keep them open after they form. When a foam is the pressurizing fluid, the settling of the proppant is reduced, which increases the chance that the proppant will distribute itself where it is needed. The small amount of liquid carried into the formation by a foam minimizes the contamination of the formation and eases cleanup when the pressure is released.

Foams as Traps

Another valuable property of aqueous foams is their ability to immobilize, that is, to contain or trap, all three states of matter: gases, liquids and solids. On the one hand, the foam can hold a desirable material. For instance, foams can distribute a material—even a small quantity of material—uniformly. Household examples include hair-setting foams and spermacidal contraceptive foams. The fabric-treating industry makes wide use of foams to distribute finishing agents such as dyes or resins uniformly. One advantage is that a foam minimizes the amount of water absorbed by the fabric. The drying of wet fabric requires time and energy.

On the other hand, the foam can blanket or trap an undesirable material. An obvious example is the trapping of a toxic material or a flammable liquid. The low density, metastability and trapping ability of a foam are all put to use in the cleanup of spills of radioactive material. The point here is to minimize the volume of contaminated material resulting from the cleanup. If a high-expansion aqueous foam is the cleanup agent, a large area can be decontaminated and yet when the foam collapses, only a small volume of contaminated material remains. In some applications the foam both contains and distributes. Crop treatment is an example: a useful chemical agent is distributed by the foam and is then kept where it will be beneficial.

Physical and chemical separation processes now exploit the trapping ability of foam. An instance of a physical process is ore refining by means of froth flotation. The ore is ground into fine particles and suspended in a foamable solution. Mechanical agitation then raises a foam. The resulting ore separation relies on the difference in wettability between the metal-poor fraction of the ore and the metal-rich fraction. The metal-poor fraction is usually hydrophilic and therefore easily wetted, so that it tends to sink. The



STIFLING OF AN EXPLOSION is an unusual use of aqueous foams now being evaluated by bomb squads and fire departments around the world. For each of these photographs five pounds of the high explosive called C4 was detonated at the Sandia National Laboratories in Albuquerque, N.M. For the second detonation the explosive was covered with 800 cubic feet of a low-density (.005 gram per cubic centimeter) aqueous foam. The foam proved to lessen the severity of the explosion by approximately 90 percent. Part of the energy of the shock wave of the explosion was consumed in converting the foam into tiny water droplets; then a further part of the shock energy was dissipated by the evaporation of the droplets.

metal-rich fraction is hydrophobic and tends to be carried off in the foam. Approximately 10^{12} metric tons of ore per year is now processed worldwide by froth flotation. An instance of a chemical process is dye separation. The process requires a fractionation tower holding a low-density foam whose surfaces preferentially attract a particular dye.

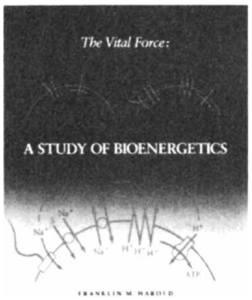
Prospective Uses

Several uses of foams are unusual. In a greenhouse designed at the University of Arizona a low-density aqueous foam helps to keep heat from escaping at night. Like most foams, aqueous foams are good thermal insulators. Each evening the foam is pumped between the two layers of translucent plastic that form the roof of the structure. In the morning the foam collapses under the heat of the sun and the liquid is collected, ready to be foamed again the next evening.

At a number of research institutions foams are being examined for their ability to decrease the pressure generated by an explosion. This dramatic ability arises from the destruction of the foam surrounding the explosive. In particular, some of the energy in the shock wave of the explosion is consumed in converting the foam into minute droplets of water; a further amount of energy is then consumed by the evaporation of these droplets. In all, a foam can consume as much as 90 percent of the pressure generated by an explosion. Bomb squads and fire departments around the world are evaluating the prospects. The foam can also act as a trap for particles propelled by the shock wave. In coal mines, for example, a foam covering the coal face being excavated by detonation reduces the pressure of the explosion in the direction of the mine workers and helps to trap coal dust, reducing the chance of a coal-dust explosion. In an extension of this technology, the U.S. Army is covering the nozzle of large guns with foam and testing the ability of foam to decrease the noise of detonations.

Aqueous foams have long been of scientific interest, at least in part because the complex chemical and physical phenomena in a form of matter notable for its great extent of surface area are well worth investigation. Certainly part of the interest comes from the fascination one finds in bubbles and films, which tend to be beautiful but evanescent. The latter property—the metastability of aqueous foams—is a clue that foams are unusual. The technological applications of foams show that the interest in foams is well justified.

JUST PUBLISHED...
from W. H. Freeman and Company



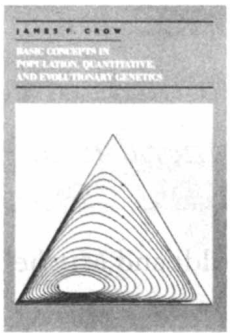
THE VITAL FORCE
A Study of Bioenergetics

Franklin M. Harold
National Jewish Center for Immunology and Respiratory Medicine and University of Colorado Health Sciences Center

"Bioenergetics is the scientific study of vitality, not a discipline so much as a point of view, complementary to that of molecular biology."
—from the Preface

Thanks to the advances made during the past quarter century, the principles that govern biological energy transduction are now fairly clear, but their significance to cellular and molecular biology is not widely appreciated. The object of *The Vital Force* is to make bioenergetics accessible to a wider audience by means of an integrated, comprehensible survey of energy transactions in metabolism, membrane transport, movement, and information processing.

1986, 557 pages, 286 illustrations

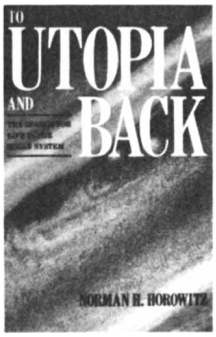


BASIC CONCEPTS IN POPULATION, QUANTITATIVE, AND EVOLUTIONARY GENETICS

James F. Crow
University of Wisconsin at Madison

This new text is a brilliant integration of classical population genetics theory and modern molecular ideas. Written by one of the premier scholars in the field, it covers significant topics of current interest such as molecular polymorphism and evolution. Emphasizing theory, but supplying enough experimental data to place the subject in a real-world context, *Basic Concepts in Population, Quantitative, and Evolutionary Genetics* makes population genetics accessible to students whose primary interest is biology, not mathematics.

1986, 304 pages, 48 illustrations

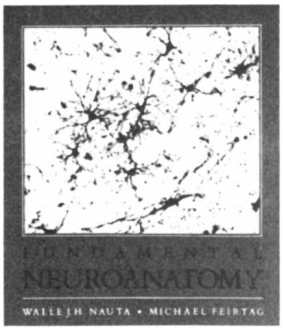


TO UTOPIA AND BACK
The Search for Life in the Solar System

Norman Horowitz
Professor Emeritus, California Institute of Technology

In this fascinating and highly readable work, the former head of the Jet Propulsion Laboratory's bioscience section for the six Mariner and Viking missions to Mars (1965–1976) provides a rare insider's view of the search for life on Mars and extends the lessons learned from the missions to the rest of the solar system. In addition, *To Utopia and Back* presents the author's first-hand knowledge and insights into the political atmosphere surrounding this scientific endeavor, some details of which are revealed in this book for the first time.

1986, 159 pages, 23 illustrations



FUNDAMENTAL NEUROANATOMY

Walle J. H. Nauta
Massachusetts Institute of Technology
Michael Feirtag
Board of Editors, *Scientific American*

Here, for the first time, is an account of the broadscale structure of the brain detailed and rigorous enough to serve the needs of medical students and researchers, yet accessible enough so that anyone who wants to learn about the brain can share in the modern understanding of the machineries underlying perception, thought, emotion, and behavior. The product of a decade-long collaboration between both authors, *Fundamental Neuroanatomy* is accompanied by more than one hundred dramatic photographs and drawings, many displayed full-page.

1985, 340 pages, 116 illustrations

Please send me

Quan.	ISBN	Title		Price
	1734-4	The Vital Force	(cloth)	$37.95
	1760-3	Basic Concepts in Population, Quantitative, and Evolutionary Genetics	(paper)	$15.95
	1766-2	To Utopia and Back	(paper)	$11.95
	1723-9	Fundamental Neuroanatomy	(paper)	$26.95
			Subtotal	$
		NY, UT, and CA residents add appropriate sales tax		$
		Add $1.50 postage and handling charge ($1.00 for each additional book)		$
			TOTAL	$

☐ I enclose a check made payable to W. H. Freeman and Company
☐ MasterCard ☐ VISA Account #_____
Expiration Date___/___ Signature_____
(CREDIT CARD ORDERS MUST BE SIGNED)

Name_____

Address_____

City_____

State_____ Zip_____

0227

Mail to: **W. H. FREEMAN AND COMPANY**
4419 West 1980 South
Salt Lake City, Utah 84104

The Cheetah in Genetic Peril

The world's fastest land animal is in a race for continued survival. An ancient population bottleneck has resulted in genetic uniformity and has made the species extremely vulnerable to ecological change

by Stephen J. O'Brien, David E. Wildt and Mitchell Bush

The cheetah, a virtual running machine, is a model of aerodynamic engineering. Its skull is small and lightweight and its limbs are long and slender, not unlike a greyhound's. Its heart, vascular system, lungs and adrenal glands are all enlarged, enhancing the animal's ability to accelerate and navigate during a high-speed chase. In addition the cheetah's claws are semiretractile: they are always extended like cleats, in contrast to the claws of other species in the cat family (*Felidae*), which are normally withdrawn and covered by a protective sheath. These various adaptations have made the cheetah a particularly effective hunter on the flat, open savannas of central and southern Africa, where it has a higher rate of successful kills than even the lion. After stalking its prey the cheetah launches a high-speed chase (often clocked at up to 70 miles per hour), pushes over or trips its winded victim and swiftly kills the prey by strangulation in its strong feline jaws.

In spite of the cheetah's impressive skill as a runner and hunter, the species seems to be heading for extinction. Where once the present-day species (*Acinonyx jubatus*) spanned the globe, now the cheetah's range is limited to a few small pockets in Africa. In all the world there are probably no more than 20,000 cheetahs.

How is one to explain the cheetah's march to extinction? Our investigations of the past five years suggest that the species has somehow lost its genetic variation. As a result of intensive inbreeding generations ago, each cheetah appears to be nearly identical with every other cheetah. Ever since Charles Darwin wrote *On the Origin of Species* a century ago it has been evident that genetic uniformity would hamper the ability of a species to adapt to such ecological perturbations as temperature shifts, drought, glaciation or the ascendance of new viruses or bacteria. Darwin's law of natural selection predicts that individuals well adapted to an environment will leave more offspring than less well-adapted individuals will. When a species has little genetic variety, its ranks are unlikely to contain many members whose genetically determined traits are well suited to withstand ecological change; the species competes poorly for survival under changed conditions and may die out.

As Darwin might have predicted, inbreeding has left the cheetah with traits that are as maladaptive as its lithe construction is adaptive. The species is vulnerable to disease and has an infant mortality rate that is estimated to be as high as 70 percent in some game preserves. In addition, although the cheetah is the world's fastest mammal, it can only run a few hundred yards. Having increased its respiratory rate from 60 to 150 breaths per minute during a typical chase, the animal often collapses for half an hour to regain its strength; during that time it can be attacked or, at the least, lose its bounty. Indeed, even when they are not exhausted, cheetahs are rather timid as defenders of their catch: fully 50 percent of their kills are snatched by more aggressive lions, leopards and hyenas.

Our investigation into the causes of the cheetah's decline began in 1981, after Frank Brand, director of the National Zoological Gardens of South Africa, invited us, as representatives of the U.S. National Zoological Park, to cooperate in studying a seemingly narrow problem: Why was it difficult to breed cheetahs in captivity? The early results of this cooperative effort yielded the first evidence that the cheetah might have a diminished gene pool and stimulated us to explore the cheetah's genetic status in detail.

Brand's problem was clear enough. In 1971 the Zoological Gardens had founded a cheetah-breeding program at the De Wildt Cheetah Breeding and Research Center, a compound near Pretoria. Ten years later the center led the world in numbers of offspring but was still frustrated by the animals' low fecundity and high infant mortality rate (37 percent).

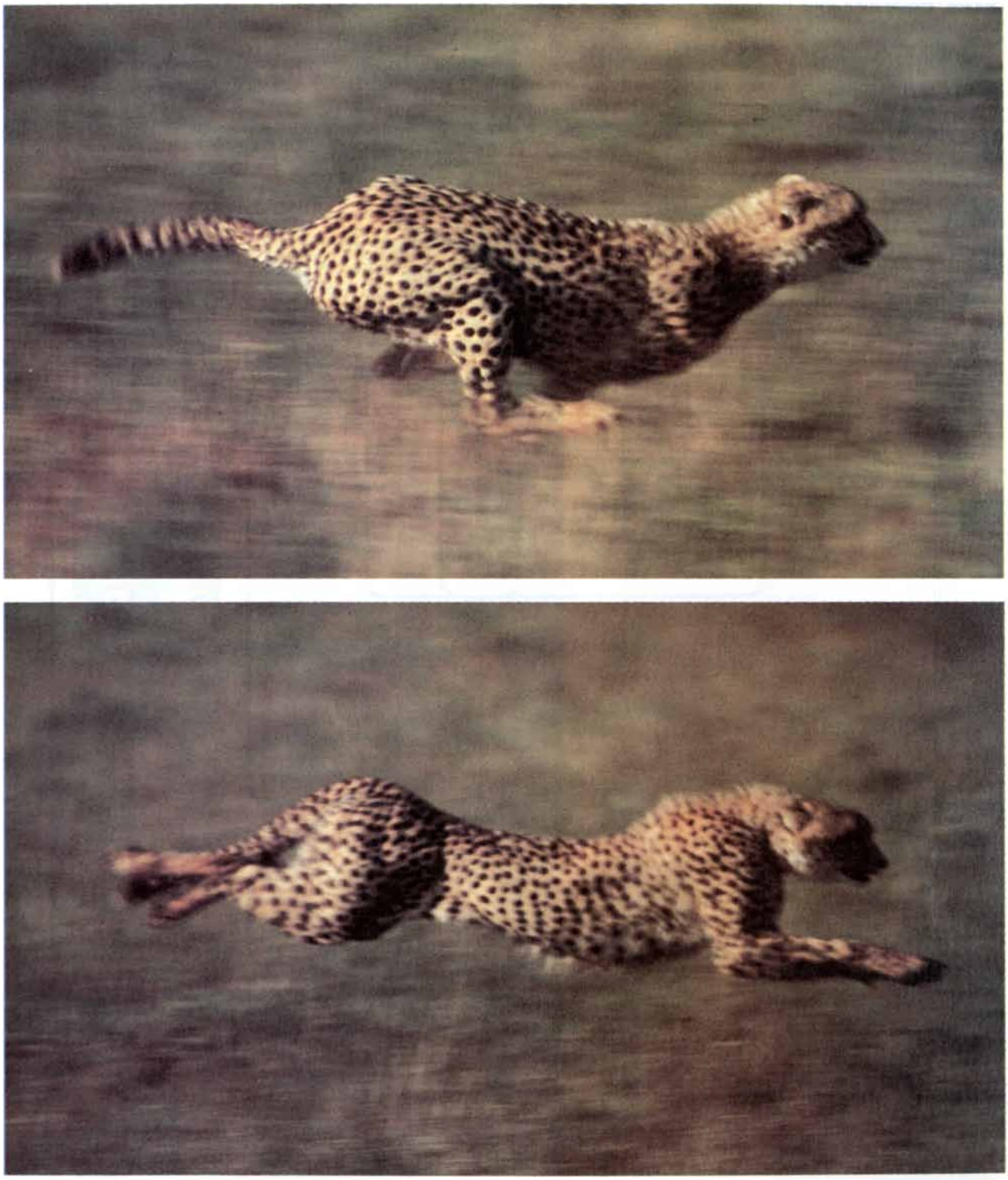
Indeed, the cheetah had a history of failure to breed in captivity dating back at least to the time of Akbar the Great, a 16th-century ruler in India. (Akbar, who had 1,000 cheetahs, was one of a long line of regal potentates on three continents who kept cheetahs as hunters and status symbols.) According to chronicles written by Akbar's son, the ruler had resorted to extreme efforts to promote breeding, including giving his regal specimens the run of the palace gardens. Even so, only a single litter was ever produced—and it was the sole recorded litter born to captive cheetahs until cubs were born at the Philadelphia Zoo in 1956.

Since 1956 a mere handful of breeding programs have been successful, and only from 10 to 15 percent of the sexually mature cheetahs caught in the wild have reproduced in captivity. Such low fecundity is often a consequence of unsuccessful mating attempts, but even after successful matings the cheetah has a low conception rate compared with that of other zoo-bred species, and about 30 percent of the cubs born in captivity die before the age of six months.

When we arrived at De Wildt to begin our diagnostic workup, including semen analyses and a study of genetic variety, the center had about 80 cheetahs. The colony had been established from two distinct regions of southern Africa: Namibia (South-West Africa) and the northern region of the Transvaal Province of the Republic of South Africa. The regions are separated by the Kalahari Desert, a

distance of 1,500 kilometers, which led the curators to conclude that the animals were representatives of separate geographic subspecies.

Our first step was to collect and analyze 40 semen samples from 18 males. The ejaculates were very different from those of other species we had studied. The concentration of sperm was only a tenth as high as the concentration usually seen in domestic cats, and the motility of the sperm was also significantly lower. Even more striking, sperm quality was invariably poor. In the average ejaculate some 71 percent of the sperm were shaped abnormally, in sharp contrast to the average of 29 percent found in domestic cats. For example, the flagella, which propel the sperm, were often coiled or bent at right angles, and many of the



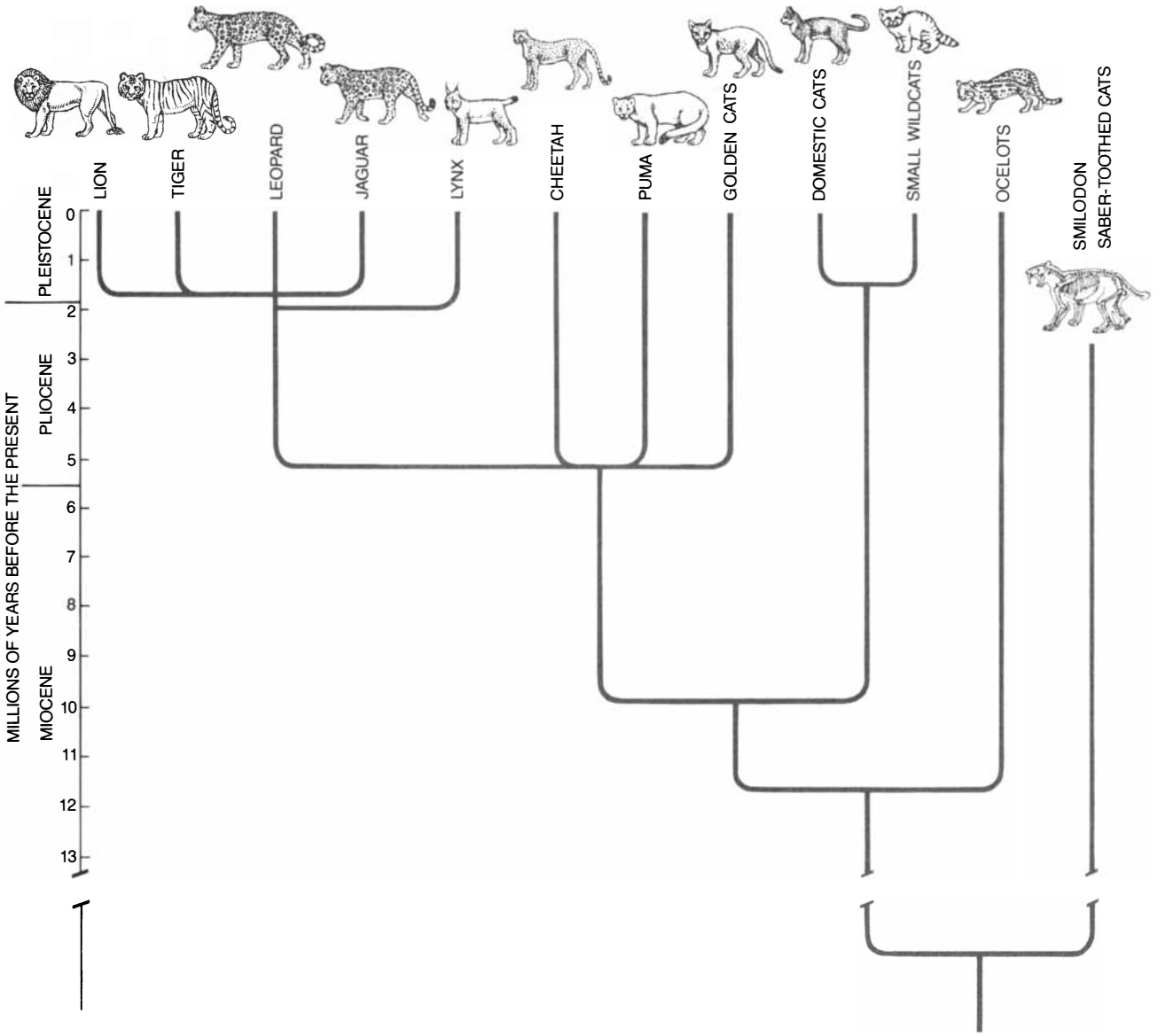
CHEETAH, the world's fastest sprinter, is shown in two of the positions it assumes while running at top speed. Often clocked at 70 miles per hour, the cheetah is airborne half of the time during a chase and can speed to 50 m.p.h. from a standing position in seconds.

sperm heads were either too large or too small.

The results of the semen studies seemed particularly significant because the vast majority of similar abnormalities (particularly at such a high incidence) in other species are associated with infertility. In the bull, for instance, the finding that 10 or 20 percent of the sperm are abnormal indicates that the animal is subfertile or even sterile. Furthermore, it seems clear that the quality of sperm morphology is under strict genetic control; increased morphological abnormalities often appear as a consequence of inbreeding of laboratory animals or livestock. The sperm data collected at De Wildt provided the first clue to the nature of the cheetah's plight.

To evaluate the extent of genetic variety in the De Wildt cheetahs we drew blood from 50 animals and arranged to have the blood-cell and plasma proteins—mostly enzymes—in the samples analyzed in the U.S. by gel electrophoresis, a standard procedure for the study of genetic variation. To make sense of the procedure one first needs to know a few basic facts about the mechanisms of genetic inheritance.

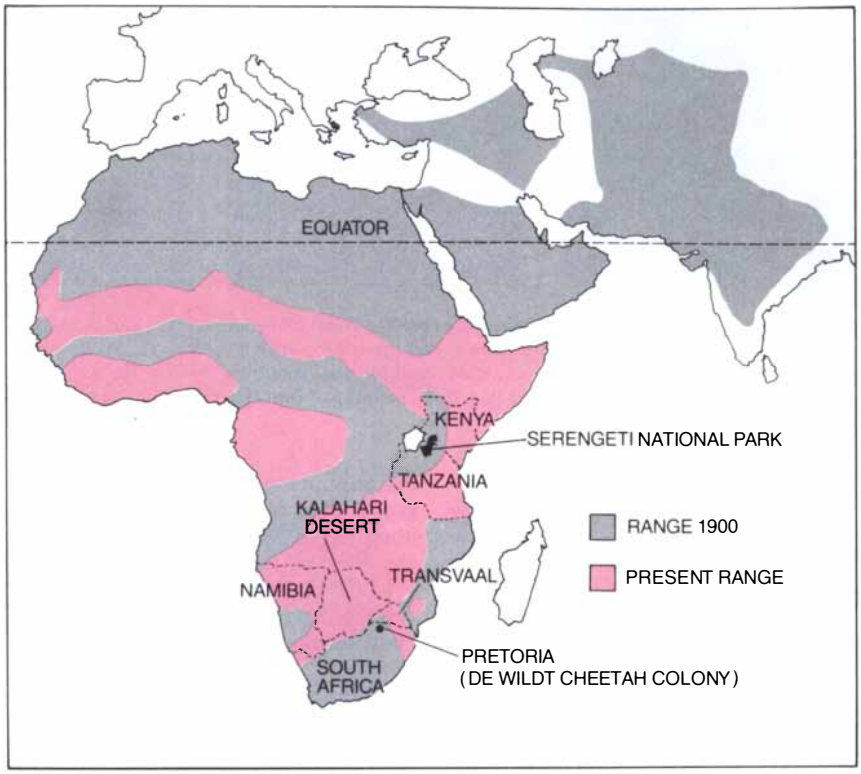
An animal receives a double set of chromosomes at conception, one set from the mother and an analogous set from the father. As a result each gene on a chromosome has an allele, or a mate, that resides on a paired chromosome and performs the same function as the first gene, such as directing the synthesis of a protein. Sometimes the alleles inherited from each parent at a given locus on the chromosome are homozygous, or identical, and sometimes they are heterozygous, or different. An individual that has heterozygous alleles for a given protein will produce two versions of that protein.



EVOLUTIONARY TREE of the cat family illustrates the relatedness of some of the 37 species that are in existence today. The fossil record shows that several families of saber-toothed cats came into being in the Miocene epoch (20 to 30 million years ago) and later died out, with one variety, Smilodon, persisting until 10,000 years ago. The modern felines began their radiation from other cats about 12 million years ago. The earliest branch led to the small and middle-size South American cats, including the ocelot and margay. The second branch began some eight to 10 million years ago and led to the domestic cat and its close relatives. The third branch began about four to six million years ago and led first to the pantherine lineage (including the golden cats, puma, cheetah, serval) and then to a split between the lynx and the modern big cats. This phylogeny is based on analyses of the protein albumin in modern cat species.

In carrying out a standard electrophoretic analysis of blood samples, workers isolate proteins from blood, place them on a gel and expose them to an electric field that causes the proteins to travel through the gel matrix. Proteins that are exactly alike (the products of identical genes) travel identical distances in the gel. Varieties of a given protein (the products of genes that encode proteins having an altered amino acid in the backbone) migrate to different positions. The proteins can be visualized with a specific stain, making it possible to compare the migration of the proteins and determine the degree of genetic variety within and between individuals [*see illustration on page 90*].

Our electrophoretic survey of the De Wildt cheetahs yielded a startling result. There was no variation in any of the 52 proteins examined; each cheetah was electrophoretically identical with every other cheetah. This was highly unusual. Similar electrophoretic surveys of genetic variation in more than 250 species over the preceding two decades had shown that between 10 and 60 percent of the genes in each species were polymorphic (had more than one electrophoretic form) and that between 1 and 36 percent of the genetic loci of the average individual were heterozygous. The genes of the cheetahs were monomorphic (had one form for each protein) and, it follows, were also homozygous at each protein locus. Such a striking level of genetic monomorphism is rare in natural populations. Like our semen analyses, it was reminiscent of the genetic homogeneity seen in purposely inbred species, such as certain laboratory mice.



RANGE OF CHEETAH is shown for today (*color*) and for the turn of the century (*shaded*). In 1900 the range of the present-day cheetah, *Acinonyx jubatus,* extended through Africa, the Middle East and India. Now the cheetah is extinct as a free-ranging species everywhere except in central and southern Africa. Estimates of the number of cheetahs alive today vary from 2,000 to 25,000. Fossils of *A. jubatus* have been found in Europe, Asia, Africa and North America. Fossils of at least three other, extinct cheetah species have also been found: *A. pardinensis,* giant cheetahs that lived perhaps four million years ago in China, India and southern Europe; *A. intermedius,* a smaller species that ranged throughout Eurasia, and *A. trumani,* which had similarity to the puma and lived in the U.S.

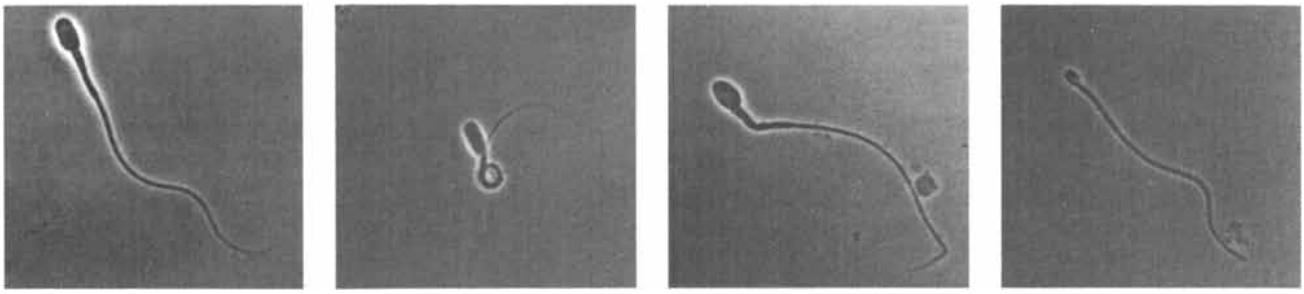
In our increasingly refined search for some indication of genetic variety, we took a somewhat different approach to protein analysis: two-dimensional gel electrophoresis. Like the technique employed in our enzyme surveys, this approach also analyzes protein migration in a gel, but it can evaluate hundreds of proteins simultaneously. We asked David Goldman and Carl R. Merril of the National Institute of Mental Health to examine proteins from the fibroblasts (connective-tissue cells) of six unrelated southern African cheetahs in U.S. and European zoos. Although these cheetahs did have a few variant proteins, the frequency of polymorphism among 155 proteins analyzed was found to be exceedingly low: 3 percent, or less than a third of what is usually seen in human populations that have been studied by the same technique.

The emerging profile of an inbred species in unusual genetic peril was corroborated by three additional studies, beginning with a comparison of cheetahs and other cats. Andrea Newman, a graduate student in the laboratory of one of us (O'Brien) at the National Cancer Institute (NCI), did a comprehensive electrophoretic survey of blood-cell enzymes from nine cat species housed in zoos and wildlife preserves throughout the world. Her results revealed moderate to high levels of genetic variety within each species studied, including some whose ranges overlap that of the African cheetah (the leopard, lion, serval and caracal) and some whose ranges do not (the tiger, ocelot, margay and domestic cat). In other words, the cheetahs we had studied were exceptional even among fellow felines in their diminished level of genetic variation.

The second set of data was obtained from an analysis of morphological traits in a group of African cheetah skulls collected in American museums. We estimated morphological variation on the basis of a measurement called fluctuating asymmetry: the extent to which features that are normally mirror images of each other, such as the left and right sides of the skull, are different sizes in a given individual. In a variety of species asymmetry has been found to increase in degree and incidence as a result of inbreeding, although just how the asymmetry comes about is not exactly clear.

Robert K. Wayne, a postdoctoral fellow in our laboratory at the NCI, evaluated fluctuating asymmetry in each of 33 museum-held skulls and compared the results with measurements from leopards, margays and ocelots, which are species we knew (from Newman's results) have abundant genetic variety. The skulls of the cheetahs did indeed display a higher level of asymmetry than the skulls of the other three cat species.

Our third—and most alarming—set of corroborative data came from a study designed to determine whether the cheetah displays any variability at the major histocompatibility complex (MHC). The MHC is a complex genetic locus in mammals that directs the synthesis of antigens on the surface



ABERRANT SPERM detected in animals at the De Wildt Cheetah Breeding and Research Center are shown, along with a normally shaped specimen (*far left*). Among the many striking abnormalities found were (*left to right*) coiling of the flagellum, bending of the midpiece and reduction in head size. The discovery that the cheetahs at the De Wildt center had a high incidence of abnormal sperm provided the first major clue that the cheetah's reproductive problems are physiological in nature and might have a genetic explanation.

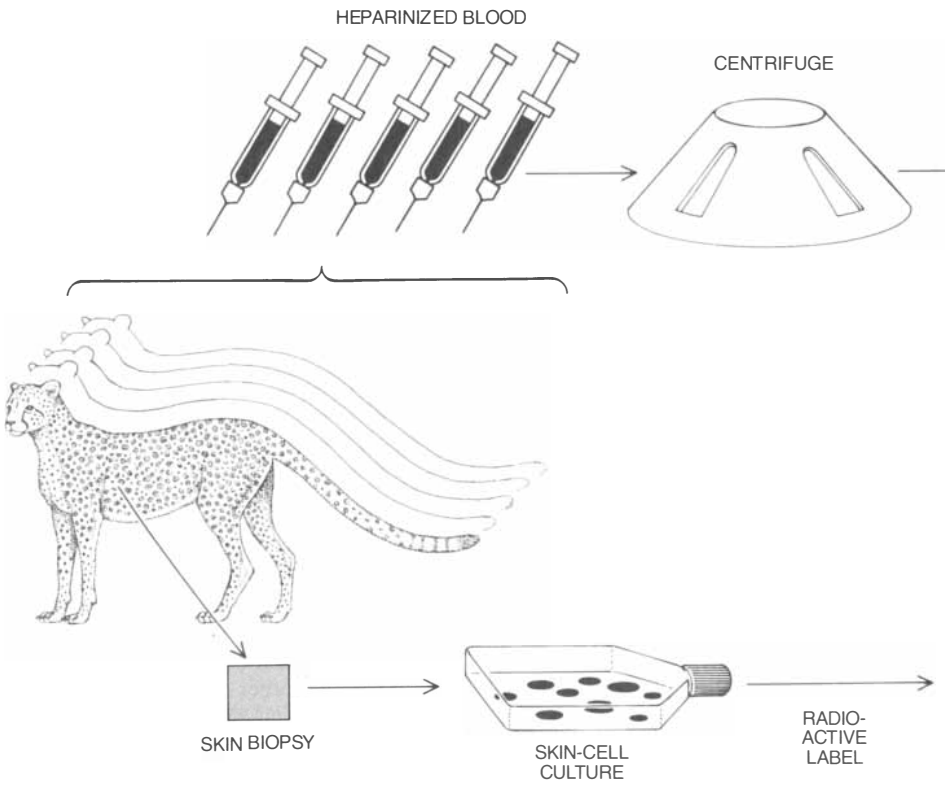
of most cells. The antigens communicate with *T* lymphocytes (circulating white blood cells) during an immune response to an infectious agent or to spontaneously arising aberrant cells, such as tumors. MHC antigens are also responsible for the immunological graft rejection that follows tissue or organ transplantation when the donor and the recipient have different MHC antigens. In the past 50 years the MHC system has been extensively studied in man and in the mouse, rat, dog, horse, chimpanzee and several other species.

The MHC is the most extensively polymorphic locus in mammals. The complex is really a chromosomal cluster having three functionally distinct groups of subloci, termed class I, II and III. In human beings the class I MHC sublocus alone is composed of three genes, each of which can be drawn from more than a dozen allelic varieties in the gene pool. As a result the number of possible combinations at the class I sublocus on one chromosome in humans is enormous (more than 12^3), and the chance of any two individuals having the same combination is slim (less than one in 10,000). Other mammalian species also display similar extreme polymorphism. The major exceptions are strains of inbred mice that are bred intentionally for homozygosity of their MHC genes and therefore accept skin grafts within but not between strains.

To assess the presence of polymorphism at the MHC in the cheetah, we exchanged skin grafts between the two members of seven pairs of cheetahs from southern Africa: six unrelated pairs and one sibling pair. A patch of skin was removed from each animal and cut in half. One piece was sutured into a graft bed in the paired recipient; it was an "allograft." The other piece was sutured into a graft bed in the animal from which it was removed; it was an "autograft." After the procedure the animals were examined and their grafts were cleaned twice a week for about eight weeks, during which time any signs of rejection were noted.

In several mammalian species, including domestic cats, the average survival time of skin grafts from an unrelated animal is some 10 to 12 days; rejection within this time is attributable to differences at the MHC. Slower, or chronic, rejection is the result of differences at other, minor histocompatibility loci in cases where there is identity at the MHC. Hence if a cheetah rejected an allograft within 10 or 12 days, we could conclude that the MHC genes of the two members of a pair were not identical.

Once again the cheetahs demonstrated their genetic uniformity. Remarkably, all allografts were accepted and indeed were indistinguishable from the autografts throughout the 10-



TWO ELECTROPHORETIC METHODS for measuring the extent of genetic variation in cheetahs are illustrated. In one method (*top*) fresh blood is treated with heparin to prevent clotting and allow different components (white cells, red cells and plasma) to be separated in a centrifuge. Soluble enzymes from the blood samples are then subjected to electrophoresis, that is, they are exposed to an electric field that causes them to migrate through a

to-12-day period. Three of the allografts did undergo slow rejection later, but several of the grafts persisted for at least 78 days, by which time they appeared to blend in with the recipient's own skin.

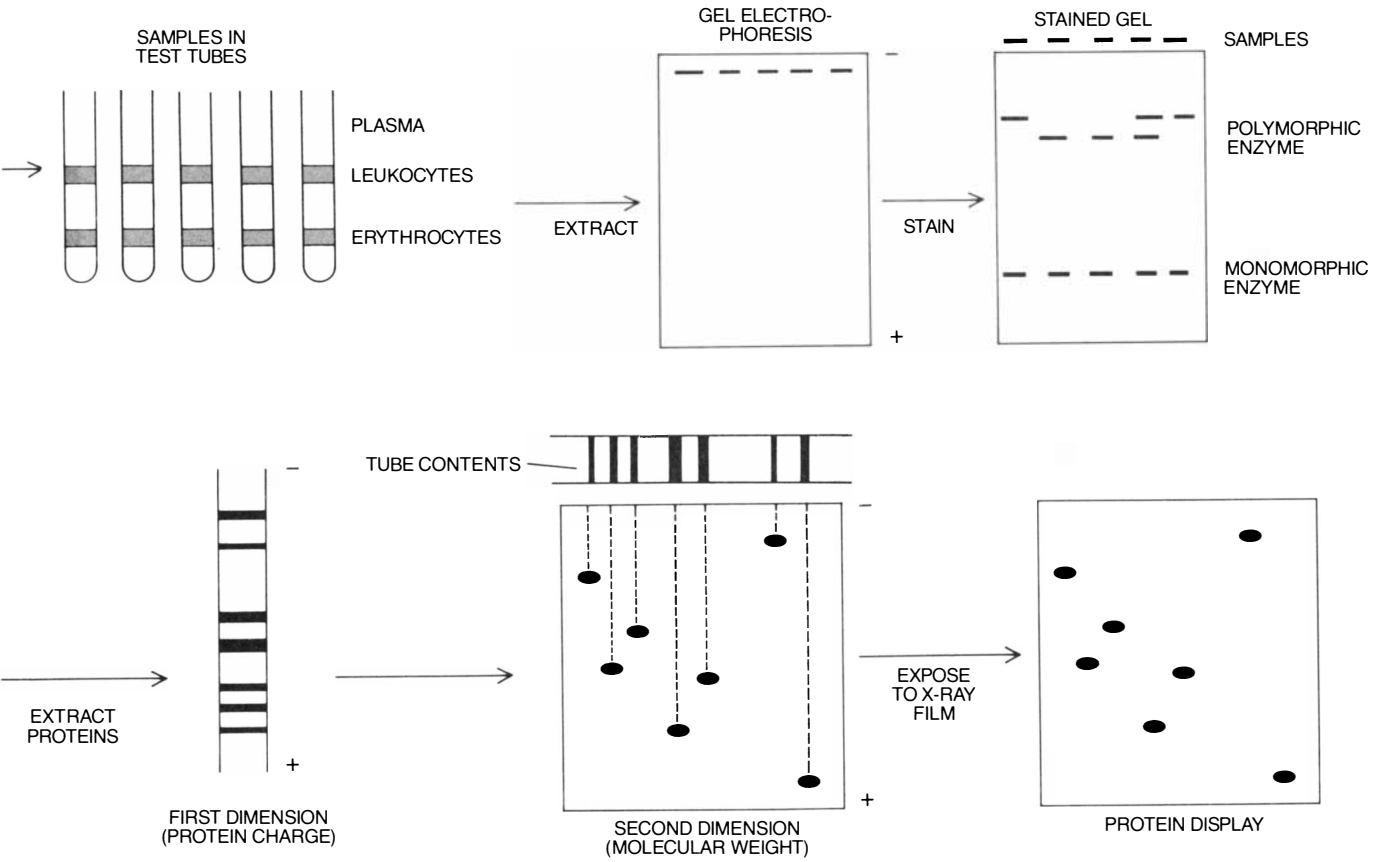
In order to be sure that the failure to reject the allografts was a result of identity at the MHC and not of a general failure of the cheetah's immune system to reject tissue grafts, we also sutured skin from a domestic cat into the graft bed of two of the cheetahs. In both cheetahs the graft from the cat underwent a classical acute rejection between days 10 and 12, whereas the autograft and the allograft continued to heal and grow. This indicated that the immune system of the cheetahs was indeed able to recognize antigens specified by foreign MHC genes. The cheetahs' immune system had simply not encountered any foreign antigens in the skin grafts from the purportedly unrelated cheetahs.

The cheetahs in these skin-graft experiments (and for that matter in all the studies we have described) demonstrate a level of genetic monomorphism that is unprecedented in any outbred mammalian species. Nevertheless, there is one possibility that must be considered before we can conclude definitively that cheetahs throughout Africa are virtual genetic twins and the products of inbreeding. The biochemical studies reported here all involved samples from cheetahs of southern Africa; most of the cheetah skulls that showed increased asymmetry (an indication of inbreeding) had been collected from eastern African populations by Theodore Roosevelt and his companions. There is a chance that the cheetahs in eastern and central Africa, although possibly inbred themselves, have gene pools somewhat different from those of the southern cheetahs. We are not optimistic that this is so, but it is a possibility we are exploring.

The fact that genetic uniformity poses a threat to the survival of a population or a species has been evident since Darwin formulated his theorem of natural selection. Genetic variation is the raw material for evolution; it is genetic heterogeneity on which natural selection operates in times of environmental or ecological change. The surprise, then, is that the monomorphic cheetah has persisted at all. It has not persisted without peril, however. Evolutionary theory predicts that a species with little genetic plasticity would be particularly vulnerable in a time of ecological perturbation, and a graphic demonstration that this is so for the cheetah occurred in 1982.

A pair of apparently healthy cheetahs arrived (on breeding loan) at the Wildlife Safari Park in Oregon, which had one of the most successful cheetah-breeding programs in the world. Within a few months the visiting cheetahs developed fever, diarrhea and jaundice, and they died a few weeks later. The cause of death was determined to be a viral infection called feline infectious peritonitis (FIP).

Although the disease spreads rampantly in colonies of domestic cats, FIP seldom kills more than 10 percent of infected animals. It was different for the cheetah. Within six months of the first two deaths at Safari Park, symptoms of FIP had developed in every cheetah in the park, and by the end of 1983 nearly half of the cheetahs had died of FIP-related disease. The epizo-



gel matrix, after which they are made visible by specific stains. Enzymes that are products of polymorphic genes migrate to different positions in the gel matrix, whereas enzymes that are products of monomorphic genes migrate to identical positions. In the second method (*bottom*) radioactively labeled proteins from skin cells that have been grown in culture are exposed to electric fields that separate them in two dimensions. The proteins are separated first on the basis of electric charge and then on the basis of molecular weight. Finally the gels are exposed to X-ray film, which reveals hundreds of proteins whose positions can be compared between cheetahs.

otic (an epidemic among animals) was the most extreme response to an FIP viral infection so far reported in any species. Since then we have learned of similar FIP epizootics at cheetah-breeding compounds in Ireland, Canada and Namibia.

One possible explanation for the Oregon epizootic was that the FIP virus was a particularly virulent strain and simply happened to strike the cheetahs first. This idea did not hold up. Several attempts to transmit FIP to domestic cats by inoculating them with virus collected from the cheetahs were unsuccessful. Moreover, 10 African lions in the Oregon compound remained free of symptoms after exposure to the virus. Such a lack of virulence in two feline groups indicated that the epizootic among the cheetahs was a result of cheetah vulnerability rather than of viral hypervirulence.

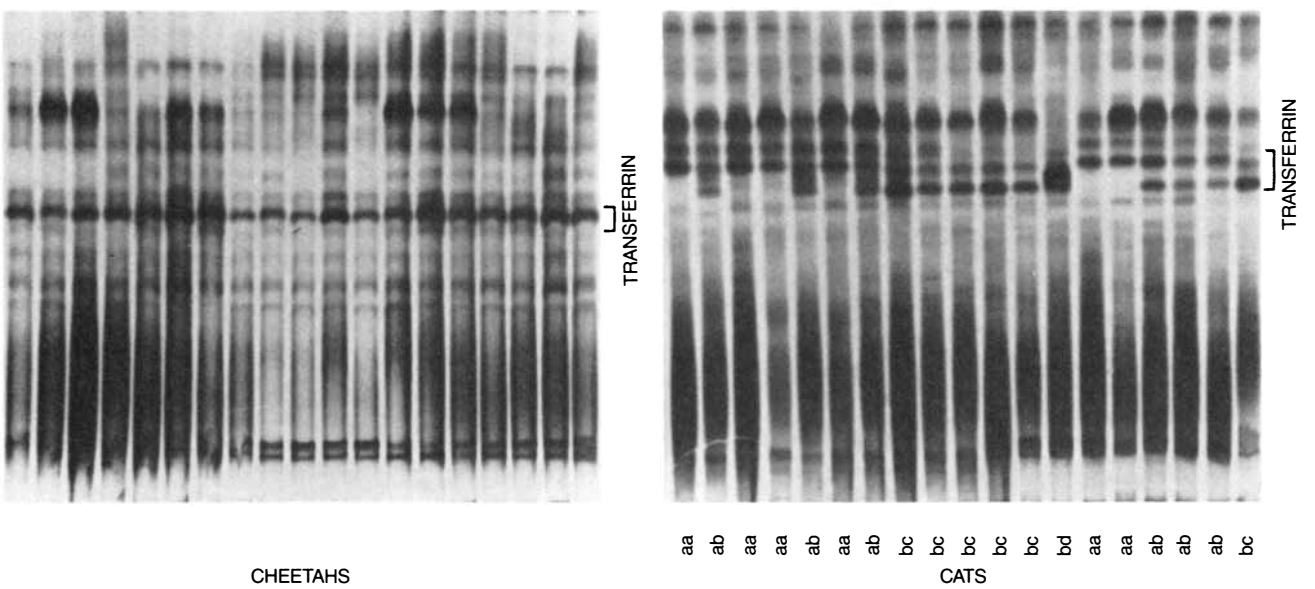
Monomorphism at the MHC, which would limit a species' repertory of defenses against a virus, is an attractive candidate for a biological explanation of the epizootic. In a monomorphic species one would expect widespread morbidity soon after a virus successfully overcomes one animal's defenses, because all the other animals would have much the same degree of susceptibility to the virus.

The event in Oregon could actually be explained by monomorphism of genes within either the class II or the class I MHC sublocus. The class II sublocus contains immune-response (IR) genes and is believed to contain viral-restriction genes also (although the two gene types may actually be the same entity). IR genes encode antigens that have been shown to elicit antibodies against synthetic antigens or viruses in laboratory animals. Viral-restriction genes have been shown in mice to determine the degree to which an organism can prevent the replication of viruses whose genetic material is RNA, and the FIP agent is an RNA virus. A population that is polymorphic at the MHC locus would be expected to have many varieties of IR and viral-restriction genes with the ability to confer protection against the FIP virus. On the other hand, a population that suddenly became monomorphic at the class II sublocus would be particularly vulnerable to a viral strain able to circumvent the immunological defenses that are controlled by the population's single, universally shared set of class II genes.

Monomorphism in the class I MHC sublocus would result in an epizootic by a different mechanism. Class I MHC genes encode antigens that must appear on the surface of a virus-infected cell before *T* lymphocytes can recognize viral antigens, which are also displayed on the cell surface, and kill the infected cell. Any virus able to alter itself in a way that prevents *T* lymphocytes from recognizing the class I MHC antigen on virus-infected cells would be effective against the host. In a polymorphic population any such adaptation to one host by a virus is unlikely to be effective against the immune system of a host having a different set of class I antigens. In a monomorphic population, however, any virus that adapts to one animal's immune-surveillance system would subsequently find every other system it encounters in the population to be identical to the first and therefore easy to get around.

It may be important to remember that exquisite strategies for abrogating immune surveillance are evolved by viruses in parallel with the defense systems of their hosts. Regardless of the exact explanation of the cheetah colony's reaction to the FIP virus, the episode in Oregon seems to emphasize the disadvantage of a genetically depauperate population in adapting to pathogens in its environment.

Having uncovered an array of evidence that the African cheetah is a genetically monomorphic species, we turned to the historical causes of the inbreeding that led to such monomorphism. Data relating to a single point in time in the dynamic evolution of a species do not give one much to go on. Nevertheless, we considered several hypotheses, each of which presumes that the ancestors of today's cheetah once had as much genetic variety as



ELECTROPHORETIC MIGRATION is shown for one protein, transferrin, in 19 plasma samples from cheetahs (*left*) and 19 samples from domestic cats (*right*). The transferrin that was synthesized by the cheetahs migrated to one position (*dark band across all samples*), indicating that all 19 cheetahs were monomorphic: they carried the identical transferrin gene. In contrast, the transferrin synthesized by the domestic cats migrated to several positions, revealing the presence of four different types of transferrin and indicating that the cat population was polymorphic: had more than one transferrin gene variant. (The letters at the bottom indicate the dual gene combinations expressed by each cat.) In the cheetah similar uniform results have been found for virtually all enzymes studied.

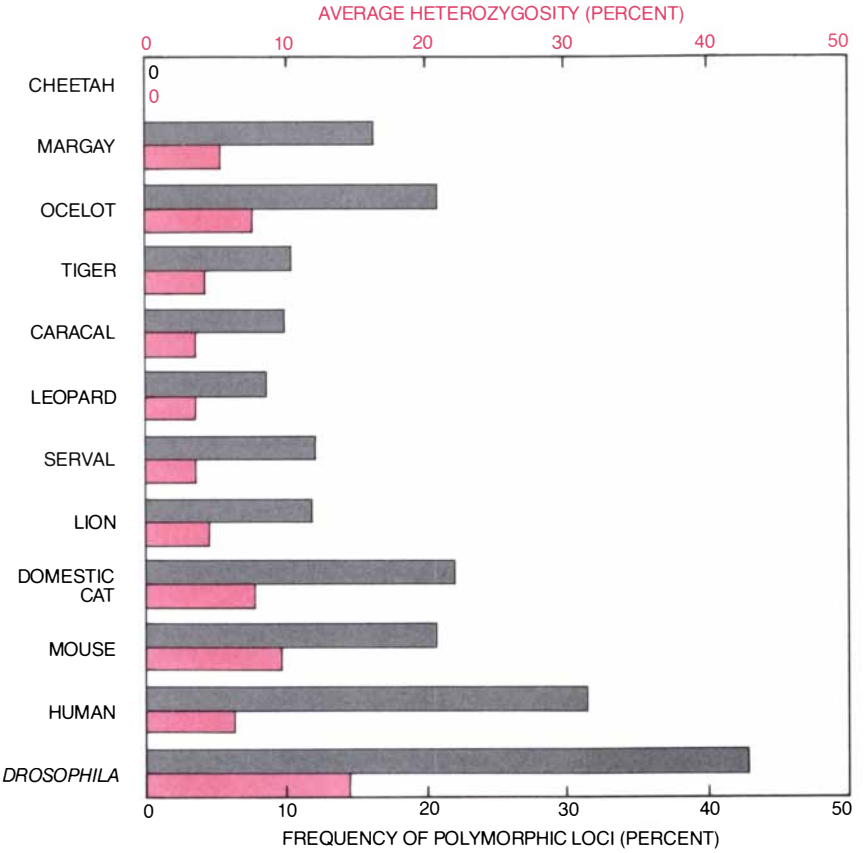
that currently enjoyed by other felines.

One hypothesis is that the cheetah is predisposed to incestuous mating, but that idea is not consistent with the findings of ethologists who have studied cheetahs of eastern and southern Africa. They find that cheetahs generally do not remain with their siblings after childhood. Male cheetahs are often territorial and occasionally nomadic, and unrelated females wander alone in and out of territories, mating with several males before moving on.

A second hypothesis is that the cheetah evolved to an adaptive optimum for a particular environmental niche and then gradually shed its variety during an extended period of niche stability. The idea is appealing to those who view the cheetah as a masterpiece of evolutionary construction, but various predictions based on this theory are not borne out by the facts. For example, a species ideally suited to a particular niche would be expected to have no closely related contemporaries, and yet the fossil record reveals that at least four species of *Acinonyx* have roamed the earth in the past few million years. One can only speculate about how many other species existed. An optimally adapted species would be expected to compete successfully with other species in its habitat, but the modern cheetah competes badly: predators always prevail in a showdown and cheetah young are often killed. A species ideally suited to its niche should have gradually shed its deleterious recessive alleles and acquired homozygosity only for highly adaptive traits, but the cheetah still has plenty of bad genes, for example those controlling sperm morphology.

Our third hypothesis, and the most plausible, is that at some point in the past the species went through an extreme population bottleneck: a severe population reduction. This was followed by inbreeding, which diminished the gene pool by the chance loss of alleles.

How extreme would a bottleneck have to be to produce a population with zero percent enzyme variation and with identity at the MHC? What caused the bottleneck in the cheetah populations and when did it occur? Theory and practice demonstrate that a population that passes through a bottleneck of a mere seven individuals can still retain about 95 percent of its original genetic variation; it can retain that variety if the survivors expand their numbers quickly and geometrically. (Slow expansion in a small population increases the likelihood that different gene types will disappear.) We therefore suspect that at least once



GENETIC VARIATION in 12 species, including nine felines, is compared based on electrophoretic surveys of enzymes. ("Heterozygous" loci are positions on chromosome pairs where homologous genes differ from each other.) Of the more than 250 species that have been studied by population geneticists, the cheetah has the least amount of genetic variety.

and perhaps several times in the past the cheetah's forerunner populations must have dropped to a very few individuals, escaping extinction by a whisker; it is also possible that the surviving cheetahs never managed to expand their numbers rapidly. Just why the cheetah population would have dwindled so severely is anyone's guess. The possibilities range from climatic catastrophe to viral or bacterial plagues to destruction of the habitat or outright killing by humans.

The timing of the first bottleneck, like the degree and cause, is difficult to determine. A prime candidate is the time between 10,000 and 12,000 years ago, at the end of the geologic epoch known as the late Pleistocene. Before this there were many species of *Acinonyx;* the modern species, *A. jubatus,* had a worldwide distribution. Then something caused a massive extinction of mammalian species, particularly ones in North and South America, destroying 75 percent of those species. Many large carnivores succumbed, including cheetah species other than *A. jubatus,* and the range of the cheetah eventually became limited to its present one in certain parts of Africa. Whether the Pleistocene's environmental catastrophe did in fact cause a bottleneck in the population of *A. jubatus* is not at all certain. As likely as the notion seems, we cannot exclude the possibility that a more recent crisis is responsible for the cheetah's vulnerability today, and we hope our ongoing studies will provide more insight into the animal's past history.

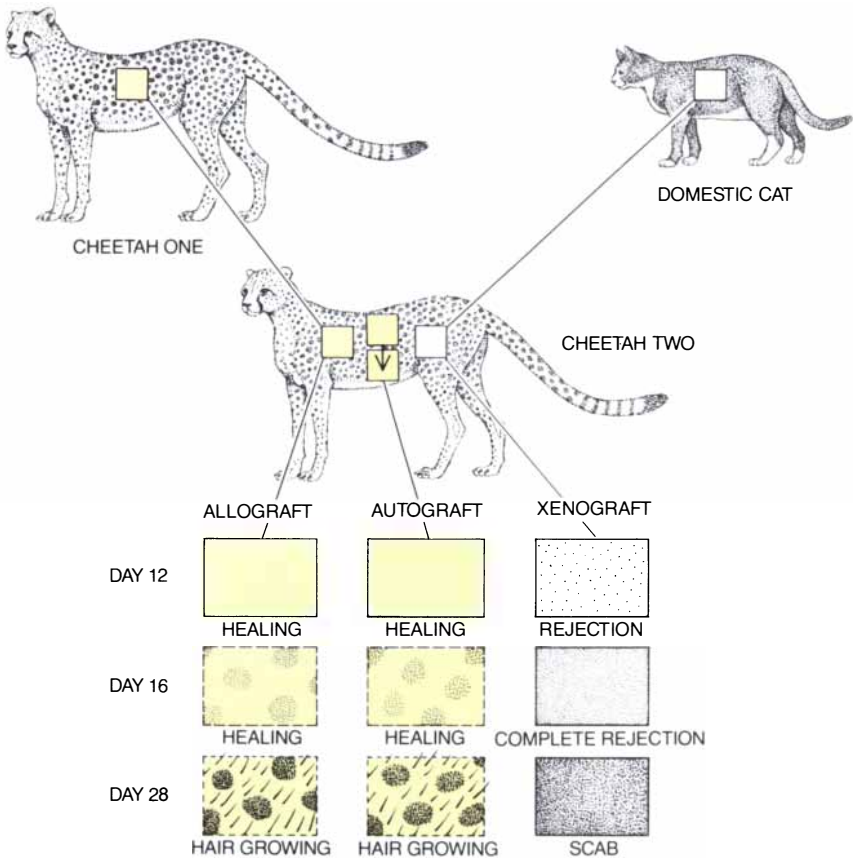
Our findings have revealed much about the past and present status of the cheetah, but an important component that must also be considered is the prospect for the future. The cheetah's genetic uniformity is certainly dangerous for the species, but we think it should not be interpreted as a death sentence. There are a number of reasons. If the proposed bottleneck did occur as long ago as the Pleistocene, then natural selection, which ensures that individuals with seriously maladaptive traits do not survive, has surely eliminated the most dramatically deleterious genes by now.

Several other animal species have gone through serious population bottle-

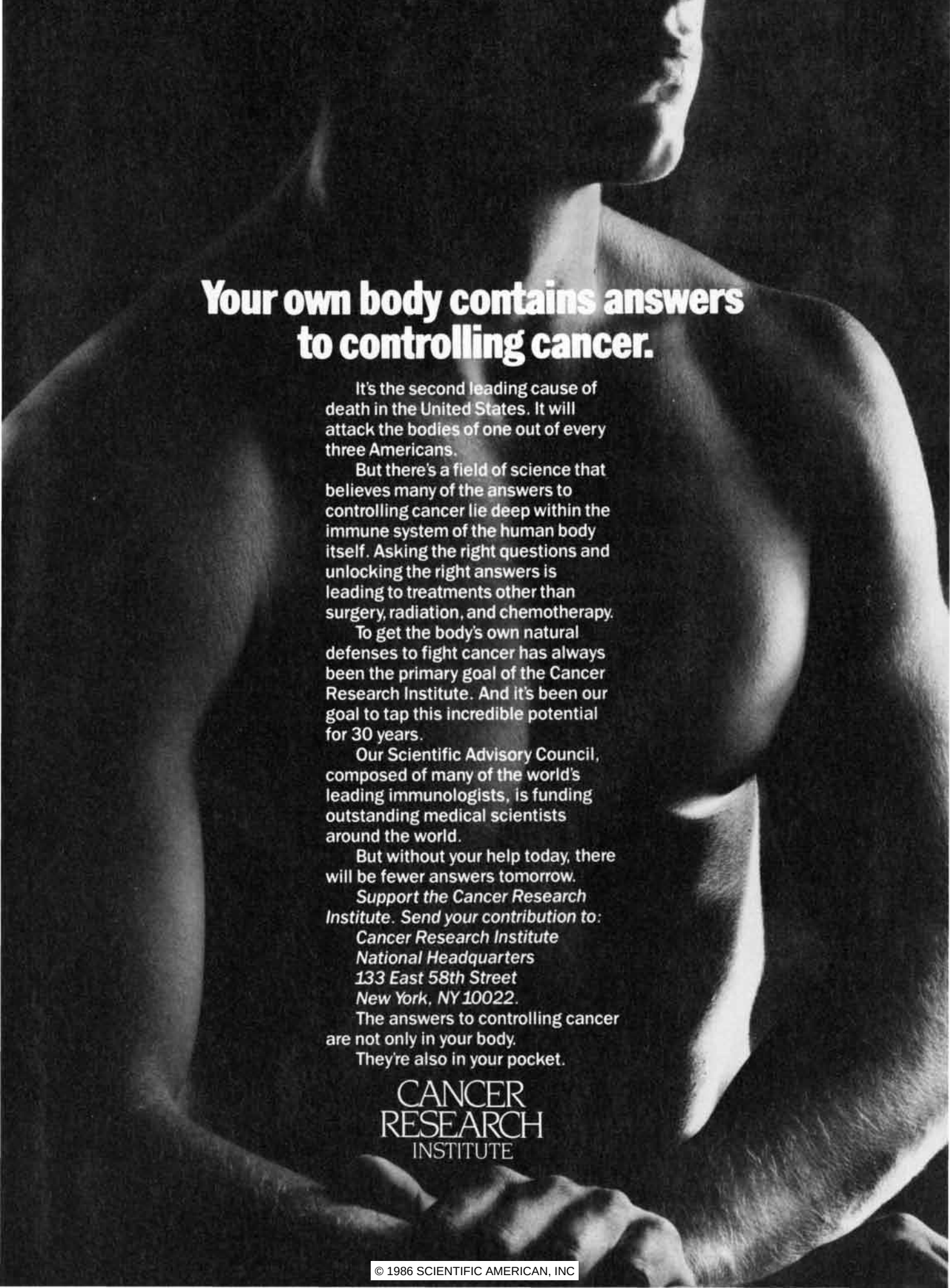
necks and seem to be recovering, and the cheetah may have the same good fortune. The northern elephant seal is an encouraging example. At the turn of the century the population was reduced to about 20 animals. Yet after (and perhaps because of) the passage of protective legislation in 1922 and later, the seal population off the coast of Mexico and California grew. Today the number of seals in the area reaches into the tens of thousands.

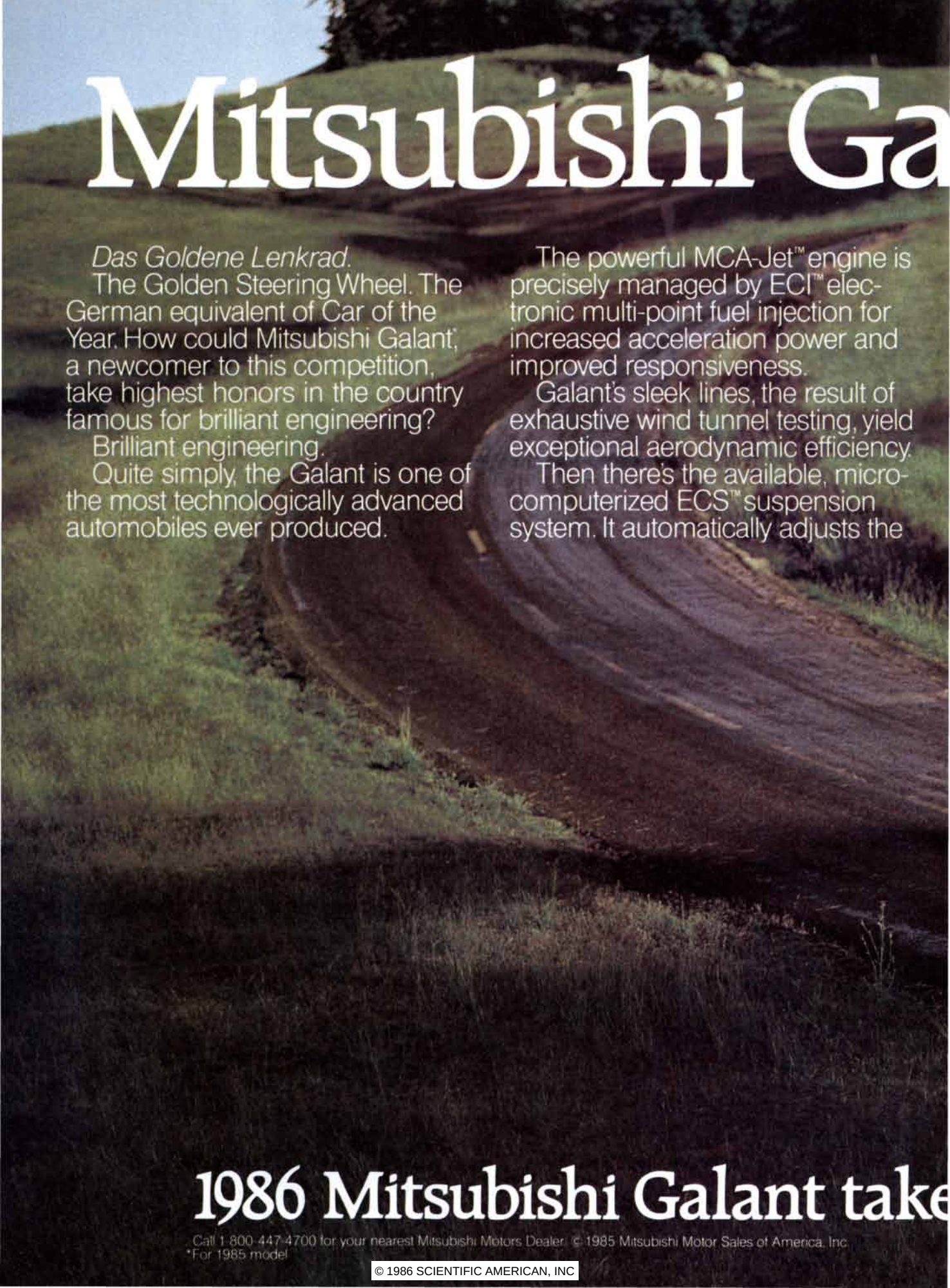
Finally, as we mentioned above, there is the possibility that the cheetah subspecies in eastern Africa is genetically different from the subspecies in southern Africa. If that is the case, the eastern African cheetahs might be introduced into captive-breeding programs, which with few exceptions have largely involved cheetahs from southern Africa. Even if the two groups seem, on the basis of our limited methods, to be genetically similar, one might still want to breed them together. This has been attempted successfully at the Whipsnade Zoo, outside London. Cheetahs that have mated there have produced litters reported to have a low infant mortality rate.

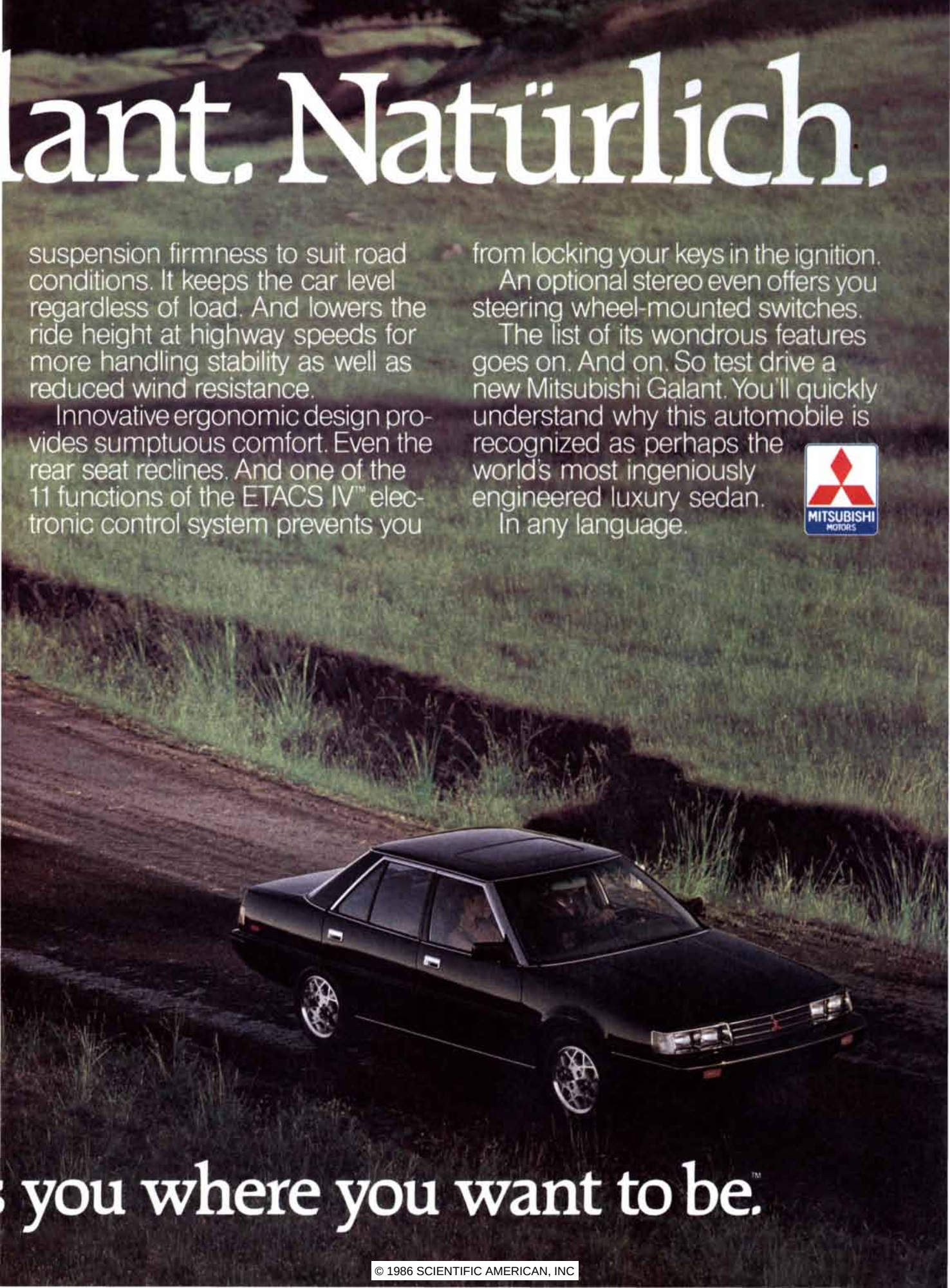
Human beings have been "civilizing" the wilderness for centuries, and the cost has too often been the demotion of evolution's most charismatic creations to endangered or threatened status. In the meantime much has been learned from the study of animals about human biology and evolution. We hope the future will see such knowledge applied to saving the world's threatened and endangered species. Perhaps science can even help the world's fastest mammal to win its race for survival.

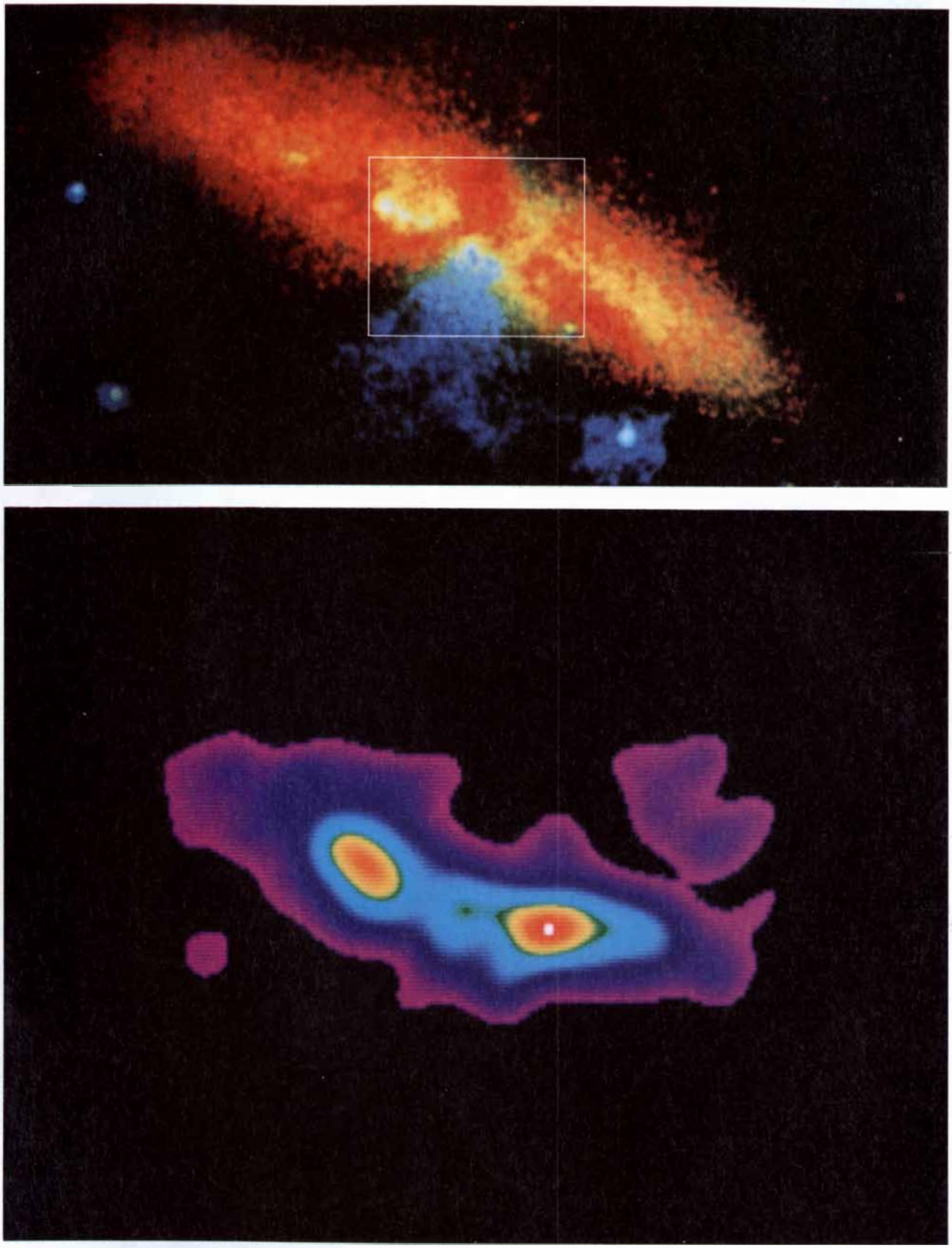


PROOF OF CHEETAH'S MONOMORPHISM at the major histocompatibility complex (MHC) is depicted highly schematically. Three types of skin grafts were surgically transplanted onto cheetahs: allografts (from unrelated animals of the same species), autografts (from self) and xenografts (from another species). Both the allografts and the autografts were readily accepted and by day 28 after grafting were developing cheetahlike spots and growing hair. The xenografts, from a domestic cat, were rejected by day 12 after grafting. MHC genes determine whether a graft will be accepted; unless immunosuppressive drugs are given, a graft normally survives only if the donor has the same MHC genes as the recipient. Therefore the acceptance of the allografts by the cheetahs in the study could be explained in two ways: either their immune system was somehow suppressed or they had identical MHC genes. Rejection of the xenografts indicated that the cheetahs' immune systems were operating adequately, and so the animals must have had identical genes at the MHC.









GALACTIC NUCLEUS is seen with an optical telescope (*top*) and with an array of radio telescopes on which superconducting tunnel detectors are mounted (*bottom*). The false-color radio image is a map of the emissions from carbon monoxide molecules in the galaxy and fits in the white rectangle in the top illustration. The optical photograph was made at the Palomar Observatory and the radio image was made at the California Institute of Technology's Owens Valley millimeter-wave interferometer. The galaxy is known as M82.

Superconducting Tunnel Detectors in Radio Astronomy

A sensitive new radiation detector enables astronomers to explore regions of the universe hidden from optical telescopes, such as sites where stars are now being born

by Thomas G. Phillips and David B. Rutledge

Astronomy often seems to be a protracted struggle on the part of the observer to collect enough of the faint electromagnetic radiation emanating from distant stars to make significant measurements. To aid in the battle, investigators have pursued two strategies: they have built larger telescopes, which gather more electromagnetic waves, and they have increased the sensitivity of detectors mounted on the telescopes to the waves that are captured. The construction of huge optical and radio telescopes can be extremely expensive. Fortunately a twofold increase in the sensitivity of a detector can have as much impact on a telescope's effectiveness as a twofold increase in the area of the instrument's aperture. For this reason astronomers have been particularly eager to improve the sensitivity of detectors.

The effort has led in the past decade to the emergence of an instrument known as the superconducting tunnel detector. Aside from its extraordinarily high sensitivity, the detector is also important because it is unusually well suited to the millimeter and submillimeter bands, or wavelength ranges, of radio astronomy. More precisely, radio astronomy is concerned with the measurement of celestial radiation having wavelengths from several thousand meters to a few fractions of a millimeter (the radio part of the electromagnetic spectrum). It is difficult to detect radiation at the short end of the range by many of the conventional methods of today.

Astronomers would like to know more about these regions for two reasons. First, the sites of new star formation lie deep within clouds of interstellar gas and are therefore obscured from the view of optical telescopes. (Optical wavelengths are typically a little less than a micron, or a millionth of a meter.) Radiation in the millimeter and submillimeter bands, however, usually passes through the clouds as if they were virtually transparent and can thereby provide clues about regions of stellar birth.

Second, radiation in the submillimeter band can yield information about the nature of the interstellar gas itself. Although most of the dense gas is composed of hydrogen molecules, nearly half of it may consist of a wide variety of trace molecules. What molecules are they? What are their relative abundances? The laws of quantum mechanics offer a way of answering these questions. In particular a given molecule can rotate only at a set of specific, discrete frequencies. The allowed frequencies depend on properties of the molecule.

Heavy molecules, for instance, rotate at relatively low frequencies and light molecules rotate at higher frequencies. When a molecule makes a transition from one allowed rate of rotation to another, it emits radiation of a characteristic frequency, by which an individual species of molecule can be identified. Much of the radiation from rotating molecules in the interstellar gas lies in the submillimeter band. The superconducting tunnel detector is therefore an ideal instrument to employ in determining the composition of the gas.

The heart of a superconducting tunnel detector consists of a thin layer of insulating material (a substance that does not conduct electricity) sandwiched between two superconductors. A superconductor is a metal, such as lead or niobium, that exhibits remarkable properties when it is cooled to temperatures near absolute zero (zero degrees Kelvin, or -273 degrees Celsius). Perhaps most remarkable of all, a superconductor carries electricity with no resistance. Such a sandwich of superconductor, insulator and superconductor is known as a superconducting tunnel junction.

Now, according to the laws of classical physics, the insulator should prevent an electric current from flowing through the junction. Quantum mechanics provides a loophole. According to this important theory, an electron is represented as a probability wave: its position is therefore indeterminate. As a result there is a small probability that an electron in one superconductor can appear in the other superconductor, as if it had tunneled through the insulator.

In the operation of a superconducting tunnel detector radio waves are directed onto a superconducting tunnel junction. Electrons in the superconductors absorb the energy of the radiation. Higher-energy electrons have a far greater probability of tunneling through the insulator. The amount of electric current passing through the insulator therefore increases with the amount and type of radio waves striking the junction. By measuring the current one can infer the nature of the electromagnetic radiation exciting the device.

In spite of the rather exotic properties of the superconducting tunnel detector, its basic principle of operation—electrical conduction induced by incident radiation—is the same as that underlying the conventional detectors used in observatories for measuring visible radiation. In a highly simplified version of an optical device, for instance, light from a telescope would be focused on a piece of semiconduct-

ing material such as silicon. The atoms of a semiconductor hold most of their electrons closely, with the important exception of a few electrons known as valence, or outer, electrons. Although valence electrons are not as tightly bound as other electrons, they can usually move only from one atom to an adjoining neighbor, helping to form chemical bonds. Light absorbed by the semiconductor can, however, promote the valence electrons into an energy level known as the conduction band: a state in which the electrons can move freely throughout all the material. If a voltage difference is applied across the semiconductor, electrons will then flow from one end of the material to the other. The amount of flow, or electric current, is proportional to the amount of light incident on the device. By measuring the current one can therefore infer the quantity and nature of the light that has entered the aperture of the telescope.

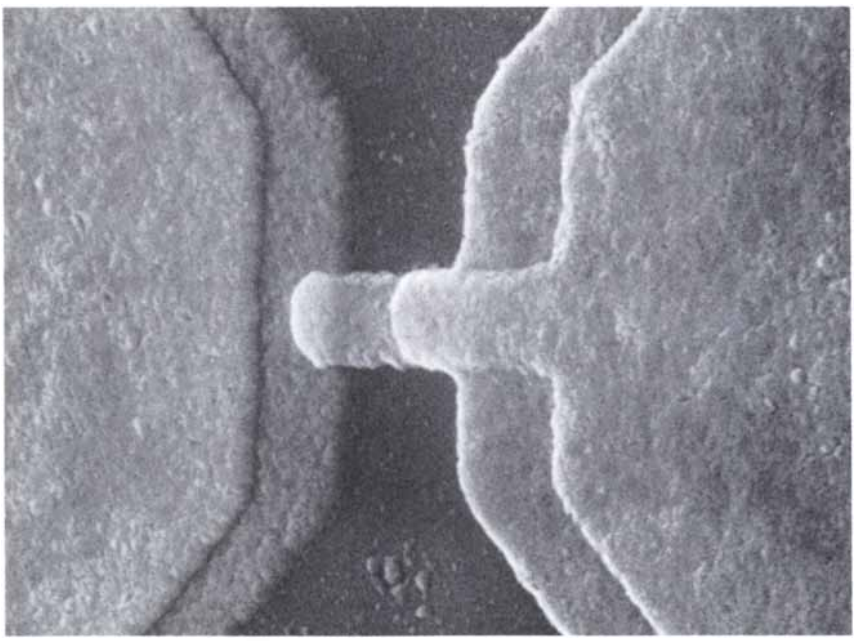
A conventional device cannot, however, detect radio waves. The reason is that such radiation lacks enough energy to promote the valence electrons of semiconductors into conduction bands. The energy of any electromagnetic wave is inversely proportional to its wavelength. Because a millimeter radio wave is roughly 1,000 times longer than a light wave, its energy is 1,000 times less. The energy needed to make a semiconductor carry an electric current is about one electron volt; the energy of a one-millimeter radio wave is roughly a thousandth of an electron volt.

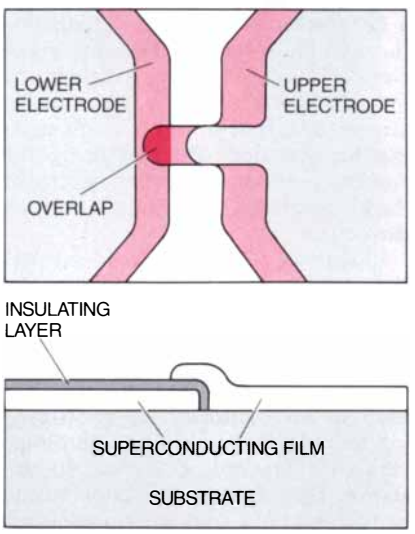
Since a superconductor is a metal, it already has many electrons in its conduction band and cannot be employed to detect light. Its electrons can, however, reside in separated energy states. The amount of energy needed to move the electrons from one state to another matches the energy of a one-millimeter radio wave and explains the suitability of the superconducting tunnel detector to radio astronomy.

The existence of the separated energy states follows from the theory of superconductivity, the underpinnings of which were first presented in a paper published in 1957 by John Bardeen, Leon N. Cooper and J. Robert Schrieffer, then at the University of Illinois at Urbana-Champaign. According to their theory, electrons in a superconductor bind to form Cooper pairs. In other words, the electrons experience an attractive force. Although electrons have the same charge (negative) and repel one another in free space, they can appear to attract one another in a metal by interacting with positively charged ions there. The positive ions are arranged in a lattice, which is distorted whenever an electron moves through it. The distortions in turn attract another electron, so that the two electrons form a pair. This cooperative interaction leads to the appearance of an energy gap in the allowed states of the electrons. Below the gap all the states are filled. Above the gap the states are empty.

In the absence of radiation, therefore, the electrons of a superconducting tunnel detector are bound in Cooper pairs. The application of radio waves to the junction splits some of the pairs in one of the superconductors and promotes one electron of each pair to a higher energy. The promoted electrons can subsequently tunnel through the insulator and enter the unfilled states on the other side, thereby increasing the flow of current through the device. If the detector is mounted on the base of a telescope, the current will contain information about the astronomical object on which the telescope is trained.

The events leading from the seminal paper of Bardeen, Cooper and Schrieffer to the realization of the superconducting tunnel detector have involved many investigators. Much of the work was done for reasons other than developing new detectors. In the early 1960's Ivar Giaver of the General Electric Company demonstrated many of the properties of the superconducting tunnel junction. A decade later the work advanced significantly when workers at both IBM and AT&T Bell Laboratories sought to make fast-switching computer elements out of

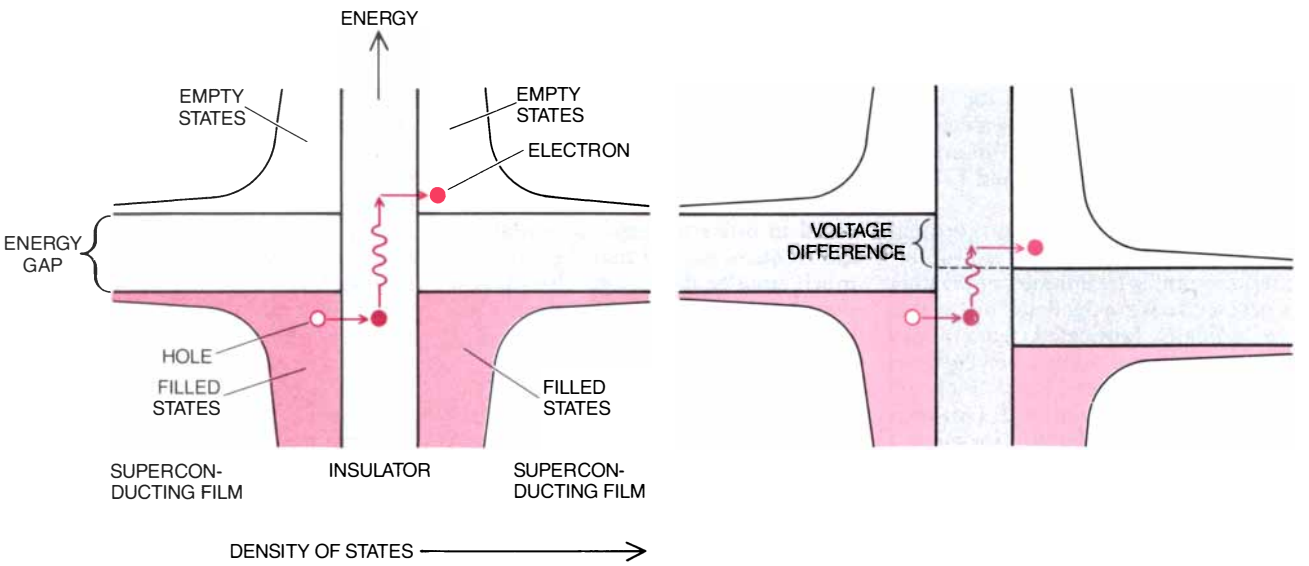




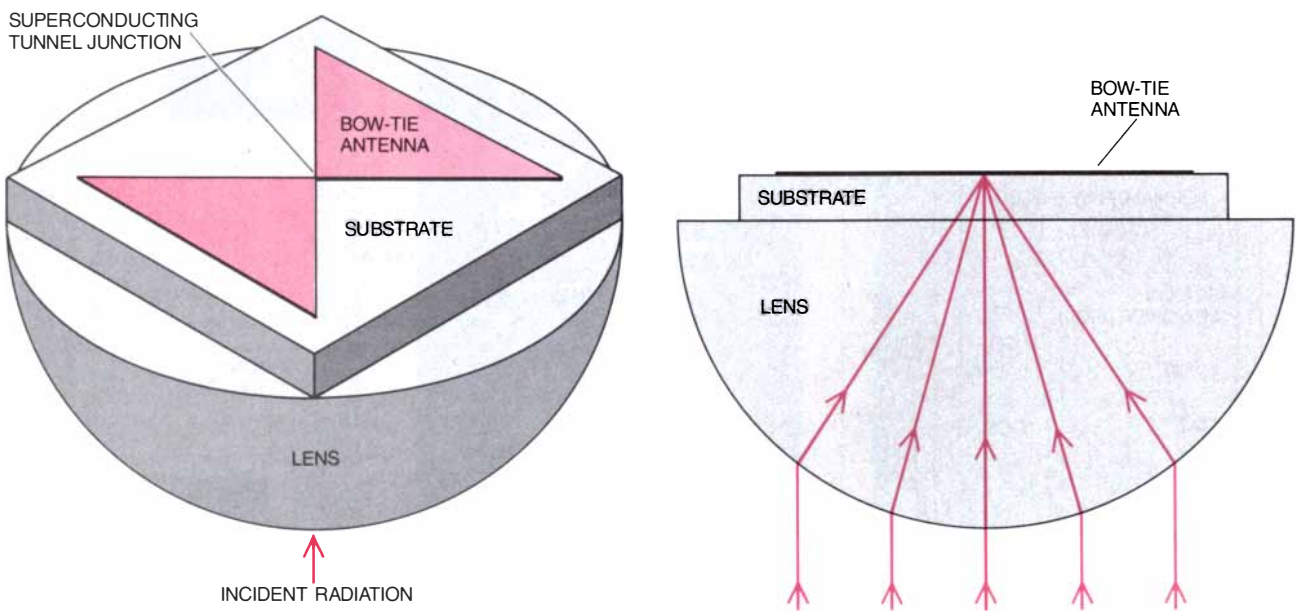
HEART of the superconducting tunnel detector consists of a thin layer of insulating material sandwiched between two superconductors. The device shown in the scanning electron microscope photograph at the top is made from two thin films of a lead alloy; lead becomes superconducting at temperatures near absolute zero (−273 degrees Celsius). The region of overlap of the films (*see top view at upper left and cross section at lower left*) forms the detector, which is about one micron (a millionth of a meter) square. The insulating layer between the two lead films was made by exposing the top surface of the bottom film to air and allowing a natural oxide layer to grow on it. Ronald E. Miller of AT&T Bell Laboratories fabricated the device. The technique of preparation, known as double-angle evaporation, makes both films appear to be double. In the operation of the detector, however, only the region of overlap is important.

superconducting tunnel junctions. It turns out that a high-speed device must be extraordinarily tiny—on the order of one square micron. To build such an element the IBM and Bell Laboratories teams first had to develop a technique known as microlithography in order to inscribe the device on a silicon chip. The achievement of that process inspired other investigators to turn to the development of radio-wave detectors, since those devices must also be extraordinarily tiny to register the relatively high frequencies that are characteristic of millimeter and submillimeter radiation.

First reports of working superconducting tunnel detectors with amazingly low noise, or unwanted disturbance, were produced simultaneously in 1979 by Gerard J. Dolan, one of us



ELECTRON TUNNELING underlies the operation of the superconducting tunnel detector. A cooperative interaction between the electrons of a superconductor leads to an energy gap in the allowed states of the electrons (*left*). Below the gap all the states are filled; above the gap the states are empty. The application of radio waves to the detector promotes some of the electrons in one of the superconductors to a higher energy. These electrons can tunnel through the insulator and enter the unfilled states on the other side. By measuring the flow of electrons, or electric current, through the insulator one can deduce the properties of the radiation incident on the device and thereby infer the nature of the radiation source. The detector can be tuned to radiation of a particular wavelength by applying a voltage difference across the junction (*right*), raising the energy levels of one side with respect to those of the other side.



BOW-TIE ANTENNA "feeds" radio waves from the telescope to the superconducting tunnel junction at the apex of the bow (*left*). A typical antenna is a few millimeters long and a tenth of a micron thick. The metal is evaporated onto the substrate, usually quartz or silicon, and the pattern is made by the same photolithographic procedures that are used for integrated circuits. Such antennas are easy to make, but they are not without problems. In particular, the antennas interact strongly with substrate modes: waves that propagate in the substrate. The interaction can result in a loss of as much as 80 percent of the incoming radiation. A solution devised by the authors is to attach a lens made of the same material as the substrate to the back of the substrate. The lens focuses the radiation onto the bow-tie antenna (*right*), eliminating the substrate modes. Such a device can be operated over a wide range of frequencies.

(Phillips) and David P. Woody of Bell Laboratories and by Paul L. Richards and Tek-Ming Shen of the University of California at Berkeley and Richard E. Harris and Frances L. Lloyd of the National Bureau of Standards. Staffan Rudner and Tord Claeson of the Chalmers Institute of Technology in Sweden also reported a working device in 1979. This is by no means the entire story. Other investigators who made important contributions to the development of superconducting detectors include John R. Tucker at Illinois and Aly H. Dayem and Raymond J. Martin of Bell Laboratories.

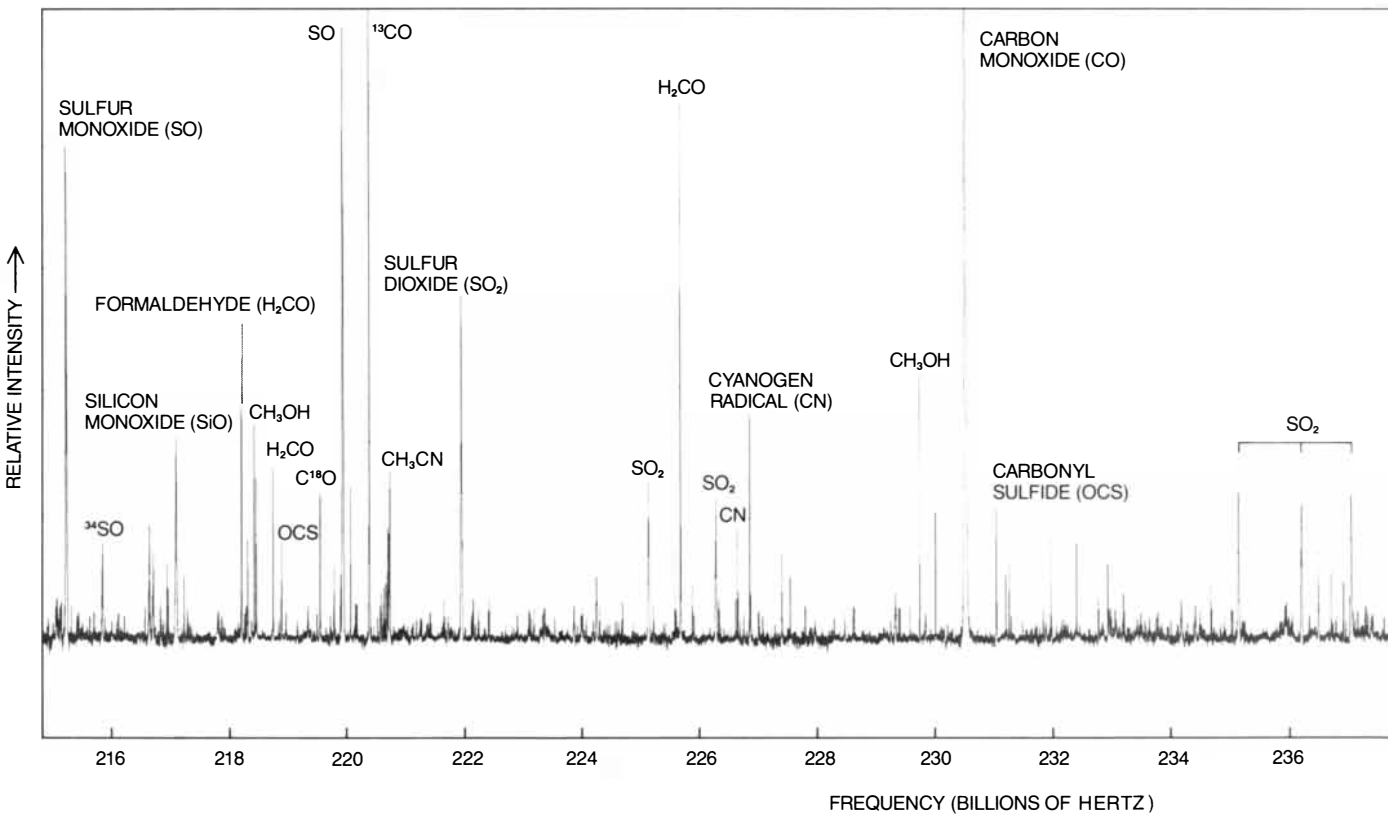
Today many laboratories around the world make use of several different lithographic techniques, or printing processes, to form the detectors, which are typically fabricated from alloys of lead. The insulating layer between the two pieces of lead has a thickness of about 20 angstrom units. (An angstrom is one ten-billionth of a meter.) This remarkable degree of thinness is achieved by simply exposing the top surface of the bottom piece of metal to air and allowing a natural oxide layer to grow on it. Lead becomes superconducting at roughly eight degrees K., so that the instruments are operated in a cryostat containing liquid helium. (Helium liquefies at 4.2 degrees K.) Future advances will no doubt involve the use of niobium or niobium nitride superconductors, which are harder to produce but should be less susceptible to chemical degradation and should have greater sensitivity.

We have not yet addressed the problem of how radio waves from an enormous telescope with a diameter of 10 meters can be "fed" into a tiny superconducting tunnel detector measuring one micron across. It is paradoxical that the detector must be small in order to respond to high radio frequencies, but that the size is so much smaller than a wavelength that the wave cannot be focused on it directly, as is the case with detectors in optical and infrared telescopes.

The solution is to make a small feed antenna that captures the waves focused by the telescope and brings them to the superconducting tunnel junction. For millimeter wavelengths the antenna is usually a metal horn connected to a hollow metal tube. The junction lies on a thin sliver of silicon mounted inside the tube. Although hollow metal waveguides transfer energy to the detector efficiently, they do have limitations. The guide is only a quarter of a wavelength high, and this makes it difficult to build the guides and mount the detectors at submillimeter wavelengths.

To prevent this problem we have made the feed antenna part of an integrated circuit that contains the superconducting tunnel junction. Instead of a machined metal tube, the feed antenna is an evaporated metal film one-tenth of a micron thick. The patterns are made by the same photolithographic procedures that are used in the integrated-circuit industry, and it is easy to control dimensions down to one micron. Moreover, there is an additional advantage. A large number of antennas and detectors can be made simultaneously on the same substrate, resulting in an imaging array that is like a television camera. Such an array would greatly speed the mapping of large astronomical sources.

Other difficulties beset integrated-circuit antennas, however. The antennas interact strongly with substrate modes: waves that propagate in the substrate. The interaction can result in a loss of as much as 80 percent of the incoming radiation. One approach



GASEOUS CLOUD of molecules surrounds part of the Orion constellation (*right*). Although most of the gas is composed of hydrogen atoms, it also contains a wide variety of trace molecules. Astronomers identify these molecules by measuring the frequencies of

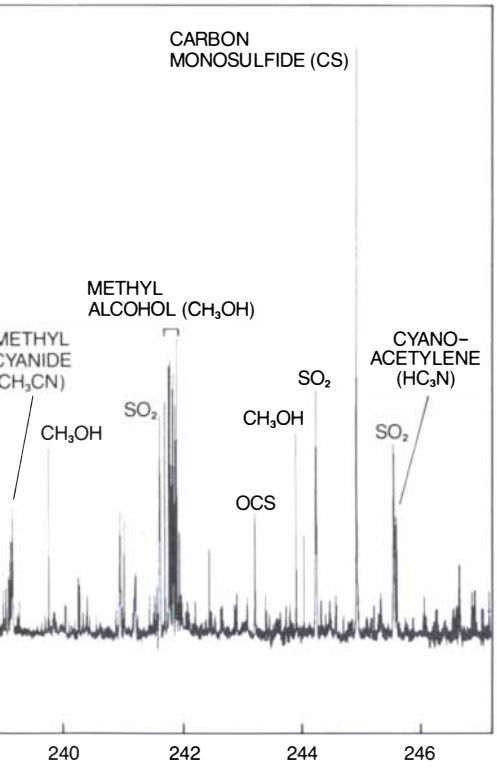
to the problem is to suppress the substrate modes by making the substrate very thin. It is now possible to make freestanding membranes of silicon and silicon nitride with an area of a square centimeter that are less than a micron thick. We are currently attempting to make an array of antennas for submillimeter wavelengths on a silicon nitride film.

Another approach is to put a lens of the same material as the substrate on the back of the substrate. The lens eliminates the substrate modes and takes advantage of the fact that antennas are much more sensitive to radiation from the substrate than they are to radiation from the air. Dean P. Neikirk of the University of Texas at Austin took this approach several years ago when he was a graduate student at the California Institute of Technology. More recently, Michael J. Wengler, one of our graduate students at Caltech, coupled a substrate lens to a bow-tie antenna containing a superconducting tunnel junction at its apex. The performance of the device was excellent over a range of frequencies much wider than that of any other radio-astronomy receiver.

When a superconducting tunnel detector is operated in a telescope, the radio waves are "mixed" with a reference beam and the resulting beam is focused on the substrate antenna. Such a process is known as heterodyne detection and effectively increases the spectral resolving power of the instrument. A device known as a local oscillator generates the reference beam, which consists of radio waves of constant frequency. The resulting beat note, or intermediate frequency, is monitored and analyzed to reveal information about the astronomical object under examination.

Superconducting tunnel detectors have already made a number of important contributions to astronomy. As disclosed by the millimeter-wave telescope at Caltech, which has such a device, the gaseous cloud around Orion is composed of at least 28 distinct molecular species. One of our graduate students, Geoffrey A. Blake, together with research fellows Edmund C. Sutton and Colin R. Masson and one of us (Phillips), made the measurements and identified more than 500 spectral lines emitted by the molecules. The detector, which was constructed by Ronald E. Miller at Bell Laboratories, is so sensitive that it enables astronomers to identify the rarest of molecules, even those containing deuterium, the heavy isotope of hydrogen. The ability to measure the abundance of so many molecular species will undoubtedly have a major impact on the important field of interstellar chemistry.

Superconducting tunnel detectors can also determine the relative motion of molecules toward or away from an observer by detecting the frequency shift of the radiation they emit. (Such a shift is known as the Doppler effect, which is familiar to anyone who has noticed that when a train passes, the pitch of its whistle changes from high to low.) In this way individual clouds of gas within a site of new star formation can be distinguished from one another and their relative motions analyzed. In the Orion cloud, for example, methanol molecules have very narrow spectral lines and sulfur dioxide molecules have wide lines. The reason for the disparity is that methanol exists primarily in the quiescent regions of the cloud, whereas sulfur dioxide is found only in the physically and chemically active regions. Processes of massive star formation take place in these



their characteristic radiative emissions. The molecular spectrum of the Orion cloud is shown above. All the data were obtained with a millimeter telescope fitted with a superconducting tunnel detector. The gas contains at least 28 distinct molecular species (not shown).

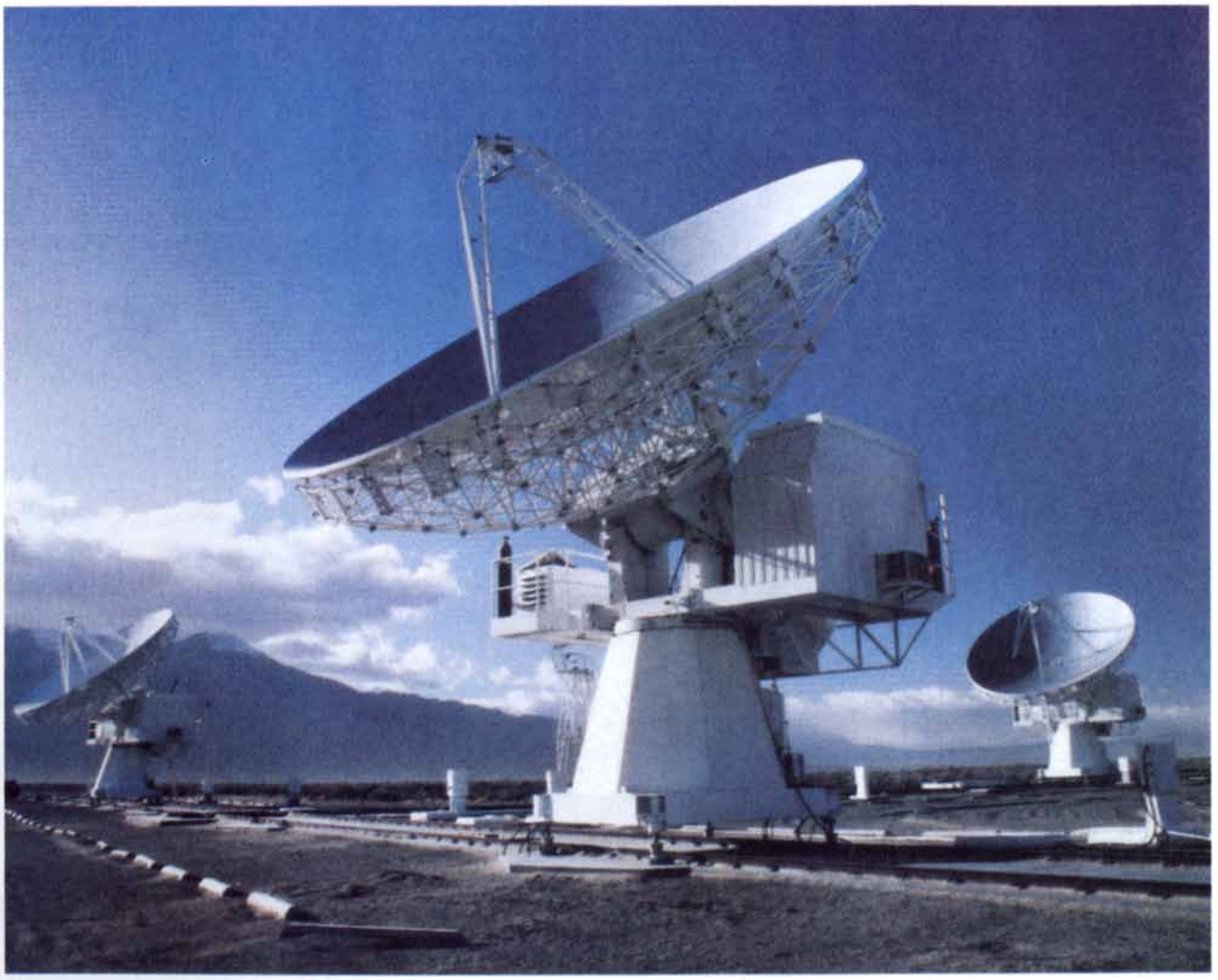
active regions, creating high winds and outflows.

In spite of the sensitivity of the detectors, however, many astronomical objects are too small to observe with a typical telescope. Owing to diffraction of radiation the best angular resolution of a 10-meter-diameter telescope for three-millimeter radiation is about one minute of arc, whereas the regions of star formation, galactic nuclei and other distant objects may have structure on the scale of seconds of arc. To get around this difficulty radio astronomers electronically link several small telescopes, in effect creating a much larger telescope. Such an approach is known as interferometry. The Caltech Owens Valley millimeter-wave interferometer exploits this technique. The facility consists of a group of three millimeter telescopes constructed by Robert B. Leighton that have diameters of 10.4 meters. Linked together, they have an effective diameter of several hundred meters. Each telescope is equipped with a superconducting tunnel detector at its base.

Probably the most exciting objects so far observed in the sky with the interferometer array are the gas condensations in the nuclei of nearby galaxies. Concentrating on the extremely strong emission from the prolific carbon monoxide molecule, we have made a map of the emission from the gas in the nucleus of the galaxy M82. Carbon monoxide contours delineate a region at the center of the galaxy where many supernova remnants are seen and where massive stars are currently forming. Optical pictures reveal that gas is streaming away from this region of intense activity.

Telescopes for the submillimeter band are also under construction. To avoid the attenuating effects of absorption by water and oxygen molecules in the earth's atmosphere, these facilities are based at high altitudes. The Caltech group and other groups are now constructing submillimeter telescopes on mountain sites, such as Mauna Kea in Hawaii. Measurements have also been made from the National Aeronautics and Space Administration (NASA) Kuiper Airborne Observatory, which consists of a one-meter-diameter optical telescope carried to altitudes of about 40,000 feet by a transport aircraft.

In the long term astronomers hope that the entire submillimeter spectrum, including the regions blocked by atmospheric water and oxygen, will be available from a large deployable reflector (LDR) in space, now in the planning stages at NASA. All these new telescope projects either now incorporate or expect to incorporate superconducting tunnel detectors.



OWENS VALLEY INTERFEROMETER, which is operated by Caltech, consists of three millimeter-wave telescopes linked electronically. A superconducting tunnel detector is mounted on each telescope. The bottom image on page 96 was made with the array. The telescopes were built under the direction of Robert B. Leighton and together have an effective diameter of several hundred meters.

TOYOTA CAMRY

ALL-SEASON CAMRY
You'll love the all-weather traction of front-wheel drive, too. And Camry's thrifty mileage: EPA estimated 28 city/34 highway.** Isn't it time you met Camry, the new friend of the family?

CAMRY. AMERICA'S MOST TROUBLE-FREE CAR* IS ONE CAR YOU CAN LOVE.

RELIABLE
The Toyota Camry is the most trouble-free new car sold in America,* the car with the fewest problems during the first 3 months of ownership. It's so reliable, roomy and comfortable. Camry can make you feel special about a car again. Camry owners say, "Love my Camry." Wouldn't you love to have a car you could feel that good about?

HOW MUCH ROOM?
The family Camry makes generous room for five adults. There's more front head room than any other compact. Soft velour upholstery, full carpeting, and reclining front bucket seats comfort you. Child protector rear door locks are for your peace of mind.

Get More From Life...
Buckle Up!

WHO COULD ASK FOR ANYTHING MORE! **TOYOTA**

*Based on problems encountered in first 90 days of ownership—1985 J.D. Power & Associates New Car Initial Quality Survey.
**Camry with 5-speed manual overdrive transmission.

Postglacial Foraging in the Forests of Europe

Hunting-and-gathering is often thought of as little more than the prelude to agriculture. A reevaluation suggests it was a parallel development that was as productive as early farming in some areas

by Marek Zvelebil

About 10,000 years ago the ice that had covered much of Europe and Asia for many millenniums began to melt rapidly. As the ice retreated, new environments came into being in northern Eurasia. Steppe and tundra, which had been common during the Pleistocene ice age, were colonized by forests made up of coniferous and deciduous trees. In the newly forested regions a new way of life arose. During the Ice Age the inhabitants of northern Eurasia had lived largely on big game, which they hunted across the frozen steppe. The forest that replaced the steppe was a complex ecosystem containing a much richer mixture of edible plants and animals, and the postglacial hunting-and-gathering bands exploited the new abundance fully, gathering plant foods, waterfowl and small game in addition to big-game species.

The postglacial foraging way of life lasted until it was replaced by agriculture, a transformation that took place in northern Europe about 5,000 years ago. Traditionally the five-millennium period between the retreat of the glaciers and the introduction of farming has been considered a humble interlude. It was thought that hunting-and-gathering was a crude mode of subsistence, capable of supporting only small groups that had to be constantly on the move to eke out a living from the forest. In the traditional view hunting-and-gathering offered little competition to farming as a mode of life. Therefore, it was thought, agriculture spread rapidly from the Near East to the forested regions of Europe, brought by colonizing farmers or adopted by foraging bands who were quick to see its potential for improving their dreary circumstances.

Recently that view has begun to change as the complexity and productivity of the postglacial foragers have come into sharper focus. It has become clear that in favorable environments hunting-and-gathering was capable of supporting populations much denser than was previously thought. In some areas foragers lived relatively sedentary lives in permanent settlements where a developed culture and the elements of social differentiation were present. Such communities did not readily adopt farming. It has become clear that in parts of northern Europe the advent of agriculture was long delayed and that the hunting-and-gathering bands took up farming only when crises resulting from climatic changes forced them to do so. The stubborn persistence of foraging long after it "should" have disappeared is one of the qualities that is contributing to a fundamental reassessment of postglacial hunting-and-gathering. I think that when the reassessment is complete, foraging in the postglacial forests will be considered a development parallel with agriculture and one that, for a time at least, was equally viable as a means of subsistence.

The Eurasian forest zone, to which the new conclusions apply, is by no means ecologically uniform. In the north the forests consist mainly of coniferous species. In more temperate latitudes deciduous species prevail, or a mixture of deciduous and coniferous species. The distribution of these components has changed considerably during the climatic fluctuations of the past 10,000 years. At the time of the flowering of postglacial hunting-and-gathering societies the forest zone extended farther north than it does today. Some areas that are now grassland or semidesert were then covered by forest-steppe and woodland.

To take account of such variation I shall employ a broad geographic definition of the forest zone rather than the strict ecological one. For my purposes the forest zone covers most of Europe north of a latitude of 40 degrees. (The 40th degree of latitude extends through central Spain, southern Italy, the Aegean peninsula and northern Turkey before reaching the Caspian Sea.) One of the most significant features of the region so defined is that during the post-Pleistocene many of its resources were concentrated near bodies of water: rivers, lakes and seas.

Another important facet of the ecology of the forest zone was the fact that the quantity of available resources fluctuated dramatically with time. The fluctuations were of several types. One type was long-term and came from the ecological structure of the Eurasian forest itself. The food chain there was dominated by species such as reindeer, hare and game birds, which feed on plants and in turn provide food for predators such as bears and lynx. Many of these lower species were opportunistic (in the zoological sense). Because they breed quickly and have large litters, their populations tend to increase rapidly if more food becomes available. If the prey species increase, so do the predators. Often the end result is a "crash" in the population of prey. The presence of many opportunistic lower species undoubtedly led to unpredictable and extreme variations in the food resources available to hunting-and-gathering groups.

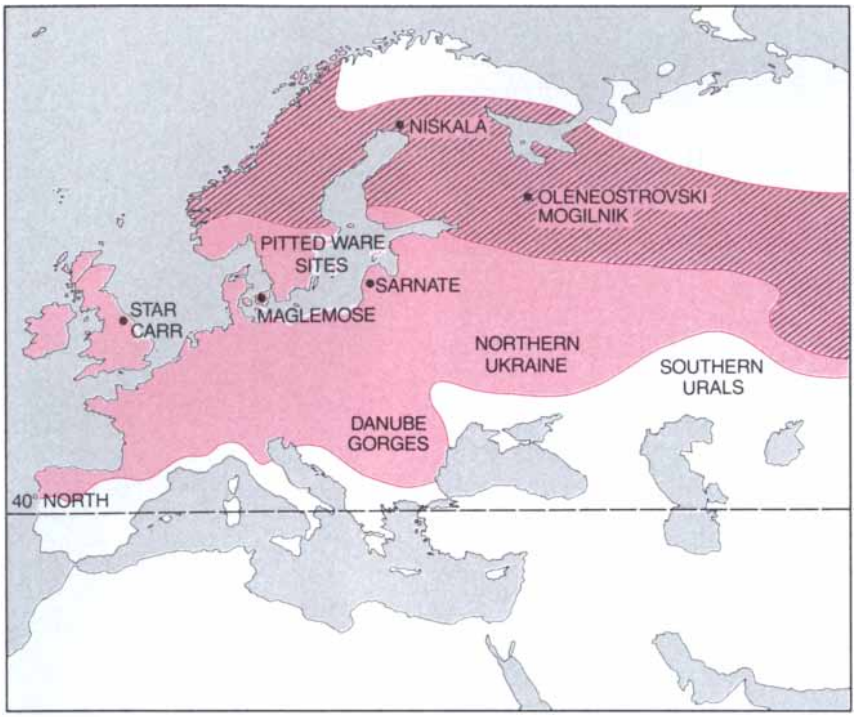
The second type of fluctuation, which is more predictable and has a shorter term, is the variation that accompanies the cycle of the seasons. Forests in the northern Temperate Zone are highly seasonal places. The type and quantity of food available varies sharply from season to season. Much of the variation is due to the

fact that some of the most significant sources of food are migratory, including many species of waterfowl and such anadromous fishes as salmon, which return every year from the ocean to breed in the streams where they were born. The pattern of migration implies that some species are available only for brief periods of the year, which tend to be in the spring and the fall. At other times of the year, particularly the coldest part of winter, food can be very scarce.

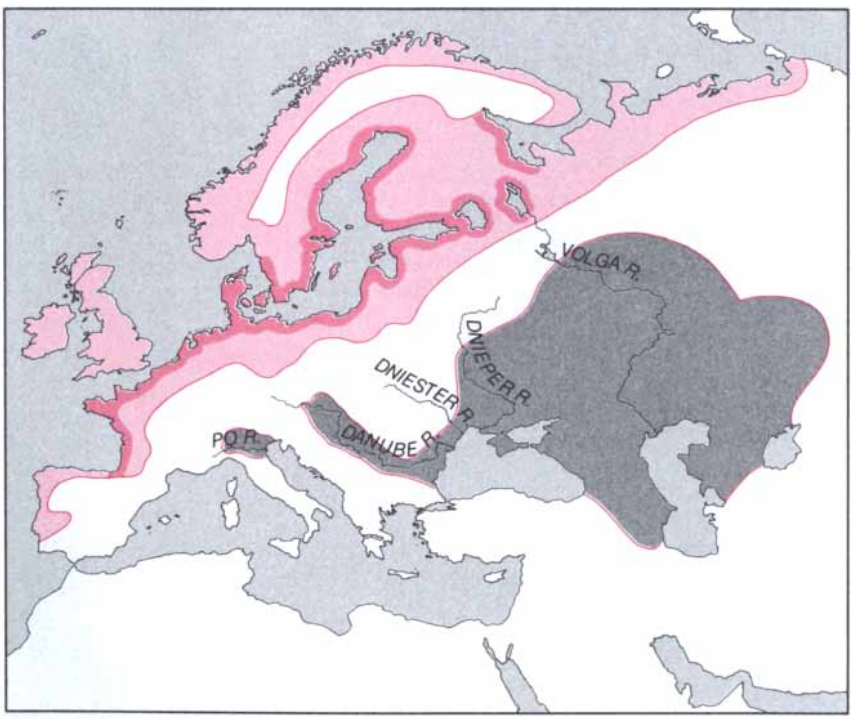
How did the postglacial inhabitants of Europe cope with such variability? One possible answer to the question might have been agriculture. Agriculture, however, was not initially an option for the inhabitants of the forest zone. The reason is that agriculture, as developed in the Near East, relies on both farming and animal husbandry. Those practices in turn depend on the presence of a combination of species that can be domesticated: plants to be cultivated and animals to be herded. In the Near East cereals provided the basis of farming and sheep and goats did the same for husbandry. The northern forest zone lacked such a propitious combination. Some species could be domesticated, but they tended to occur in isolation and not as part of the range of species that was needed to make agriculture possible.

Yet agriculture is only one strategy for increasing the yield of resources and overcoming the inherent instability of the environment. Other strategies are possible, and what they share with agriculture is an increased investment of labor power in obtaining food. For example, the increased investment, which is called intensification, might begin with the accumulation of much knowledge of the life cycle of a particular species. Based on that knowledge special methods might be developed for capturing the prey. In this way a food species might be managed so that yields are increased without full domestication.

SLATE OBJECT with the head of an elk comes from a site near the village of Niskala in northern Finland. The beautifully striated object was carved in the third millennium B.C. by a member of a late Stone Age hunting-and-gathering band. Some archaeologists believe the artifact was a knife, but the author holds that its function was mainly social: possessing it was a means of increasing one's social status. Late in the postglacial period such status objects appeared widely in the forests of Europe, which may indicate a heightening of social competition within the foraging groups living there. The elk-headed object is in the collection of the National Museum in Helsinki.



EURASIAN FOREST ZONE, where complex foraging cultures arose, includes several components. Much of northern Europe and Asia is covered by coniferous forests, deciduous forests or forests made up of a mixture of the two types (*color*). The northern part of that zone is covered by forests made up solely of conifers (*hatched area*). The extent of coniferous forest and the extent of the overall forest zone have both fluctuated considerably in the past 10,000 years. Partly for that reason the author has chosen to employ a broad definition of the forest zone that includes most of Europe north of the 40th degree of latitude.



DISTRIBUTION OF RESOURCES in the forest zone is uneven: animals and plants that provide food are concentrated near bodies of water. In northern and western Europe the richest areas are near the Atlantic Ocean (*color*). Within that region the most productive areas are found along certain coastlines (*dark color*). In eastern Europe resources are concentrated in the basins of the Volga, the Dnieper and the Dniester as well as in the valleys of the Po and the Danube (*gray*). The rich resources available in these areas provided the basis for complex foraging cultures, which became specialized for exploiting aquatic species.

The archaeological record suggests that such specialization was the strategy of intensification adopted by many postglacial groups. The organisms they specialized on were generally aquatic because the trend toward specialization took place in the context of an overall shift from the resources of the land to those of the water. Early in the postglacial period the main sources of food at most sites were ungulates, such as deer or wild cattle, and other large game. Later in postglacial times there was a tendency toward the establishment of specialized sites where a significant fraction of the diet was made up of one or a few species. Generally those species were aquatic: fish, shellfish, migratory waterfowl or such marine mammals as the seal. The shift to aquatic resources did not take place simultaneously throughout the Eurasian forests. In the southern Urals its beginnings can be detected among remains dating from between 7000 and 5000 B.C.; in the Danube gorges of Yugoslavia the corresponding date is about 5500 B.C.; in peninsular Scandinavia, 4000 B.C., and in the eastern Baltic region, 3000 B.C.

The Pitted Ware culture of southern Sweden provides a good example of a specialized economy. The Pitted Ware groups, which flourished along the Swedish coasts during the third millennium B.C., lived mainly on the ringed seal. Several significant cultural adaptations improved the efficiency with which these groups hunted seals and used the seal catch. Pitted Ware hunters had a highly developed harpoon technology, which made a kill likely if the seal was within range. After the hunt seal oil and blubber were stored in large clay jars that may have been designed specifically for that purpose. Even the location of the Pitted Ware communities reflected the specialization on seals. Ringed seals are hunted in the spring on the coastal ice where they bear their young, and the settlements were built near coves and other natural coastal features that provide easy access to the seal lairs.

The Pitted Ware culture exemplifies several essential features of the specialist strategy. Maximizing the catch during the brief period the prey is available often implies developing specialized tools, such as the harpoon technology of the Pitted Ware hunters. Specialized tools generally take longer to make and repair than general-purpose tools, but that work cannot be done during the prime hunting season. Hence for specialists the budgeting of time becomes crucial: blocks of time for tool maintenance must be found in the slack periods before and after the "harvest." Moreover, if the catch from

No other system of keeping up can compare with ours.

YOUR SYSTEM: a time-consuming, futile struggle to keep up with the information explosion.

The classic texts are convenient references—but the information they contain is obsolete before publication.
Like many physicians, you probably rely on the texts you first used in medical school. But even using the most recent editions, you find material that no longer reflects current clinical thinking—and they lack the latest information on such topics as herpes, oncogenes, AIDS, and photon imaging.

Reading stacks of journals alerts you to recent developments—but can't give you quick answers on patient management.
Struggling through the hundreds of journal pages published each month—even on only the really significant advances in the field—is arduous and memory-taxing. And it's a task that costs physicians valuable time—their most precious resource.

Review courses cover clinical advances—but, months later, do you recall the details of a new procedure or unfamiliar drug?
Seminars can also be costly and make you lose valuable time away from your practice—expenses that may amount to several thousand dollars. And, the speaker's skill often determines how much you learn.

OUR SYSTEM: a rewarding, efficient way to keep yourself up-to-date—and save hundreds of hours of your time for patient care.

A comprehensive, 2,300-page text in two loose-leaf volumes, incorporating the latest advances in medical practice as of the month you subscribe.
This superbly designed, heavily illustrated resource, called "the best written of all [the internal medicine] books" by *JAMA* (251:807, 1984), provides a practical, comprehensive description of patient care in 15 subspecialties. And, because the text is updated each month, the clinical recommendations reflect all the current findings. A practice-oriented index and bibliography of recent articles further enhance the efficiency of the text.

Each month, six to nine replacement chapters to update your text *plus* new references, a four- to five-page news bulletin, and a completely new index.
You'd have to read hundreds of journal pages each month—and memorize the contents—to get the same information SCIENTIFIC AMERICAN *Medicine* contains. With our updated text, you read only the information you really need. Our authors, largely from Harvard and Stanford, sort through the literature and monitor developments, incorporating the significant advances into our chapters.

At no additional cost, a 32-credit CME program, to save you valuable patient-care time and the expense of attending review courses.
Earn 32 Category 1 or prescribed credits per year with our convenient self-study patient management problems; each simulates a real-life clinical situation. Choose either the complimentary printed version or, at a modest extra charge, the floppy-disk version for your IBM® PC/PC Jr., Macintosh,™ or Apple® IIe or II+ (or compatible).

☐ **Yes,** I'd like to try the SCIENTIFIC AMERICAN *Medicine* system. Please enter my subscription at a first-year price of US$245.*
☐ Also enroll me in the CME program. I'd prefer:
☐ the printed version at **no additional charge.**
☐ the floppy-disk version at $US97* additional. Computer type or compatible:
☐ IBM® PC/PC Jr. (128K RAM, DOS 2.0/2.1) ☐ Macintosh™ (512K RAM)
Apple® II+ (64K RAM) 80-col. card by: ☐ Apple® ☐ Videx® ☐ Apple® IIe (64K RAM)
☐ Enroll me in the floppy-disk CME **only** at US$152.* (Note model above.)
☐ Check enclosed* ☐ Bill me Signature _____
☐ VISA ☐ MasterCard Exp. Date _____ Account No. _____

Name _____ Specialty _____
Address _____
City _____ State _____ Zip Code _____

Or, call toll-free: 1-800-345-8112 (in Penn. 1-800-662-2444). 7D

*Add sales tax if for Calif., Ill., Mich., or N.Y. Allow 8 weeks for delivery. Add US$10 for shipping to Canada.
IBM is a registered trademark of International Business Machines Corporation. Apple is a registered trademark of Apple Computer, Inc. Videx is a registered trademark of Videx, Inc. Macintosh™ is a trademark of McIntosh Laboratory, Inc. and is used by Apple Computer, Inc. with its express permission.

Your subscription includes:

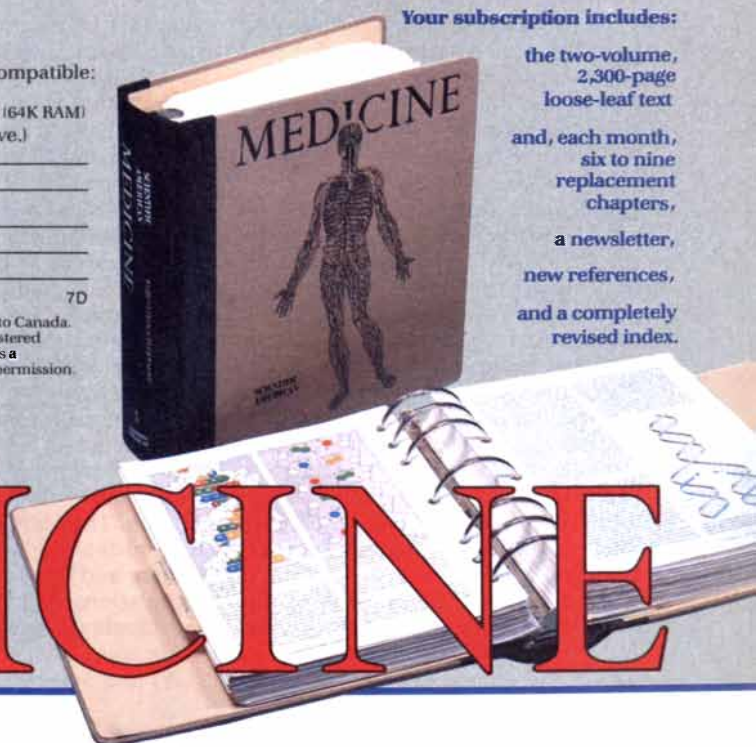
the two-volume, 2,300-page loose-leaf text

and, each month, six to nine replacement chapters,

a newsletter,

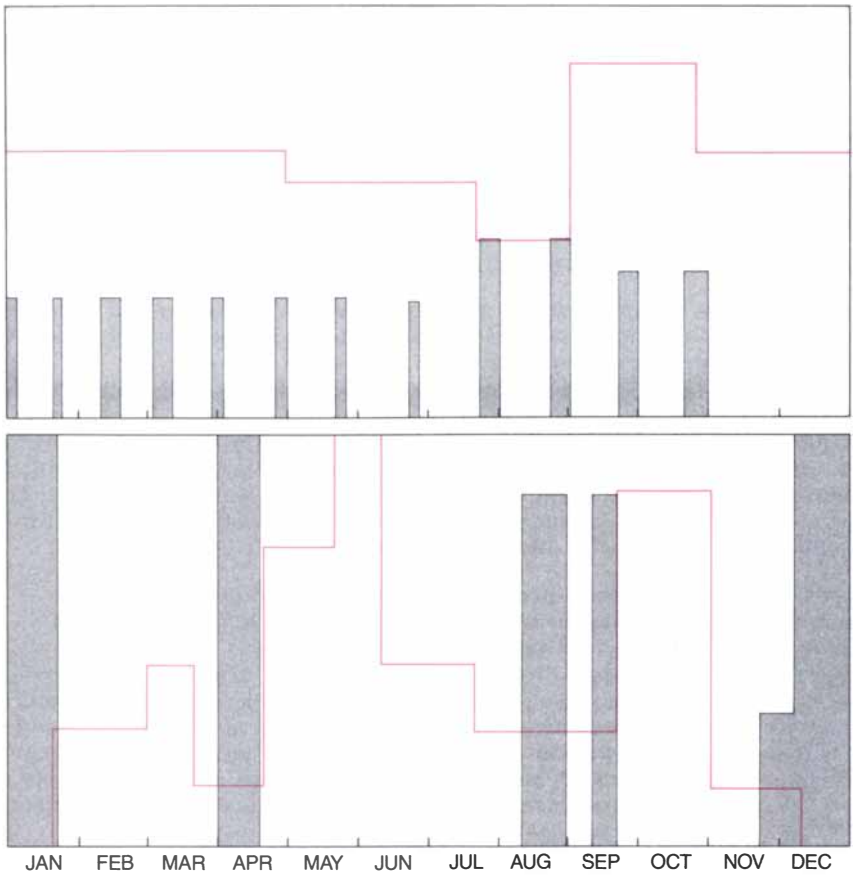
new references,

and a completely revised index.



SCIENTIFIC AMERICAN
415 MADISON AVENUE
NEW YORK, N.Y. 10017

MEDICINE



BUDGETING OF TIME among hunter-gatherers is quite different in the Eurasian forest zone than it is in the Tropics. In each panel the area of the gray bars corresponds to time spent preparing tools and the area within the colored line corresponds to time spent foraging. In a hypothetical tropical environment (*upper panel*) the amount of food available does not fluctuate much, and so tool preparation can be spread through the year. In a hypothetical forest environment (*lower panel*) the amount of food fluctuates greatly with the seasons. Tool preparation must be concentrated in slack periods so that during the periods of peak availability the maximum amount of time can be spent foraging. Such a pattern encourages the development of specialized tools to increase the yield of prey during the peak period.

a brief period is to provide a large fraction of the diet for an extended period, storage is crucial. The Pitted Ware culture was only one of many postglacial foraging groups that had reliable methods for storing food.

Specialized technology, particularly when combined with methods for storing food, helped to stabilize the food supply. Yet the specialist strategy was not without risks. Specifically, concentrating on one source of food rendered the group vulnerable to any fluctuation in the population of that species. Because of the instability of the forest environment, in the long term the population of any species was likely to fluctuate. One solution was to diversify the diet by expanding the range of plants and animals that were exploited for food.

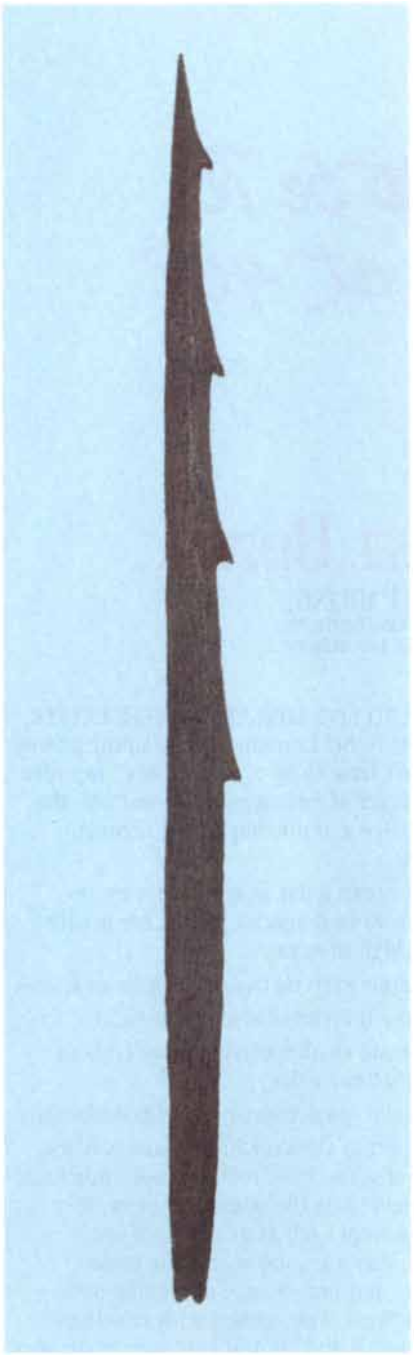
Diversification accentuated the contrast between the postglacial foragers and their Upper Paleolithic predecessors of the Pleistocene ice age. Many Eurasian Upper Paleolithic groups relied on a few species of big game, such as reindeer, wild horses and woolly mammoths. Early in the postglacial period that narrow range was greatly augmented. The Maglemosian culture, named for the Danish peat bog in which the original cultural remains were found, offers a good example of diversification. During the eighth and seventh millenniums B.C. the Maglemosian culture stretched across eastern Britain, the North Sea basin (which was then dry), Denmark, northern Poland and the eastern Baltic region. The people of the Maglemosian culture fed on large game such as aurochs (wild cattle), red deer, roe deer and wild pigs. They caught fish, including pike and salmon. They hunted seals and migratory birds along with small game such as hare and beaver. They also exploited a variety of plants as foods, including hazelnuts, water chestnuts and in some areas even water lilies.

As the specialist strategy brought with it tools intended for ever narrower tasks, the strategy of diversification required tools that could be applied with equal efficiency to many tasks. Such a technology was soon developed, and its most important component was a type of small stone blade called a microlith. Microliths, which are found at almost all postglacial sites, are generally between one and two centimeters long. Their shape varies from trapezoidal to triangular to lunate (half-moon-shaped). Microliths were made by breaking up a larger blade, which was often of flint. After being broken from the larger blade, the sliver of stone was worked into the desired shape. Once shaped, the stone was often embedded in a handle of wood, antler or bone. The great virtues of the microlithic technology were the ease with which it could be repaired and its versatility. A blade could easily be removed from its handle and sharpened or replaced by a new blade. In this way a hunting party might reequip itself in the midst of a hunt on sighting an unexpected quarry.

Indeed, the capacity for retooling during the hunt may well have been at the heart of the advantages of the microlithic technology. Microlithic technology was probably employed in situations where hunters set out in search of food without having a specific type of prey in mind. If game was sighted, the small stone blades could in a very short time be reshaped for capturing that prey. Thus microlithic technology, which required a relatively small investment of labor, was well suited to situations where the type and quantity of game are unpredictable. In contrast, the specialized technology, which required a considerable investment of labor time, was much better suited to situations where resources recur in a predictable pattern.

Although the specialized tools and the multipurpose ones operated on different principles, it should not be thought they were contradictory. Indeed, it appears they complemented each other nicely. The strategy of diversification (carried out with microlithic technology) reduced the risks inherent in the specialized strategy. By combining the two approaches, a postglacial group could have ensured an adequate and reasonably reliable food supply.

Physical evidence for this hypothesis has come from many postglacial sites where specialized tools and microliths have been found side by side. One example is Star Carr in northeastern England, which was the site of a Maglemosian community. Digging there in the 1950's, Grahame Clark of the University of Cambridge uncov-



BARBED SPEARPOINT made of antler was a specialized tool. It was found at Star Carr in northern England. Early in postglacial times Star Carr was the site of a community of the Maglemosian culture. Its members may well have used the spearpoint for hunting elk or deer, among other large game. In the Vale of Pickering in Yorkshire, where Star Carr is situated, the peak hunting season for those species probably came during the winter. Such seasonality made deer and elk prime candidates for hunting with specialized tools. Specialized tools, which require considerable preparation, are used most efficiently for exploiting resources that appear predictably but briefly and offer the potential of high yields. The barbed antler point is in the collection of the Rotunda Museum at Scarborough in England.

ered about 200 barbed antler spearpoints next to microlithic tools. One plausible interpretation is that the barbed antler points were used for killing game such as elk or deer at their seasonal peak, and that the microlithic weapons were used for hunting expeditions where the prey could not have been predicted easily and any one of a broad range of species might have been encountered.

The amalgam of specialized tactics and diversified ones proved to be highly effective in damping the oscillations of the forest zone. Perhaps the best evidence that the food supply had become more reliable comes from the postglacial residential sites, which often appear to be as permanent and complex as early farming villages. Indeed, some of the sites, such as Sarnate in Latvia, would seem to have been occupied throughout the year. At other sites excavators have found the remains of dwellings with floors, built-in hearths, storage facilities and partitions for dividing the internal space. Such structures are far removed from the usual picture of hunting-and-gathering life. Their relative permanence suggests a fundamental change had taken place in the lives of some foragers. Rather than moving constantly in search of food, the group now stayed in one place for much (if not all) of the year, sending small parties into the surrounding territory to carry out specific hunting or gathering tasks.

Along with greater residential permanence came social developments that ultimately lifted the foraging groups beyond the rudimentary social level that is generally attributed to hunting-and-gathering communities. One of those developments was the emergence of distinctions in wealth, power and status among members of the community. Social differentiation is best observed in the postglacial cemeteries, of which some 20 have now been excavated in Europe. The largest and the richest of them is Oleneostrovski Mogilnik (Deer Island Cemetery) on Lake Onega in northern Russia. Deer Island, which dates from the sixth millennium B.C., was excavated by Nina Nikolaevna Gurina of the Institute of Archaeology in Leningrad in the 1950's. Recently John M. O'Shea of the University of Michigan and I analyzed the findings from Deer Island to see what they could tell us about social differentiation.

O'Shea and I were particularly interested in the distribution of artifacts among the graves. We wanted to know whether the distribution was unequal and was systematic enough to indicate the existence of status groups within the society. We found that status varied along several social dimensions. The first dimension was wealth. Three groups could be distinguished on the basis of the ornaments recovered from their graves. Many of the ornaments are necklaces made from animal teeth, and the type of teeth can serve as an index of material wealth. The graves of the wealthiest group contained necklaces made of the most precious commodity: bear teeth. An intermediate group were buried with necklaces made of beaver incisors or elk teeth. The poorest group were buried with no ornaments at all.

At least two other forms of social differentiation were evident among the graves at Deer Island. One is the existence of special roles within the community. One set of graves, occupied exclusively by male skeletons, contained rich deposits of bone spearpoints but no other artifacts. These would seem to be the remains of men who had special responsibilities as hunters conferred on them by the community. Another set of graves, which contained an unusual abundance of ornaments, may have belonged to specialists in ritual. Unlike the ordinary members of the community (who were buried lying flat in horizontal graves), the putative ritual specialists were interred in vertical shaft graves. Furthermore, some distinctions of wealth and status appear to have been passed from generation to generation, as is shown by the presence of graves containing impressive burial equipment along with the skeletons of children.

The conclusion that postglacial foraging societies underwent a process of social differentiation that lifted them above the egalitarian level is reinforced by some evidence that is quite generally distributed throughout the forest zone. The northern forests are replete with objects carved of wood, bone or stone dating from the postglacial period. The subjects most often depicted are forest animals, such as elk or bear, or waterfowl; anthropomorphic figurines are also found. One intriguing aspect of the carvings is that as the postglacial period went on more of them were made and they were more widely distributed throughout the forested regions.

For the most part these carvings had no clear utilitarian application. Their function seems to have been social: in a culture without money, beautifully carved objects, which required considerable time and skill to make, served as tokens of value. Having them in one's possession was impressive to other members of the group. Barbara Bender of University College London has

Touch the new Buick Riviera. And it touches you.

The name Riviera has always represented a very special sort of automobile.

And now there is a new inheritor of the Riviera legend, so advanced it turns driving into a sensory experience of sound, light, motion and touch. It even establishes a unique two-way communication link between driver and car.

Touch the screen of the Graphic Control Center and it responds with information, it provides control of audio and climate control systems, it even tells you about systems that need your attention.

Touch the controls of the Power Comfort Seat in the T Type and make all the conventional adjustments, plus thigh support, side support, lumbar support and head-restraint height adjustments.

Now reach for the wheel and find yourself thrust down the road by the power of an advanced 3.8-litre engine with sequential-port fuel injection.

Feel the handling of Riviera with disc brakes and fully independent suspension.

Sense a finely calibrated change as the self-leveling suspension resets itself for different load weights.

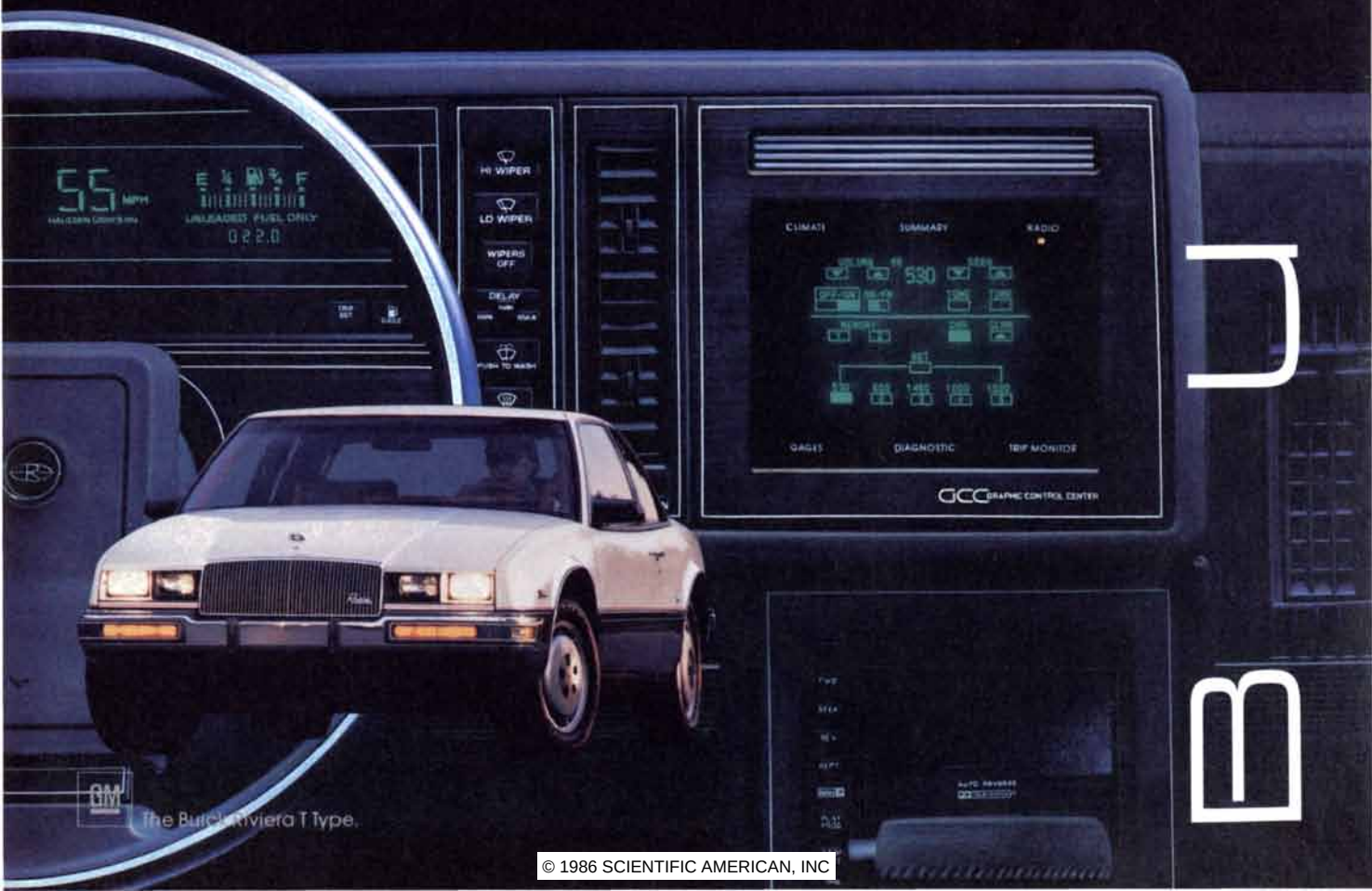
Take in the whole sensory experience of this Buick. It is an experience that comes from an amalgam of high technology, utter luxury and raw, pure, scalp-prickling performance.

Buckle up and see your Buick dealer for a test touch of the automobile that touches back. Riviera. For more information, a test drive or a brochure, call:

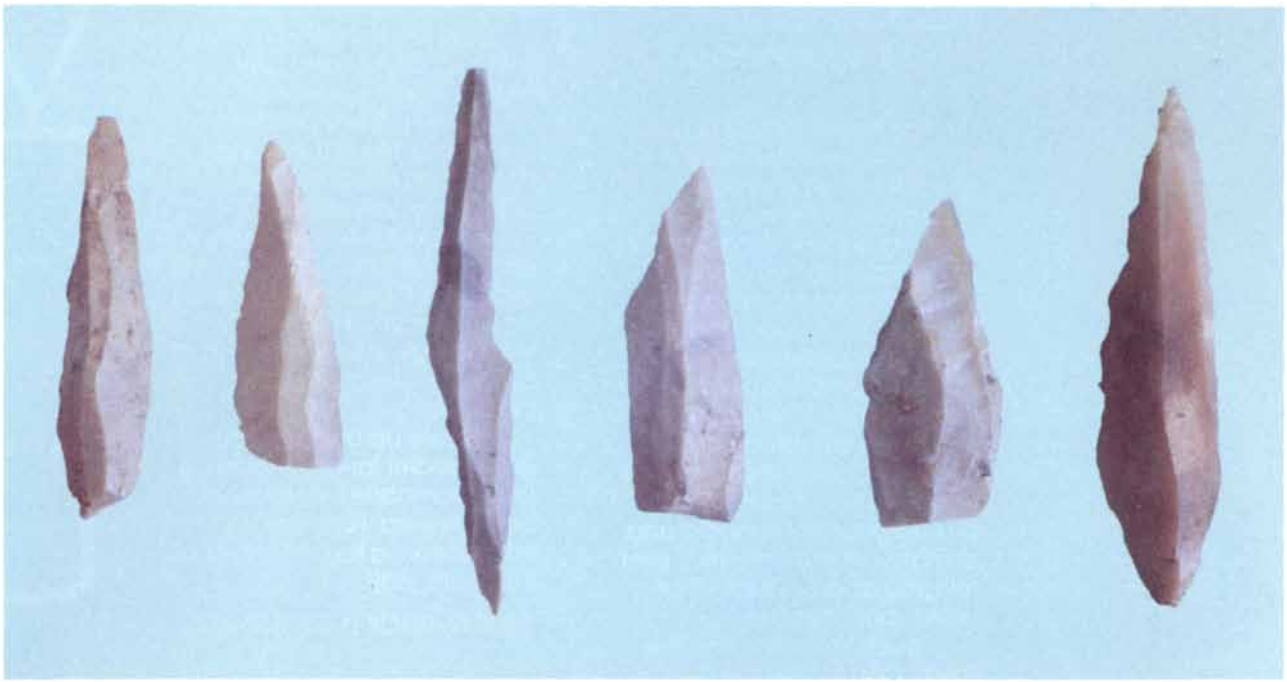
1-800-86-BUICK (1-800-862-8425).

Wouldn't you really rather have a Buick?

BUICK



The Buick Riviera T Type.



MICROLITHS are fragments of stone that were sharpened and set into bone, wood or antler handles. The microliths shown here come from Deercar in Yorkshire. They were made in the eighth or early seventh millennium B.C. Such small blades, which are found at almost all postglacial sites, were the main component of a multipurpose technology. On sighting prey a hunter could remove the stone chip from the handle and reshape it, thereby reequipping himself for killing a wide range of game. Microlithic technology was probably employed when a hunting party set out without a particular prey in mind. It complemented the specialized technology, which was appropriate for carefully planned expeditions. The microliths are from the collection of the Sheffield City Museum in England.

pointed out (in a slightly different social context) that the spread of value tokens may well reflect increased social competition. O'Shea and my colleague Paul Halstead have suggested that such tokens may also have served as a form of "social storage," by being accumulated in times of abundance and then exchanged for food or other basic commodities in times of need. Thus for an individual the tokens might not only have expressed wealth and status but also have provided additional insurance against fluctuations in the food supply.

The emergence of a social hierarchy completes the sketch I have been drawing of the postglacial foraging groups. Social differentiation (and the attendant increase in social competition) complemented the combination of specialized and multipurpose technology, food storage and permanent or quasi-permanent settlements. That constellation constitutes a complex adaptation to the ecology of the Eurasian forest zone. In my view the complex foraging adaptation ought to serve as the defining characteristic of the Mesolithic period. The Mesolithic is the time between the end of the Upper Paleolithic cultures of the Pleistocene glaciation and the introduction of agriculture, which marks the beginning of the Neolithic period. The Mesolithic is sometimes given either a chronological definition or a relatively simple economic one that defines it as the period of postglacial foraging. Neither definition captures the full achievement of Mesolithic social development as represented by the complex hunting-and-gathering adaptation.

Since the complex foraging adaptation was so successful in stabilizing the food supply, the question naturally arises of why it ultimately gave way to farming in every major region of Europe. According to the traditional view, agriculture spread rapidly and uniformly from the Near East, either carried by immigrants or diffused by cultural imitation. Recently Albert J. Ammerman of Princeton University and Luigi L. Cavalli-Sforza of Stanford University have constructed a mathematical model based on those assumptions. What is known about the diffusion of farming into southeastern and central Europe fits the model quite well. In northern and eastern Europe, however, the fit is not nearly as good.

Along the northern Atlantic coast and in the forests of eastern and northern Europe, farming spread much more slowly than the model would have predicted. Indeed, on the whole it appears that the forward progress of agriculture did not resemble a steady march. Instead it seems to have been more like a series of dashes, punctuated by long periods of waiting. For example, agriculture was being practiced in northern Germany by about 4500 B.C. After that, however, there was a long delay, and not until 1,200 years later did it appear in Denmark and southern Sweden. According to the model, farming should have taken hold in the forest zone of eastern Europe between 5000 and 4000 B.C.; in actuality it was adopted between two and three millenniums later. Farming spread particularly late to regions such as the eastern Baltic area, northern Spain and the northern Ukraine, which are rich in aquatic resources. It is notable that those are the environments where the complex foraging adaptation was most highly developed.

When one examines the archaeological record on a smaller geographic scale, it becomes clear that in some parts of the forest zone agriculture was adopted even later than a general survey of Europe would suggest. Some scholars have been tempted to take the first bone of a domesticated animal or the first cereal grain found at a prehistoric site as evidence for agriculture. Yet such items could easily have been obtained by trading. It makes much better sense to take the consistent appearance of domesticated species as the criterion for the transition from foraging to agriculture. In some areas

Here are 128 reasons to buy at Elek-Tek, not to mention the fastest delivery anywhere.

Save 30% to 43%

off Manufacturer Suggested Retail prices on

EPSON • OKIDATA • STAR • TOSHIBA • DIABLO •

IOMEGA BERNOULLI BOX (Modular)

1. 10 meg ½ height Drive for IBM-PC/XT/AT & compatibles $1675
2. 20 meg ½ height Drive for IBM-PC/XT/AT & compatibles 2335
3. 40 meg. (20 + 20) ½ Height Drive for IBM PC & compatibles . . 3075
4. 10 meg cartridges for above (3 pak special) 125
5. Head Cleaning Kit 69

Interfaces priced separately

EPSON FX 85

16. LX 80 . . . . . $ CALL
17. RX 100+ 299
18. FX 85 370
19. FX 286 550
20. LQ 800 parallel 590
21. LQ 1000 parallel 699
22. DX10 Daisy Wheel 10CPS. . . . 230
23. DX20 Daisy Wheel 20CPS . . . 320
24. Comrex CR420 400CPS . . . 900

Huge Discount on TOSHIBA

OKIDATA star

SPECIAL PURCHASE!
Diablo P-38 Dot Matrix 400 CPS
Mfr. Sugg. Ret. $1495
Elek-Tek Price $600

SPECIAL PURCHASE!
Diablo D-36 Daisywheel 35 CPS
Mfr. Sugg. Ret. $1495
Elek-Tek Price $450

HEWLETT-PACKARD

Lower Prices on 6 Hewlett-Packard "Best Seller" Calculators

- HP11C $41
- HP12C 72
- HP15C 72
- HP16C 86
- HP41CV 126
- HP41CX 179

70. Optical Wand 91
71. Card Reader 139
72. Printer 275
73. Ext. Memory Module 60
74. Ext. Function Module 60
75. Time Module 60
76. HP-IL Loop Module 95
77. Digital Cassette Drive . . 400
78. Printer/Plotter (HP-IL) . . 329
79. HP-9114A Disk Drive 575

HARD DRIVES

SEAGATE

Internal

Hard Disk Subsystems

6. **20 MEGABYTE $500**

Each system includes:

- Seagate Drive • Controller
- 2 Cables • Documentation
- 1yr limited warranty
- For use in IBM PC, XT, AT&T, most compatibles

PRINTER BUFFERS

7. **DDP-64** 64K Par./Par. $80

QUADRAM Microfazers

8. **QRMP8** 8K Parallel/Parallel 125
9. **QRMPS8, SP8 & SS8** Parallel/serial, serial/parallel & serial/serial 139

POWER PROTECTORS

10. **PowerMite PG600-S9** 6 outlet Surge Suppressor . . $30

TRIPPLITE

11. **ISOBAR 4** 4 outlet Surge Suppressor . . . 42
12. **ISOBAR 8** 8 outlet Surge Suppressor . . . 50
13. **Line Conditioner** (4 outlet) 135
14. **Backup Power Supply** (425 watt) 360
15. **Backup Power Supply** (1000 watt) 850

PRODUCTS FOR IBM-PC

25. **Intel Above Board** All Products . . . . . $ CALL
26. **Generic Multi** Multifunction Board, 64K 135
27. **Generic Multi 384K** Multifunction Board, 384K 175
28. **AST Six Pak +** Multifunction Board, 64K 225
29. **AST Six Pak + (loaded)** Multifunction Board, 384K 290
30. **Quadram Quadboard** Multifunction Board, 0K 195
31. **Quadram Quadboard** Multi. Board, 64K/384K 215/260
32. **Orchid Tech.** PC Turbo 186 350
33. ECCELL 0K 399
34. ECCELL Daughterboard 135
35. ECCELL I/O Board 99
36. **Hercules** Monochrome Card 299
37. **Generic version of above** 100
38. **Hercules Color** Color Graphic Card 155
39. **Generic version of above** 90
40. **Novation 490605-1** 2400BPS inc. Mite Software 540
41. **Novation 490592-1** 1200B Int. Modem w/ Mite software . 165
42. **Novation 490603** ½ Card Modem 2400BPS No Software 425
43. **Hayes 1200** External modem 380
44. **Hayes 1200B** Internal modem w/software 359
45. **Hayes 2400** External modem 599
46. **Hayes 2400B** Internal modem w/software 525
47. **US Robotics Courier 2400** Ext. 2400B Smart Modem 460
48. **Toshiba ND 04D** ½ ht. DSDD Disk Drive 100

PRODUCTS FOR APPLE

49. **U—Print IIc-16K** Par Interface ext. for IIc 85
50. **Grappler PLUS** Graphics Interface 80
51. **Buffered Grappler** 16K Buffer (Expand to 64K) . . 150
52. **Kensington System Saver** Fan and Surge Suppr 69
53. **Super Cooling Fan** Fan and Surge Suppr. 29
54. **80 Column Card** for Apple II 55
55. **Gamma** Disk Controller Card 48
56. **Gamma II** SS Disk Drive for II, IIe, IIplus . . 125
57. **Gamma IIc** SS Disk Drive for IIc 125
58. **Rana ELITE 2** DSSD Disk Drive 290
59. **Rana ELITE 3** DSDD Disk Drive 380
60. **ERAM 80** 80 col; 64K RAM for IIe 115
61. **Novation Apple Cat II** 300 Baud Modem 200
62. **Novation 212** 300/1200 Baud Modem . . . 359
63. **Nov 49059112** Mod. & Soft. for MacIntosh . . . . CALL
64. **Hayes Micromodem IIe** 300 Baud Modem 135
65. **Hayes Smartmodem IIc** 300 Baud Modem for IIc . . . 175
66. **Wico 501030** Analog Joystick 36
67. **Amdek Monitors** . . . . . CALL
69. **Zenith Monitors** . . . . . CALL

Hewlett-Packard QUAD MEMORY MODULE

$25 Mfr. Sugg. Ret. $75
Upgrades the HP41C to the power of the HP41CV

SLIMLINE Shirtpocket Styled Power Packed Programmable LCD PROBLEM SOLVERS

80. **HP-11C** Scientific 41
81. **HP-15C** Scientific 72
82. **HP-12C** Financial 72
83. **HP-16C** Programmer 86

HP-ThinkJet

HP-2225B
$350

84. **HP-71B Computer** . . . . $ 375
85. **HP-82400A** Series 70 card reader 115
86. **HP-82401A** HP-IL interface 95
87. **HP-82402A** 4K memory module 60
88. **HP-82700A** 8K memory module 150

HEWLETT PACKARD

IBM is a trademark of International Business Machines Corp.

HUGE SAVINGS ON DISKETTES

	Dysan	maxell	3M	ELEK-TEK	XIDEX
3½" SSDD	$ 20.00	$ 19.00	$ 21.00	—	—
DSDD	27.00	26.00	29.00	—	—
5¼" SSDD	15.00	12.00	12.00	—	—
DSDD	19.00	15.00	15.00	$ 12.50	$ 7.00
SSDD96TPI	24.00	22.00	23.00	—	5
DSDD96TPI	33.00	29.00	25.00	—	Box
5¼" DSDDHD (For IBM AT)	33.00	28.00	30.00	—	Min.
8" SSDD**	22.00	23.00	22.00	—	
8" DSDD**	26.00	27.00	24.50	—	

**Unformatted

Call for Quantity pricing for 10 boxes or more.

3M DATA CARTRIDGES

3M	1	10	25
122. DC100	$13.50	$13.00	$12.00
123. DC300A	17.50	17.00	16.00
124. DC300XL	20.50	20.00	19.00
125. DC300XL/P	21.50	21.00	20.00
126. DC600A	23.00	22.50	21.50
127. DC1000	14.50	14.00	13.50
128. DC2000	19.50	19.00	18.50

CALL TOLL FREE 800-621-1269 EXCEPT Illinois, Alaska **CANADIAN TOLL FREE 800-458-9133**

ELEK-TEK, inc. 6557 N. Lincoln Ave. Chicago, IL 60645 (312) 631-7800 (312) 677-7660

Corp. Accts. invited. Min. Ord. $15.00. Visa or MasterCard by Mail or Phone. Mail Cashier's Check, Mon. Ord. Personal Check (2 wks. to clear). Add $4.00 1st item, Alk. HI, PR, Canada add $10.00 first item, $1.00 ea. add'l. Shipping & Handl. Shipments to IL addresses add 7% tax. Prices subj. to change. WRITE for free catalog. RETURN POLICY: Defectives Only. Most products replaced within 30 days of purchase with identical merchandise only. Computer and large peripherals replaced only when defective on arrival (within 3 work days of delivery). Other problems covered by the warranty. ALL ELEK-TEK MERCHANDISE IS BRAND NEW, FIRST QUALITY AND COMPLETE. Delivery subject to availability. DUNS #09-718-0617

the difference between the two dates is considerable. In Finland, the eastern Baltic region and many parts of Russia isolated remains of domesticated species were deposited as much as 1,000 years before systematic farming began.

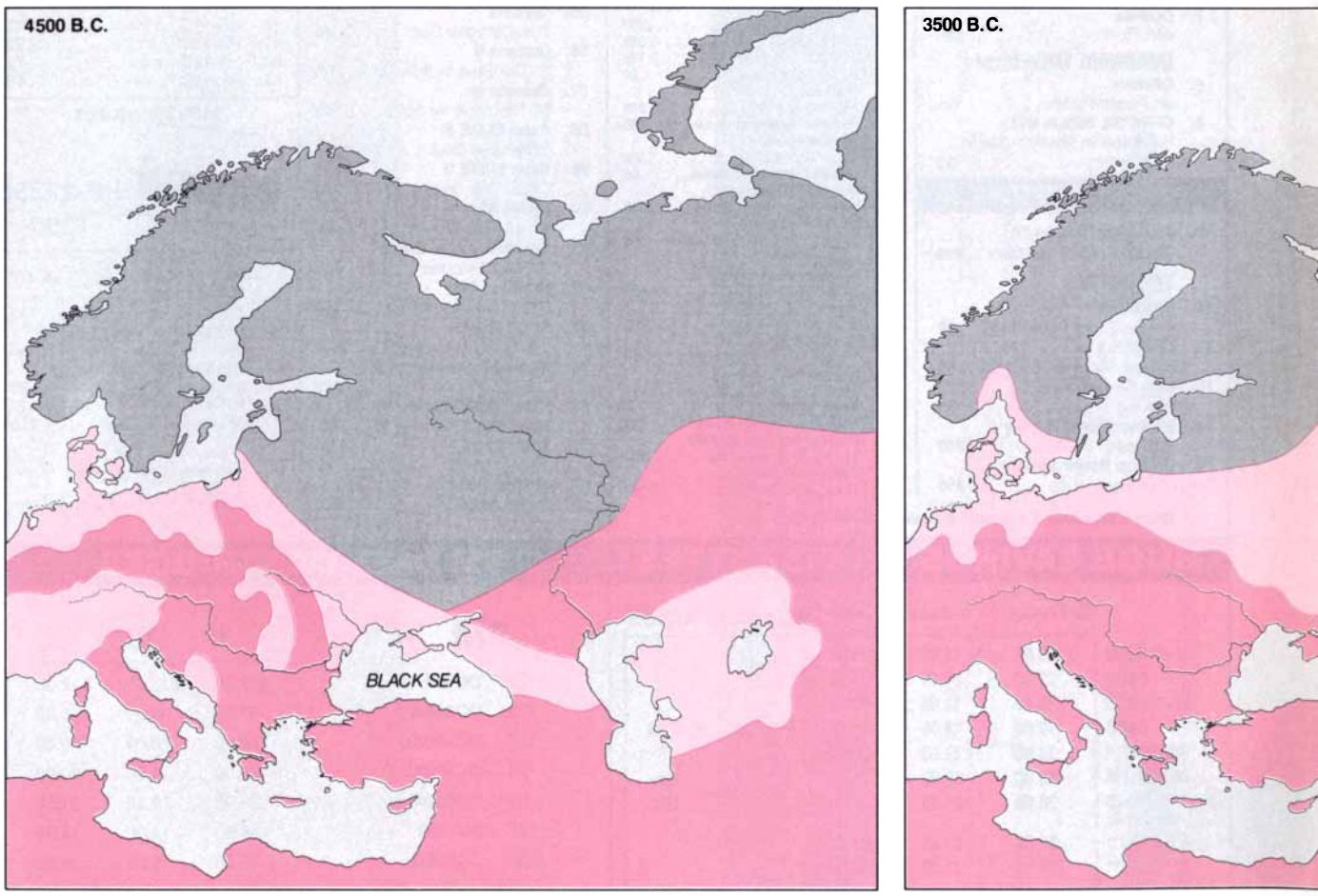
What emerges from an examination of the frontier zone dividing farmers and foragers, then, is a picture of a long period when agriculture was available to the foragers but was not put into practice. Such an image raises the question of why the foragers were so slow to adopt agriculture. I think the answer is that farming was not necessarily advantageous, particularly for communities that specialized in exploiting aquatic resources.

In such cultures adopting agriculture would have required abandoning a large investment in specialized technology and forms of social organization adapted specifically to complex foraging. Furthermore, the introduction of farming would have caused scheduling problems, because in the Eurasian forest zone most cultivated species are harvested in the fall, which is a peak period for hunting. In any case, the incentive for taking up farming must have been small, at least initially, because there was no pressing need to increase productivity. The specialized foraging strategy was capable of supporting relatively high population densities and of absorbing some fluctuations in resources. Therefore the explanation for the transition to agriculture must be sought elsewhere than in the inherent superiority of farming.

One plausible explanation is that farming was introduced to compensate for failures of specific resources. The essence of the complex foraging economy was the balance between the few species that were intensively exploited by means of specialized technology and the larger group of prey exploited on a "chance encounter" basis using multipurpose tools. As long as the delicate balance was maintained, the complex foragers were able to reduce risks and maintain productivity. The disappearance of a single intensively exploited resource, however, could destroy the balance and throw a foraging community into crisis. Several areas of Europe have yielded evidence of such crises. In Denmark the vanished resource was the oyster; in southern Finland, the seal; in northwestern Russia, the water chestnut along with wild cattle and the wild pig. Farming appears to have been taken up in all three regions as a stopgap intended to overcome the crisis caused by resource failure.

The integrity of foraging economies may also have been disrupted by competition with groups that practiced agriculture. In the frontier zone foraging and farming communities must have competed for a variety of resources; among the most significant of them was territory. The end of the Pitted Ware culture provides an example of how loss of territory could bring with it dire consequences for hunters-and-gatherers.

In about 2600 B.C. Pitted Ware bands occupied the interior of middle Sweden along with the coastal areas and the islands between Sweden and Finland. That broad territory supplied



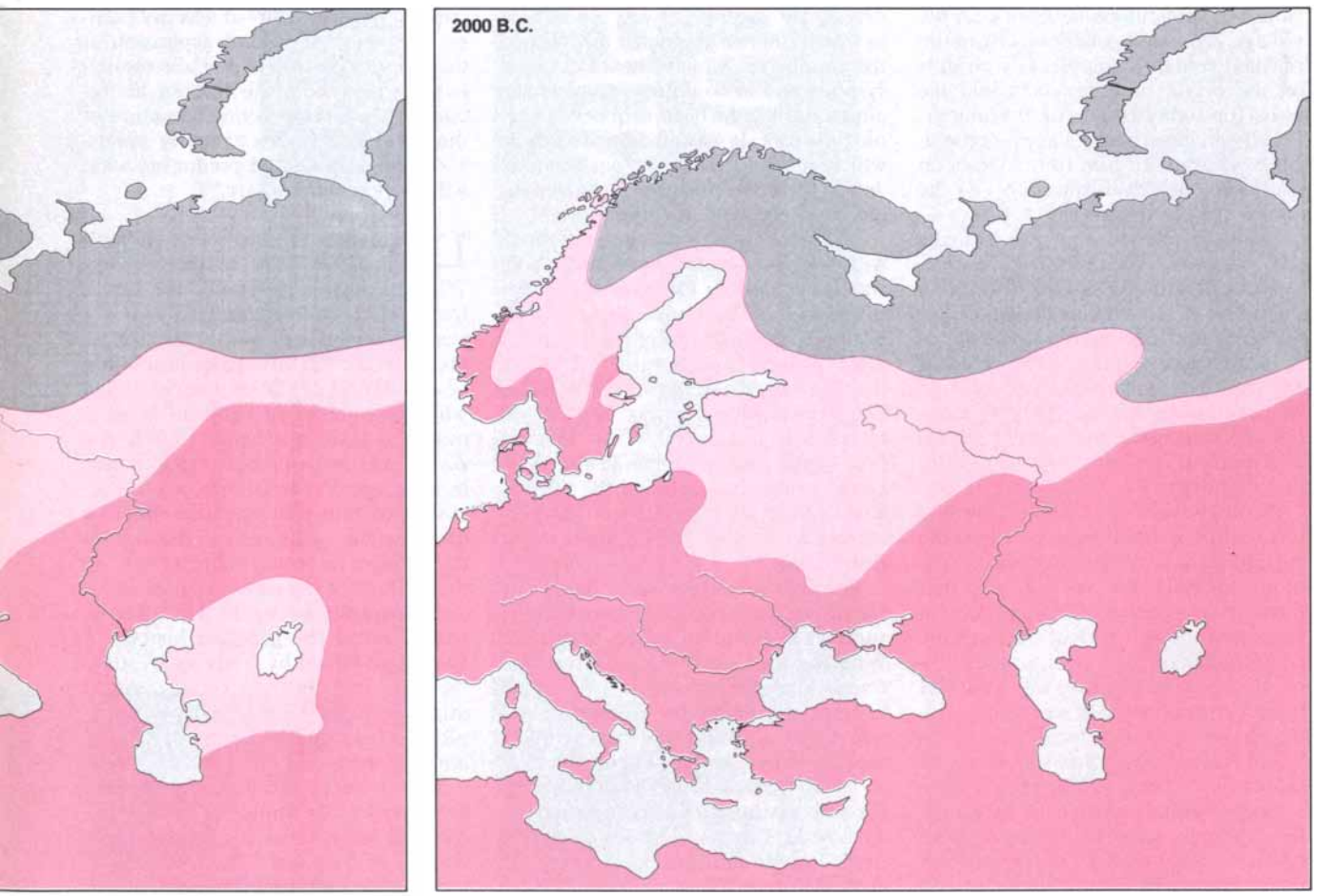
SPREAD OF FARMING from the Near East through Europe was delayed in the forest zone. On each map the area in dark color corresponds to the territory where agriculture had taken hold and the gray area corresponds to the territory of hunting-and-gathering groups. The area in light color represents an "availability zone," where hunting-and-gathering groups had had contact with farmers and knew of the techniques of agriculture but had not yet adopted farming. In the northern Atlantic region, including Denmark and

a variety of terrestrial game together with the seals on which the Pitted Ware groups specialized. By 2300 B.C. farming had been introduced in the Swedish interior. As a result Pitted Ware communities were restricted to the coast, and their dependence on seals increased greatly. As a result of the destruction of their balanced economy Pitted Ware groups were poorly equipped to cope with the fluctuations in the seal population that accompanied a change in climate between 2000 and 1800 B.C. The foraging economy soon disappeared, and the people of the Pitted Ware culture were probably absorbed by the farming population of the interior.

Another form of competition may also have had a role in the transition to agriculture: competition within a foraging society. As I have described above, in the late Mesolithic period there seems to have been an increase in social differentiation and in striving for status, much of it directed toward objects serving as value tokens. In the frontier zone where farming and hunting communities interacted, foragers would have become aware of objects that were exotic to them and on which they may have put great value. For example, in the third millennium B.C. trade among a chain of farming communities might have brought well-fired pottery or polished axes north from central Europe to southern Scandinavia. Foragers there might have seen such goods and coveted them as symbols of status. To get them, the foragers would have had to intensify their search for the kind of goods they could offer in return, such as beaver pelts. Quite soon, however, the foragers would deplete the stock of beaver or reach the point where further intensification of hunting became uneconomic. In such cases they might well have turned to farming's greater productive capacity as a means of satisfying the desire for acquisition.

As I hope the preceding paragraphs make clear, I do not think the transition to farming took place everywhere in the forest zone for the same reason. On the contrary, in each region any of several mechanisms could have been operating, and the immediate causes varied from region to region. It follows that the attempt to find a single, universal explanation for the adoption of agriculture is bound to fail. In my opinion a more fruitful program of research would be to examine the specific circumstances of the transition in many smaller areas.

Such an examination might change our picture of foragers in several respects. I think it would be found that the persistence of the hunting-and-gathering adaptation has been underestimated. Furthermore, the notion that the replacement of foraging by farming constitutes a linear march of progress might be amended. In chronology and in function the development of the complex hunting-and-gathering adaptations was parallel to that of agriculture. Sooner or later the parallelism will be recognized and the two forms will be seen for what they were: attempts to reduce risks and increase the productivity of the environment by intensifying labor in the world left behind by the retreat of the glaciers.



the nearby lands, agriculture was available beginning in about 4500 B.C. but was not adopted until shortly before 3000 B.C. Knowledge of agriculture spread through eastern Europe north of the Black Sea between 5000 and 4000 B.C., yet farming was not adopted there until 2500 B.C. or even later. These were the regions where the complex foraging cultures flourished. The long delay before farming was taken up suggests farming was not necessarily advantageous to groups practicing the complex hunting-and-gathering adaptation.

Darwin as a Geologist

He tends to be viewed now as a biologist, but in his five years on the Beagle *his main work was geology, and he saw himself as a geologist. His work contributed significantly to the field*

by Sandra Herbert

One tends to see Charles Darwin today essentially as a biologist, concerned with the origin and evolution of plant and animal species. Yet it was as a geologist even more than as a naturalist that he took part in the famous voyage of the *Beagle* from 1831 to 1836, and it was as a geologist that he saw himself in the years immediately following the voyage. His contributions to geology were significant. His meticulous fieldwork on the voyage produced collections of material that remain valuable. His insights on the origin of coral reefs laid the basis for today's view of the matter. Finally, in large part it was his geological work that led him to his views on evolution, or "transmutation" as the vocabulary of the time had it.

Darwin's interest in geology during the voyage did not stem from long involvement with the subject. Indeed, at the outset of the voyage he was better prepared to collect insects and invertebrate animals; he had been an amateur entomologist since his youth and had been well trained in the study of invertebrates during the two years he spent as a medical student at the University of Edinburgh.

Geology took precedence on the voyage for several reasons. First, the Admiralty wanted a man trained in geology to make the voyage, and that must have affected the way Darwin presented himself to Robert FitzRoy, commander of the vessel, in their initial discussions. FitzRoy was aware of the economic interests served by geology. In 1830, after an earlier mission to South America with the *Beagle,* he had pointed out that a geologist or a mineralogist would be able to ascertain, for example, whether the mountains of Tierra del Fuego did indeed contain metal. Similarly, when it came time to commission the *Beagle* for its second voyage, Francis Beaufort, the hydrographer to the British navy, stressed the importance of a geological investigation of coral islands. He noted that they "occasionally afford excellent landlocked harbours," and he also pointed out that they were currently of interest to geologists.

The second reason for Darwin's bending toward geology during the *Beagle*'s voyage had to do with the particularly attractive state of the field at the time. The 1830's were a golden decade for geology. It was the decade in which English geologists established the Cambrian, Silurian and Devonian systems and in so doing completed in global outline the basic sequence of geologic strata. It was also the decade in which most English geologists abandoned the attempt to synchronize biblical and geologic history.

The person who put Darwin in touch with the new developments in the field was John Stevens Henslow, professor of botany at the University of Cambridge (which Darwin had entered after leaving Edinburgh), fellow of the Geological Society in London and Darwin's mentor. It was in 1831, at Henslow's instigation, that Darwin first began serious study in the field. Until then he had avoided the subject, having been disappointed in the geological lectures of Robert Jameson at Edinburgh.

Through arrangements made by Henslow, Darwin did fieldwork in the summer of 1831 with Adam Sedgwick, Woodwardian professor of geology at Cambridge. This was the work that led Sedgwick to define the Cambrian system. So it was that Darwin, only a few months before embarking on the *Beagle,* was trained in field methods by the most distinguished geological fieldworker in England, who was at the time engaged in his greatest work. On arriving home from his travels with Sedgwick, Darwin found the letter containing the offer to join the *Beagle* awaiting him.

The third reason Darwin chose geology was surely its suitability to his own makeup. Geology in the 1830's combined the traditional emphasis of natural history on the particularity of things with an openness to theorizing—a combination that suited Darwin's abilities and tastes extremely well. In a short autobiographical note written soon after the voyage he recalled a childhood desire "of being able to know something about every pebble in front of the hall door—it was my earliest and only geological aspiration at that time." His talent for abstraction led him to write of his interest in "recording the stratification and nature of the rocks and fossils at many points, always reasoning and predicting what will be found elsewhere."

Darwin's decision to focus on geology came early in the voyage. What seems to have caused the forces from within and without to coalesce was the prospect of authorship. As he prepared for the voyage he noted that he was afraid of "being overwhelmed with the number of subjects which I ought to take into hand." While the *Beagle* was in its first port, São Tiago in the Cape Verde Islands, a plan occurred to him that had the effect of ordering his activities for the rest of the voyage: he would write a book on the geology of the points visited. In his autobiography he wrote that after he had deduced the geologic history of São Tiago Island by studying its rocks "it then first dawned on me that I might perhaps write a book on the geology of the various countries visited, and this made me thrill with delight."

That Darwin should have thought of himself as the author of a major geological work at such a young age—he was nearing his 23rd birthday—is surely a sign of a powerful ambition at work. It was, however, a realistic ambition, because Darwin knew that the geology of South America, particularly the southern extremities of the conti-

nent, was so little explored that his work could not fail to be original. Moreover, he had the influence and inspiration of two recently read masterpieces before him: Alexander von Humboldt's *Personal Narrative* of his travels in South America, which was heavily geological but dealt mainly with the northern portion of the continent, and Charles Lyell's *Principles of Geology,* the first volume of which had appeared in 1830 and appealed to Darwin for its bold style and provocative content.

With future publication in his mind, then, Darwin set out from Cape Verde to collect thoroughly in geology and to observe as completely as he could. As he wrote to Henslow in August, 1832, "I have endeavoured to get specimens of every variety of rock, and have written notes upon all." Eventually his work led to a three-part publication on the geological findings of the voyage: *The Structure and Distribution of Coral Reefs* (1842), *Geological Observations on the Volcanic Islands Visited during the Voyage of H.M.S. Beagle, together with Some Notices of the Geology of Australia and the Cape of Good Hope* (1844) and *Geological Observations on South America* (1846).

In the course of collecting during the voyage and of writing up his findings Darwin became a geologist. That was how he viewed himself and that was how others viewed him. At the end of the voyage he referred to himself in a private notebook as "I, a geologist." His identity had formed around his ambition.

It was typical for him that his ambition operated on more than one level. Each day he pursued his original intention to systematically collect and observe. Here his methodical habits stood him in good stead. As a sportsman he had recorded his quarry; as a geologist he was equally diligent in numbering his specimens, recording their locations and keeping up with his note-taking. On quite another level he pondered the significance of his findings and developed the hypotheses that have led to his enduring reputation as a profoundly original thinker.

Darwin's notes from the voyage, which are in the Cambridge library, reveal his daily habits of work. Unlike his peers in England, who often returned each summer to the same site, Darwin usually had to get what he wanted the first time through an area. The notes he made during the voyage on his first geological excursion illustrate his approach. On January 17 and 18, 1832, the stop was at tiny Quail Island (now Ilhéu de Santa Maria) in the harbor of Praia in São Tiago. The notes begin:

"Quail island a small desolate spot lying close to Porto Praya.—Its shape is oval from N. to S. & barely a mile in circumference. There are round Porto

ISLAND OF MOORÉA, which is in the South Pacific near Tahiti, influenced Darwin's thinking about the origin of coral reefs. When the *Beagle* stopped at Tahiti in 1835, Darwin climbed a hill and saw approximately this view of Mooréa (then Eimeo) some 15 miles away. The island is encircled by coral reefs. Darwin speculated that if the island were to sink slowly, a coral atoll would form in its place.

SEA-LEVEL VIEW of Mooréa near the side facing Tahiti shows part of the coral reef. It is the tree-covered islet at the center. Darwin subsequently enlarged his theory of reef formation to encompass three classes of reefs: fringing reefs, barrier reefs and atolls.

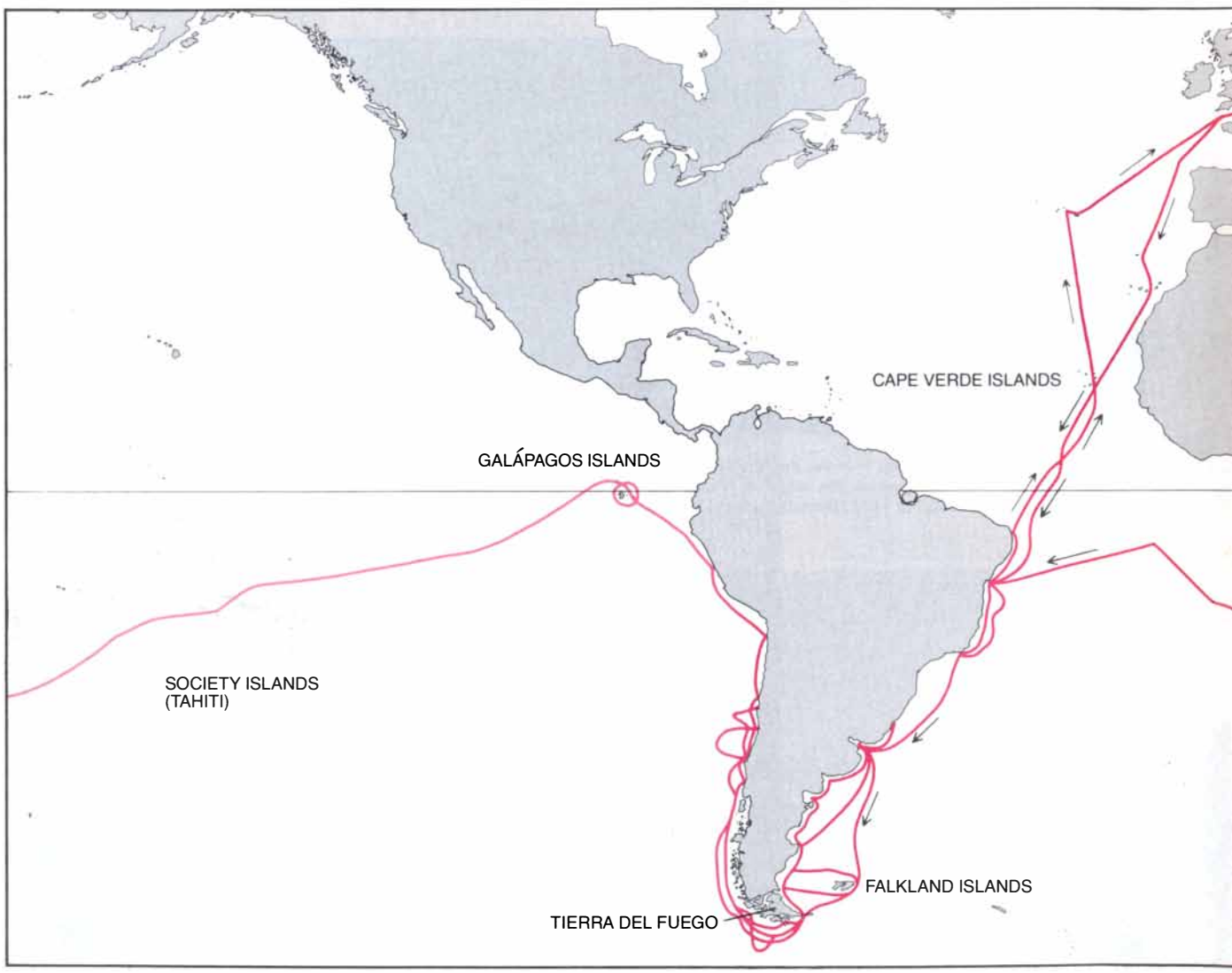
Praya several truncate conical hills: this island may be considered one of the set, but with sea instead of a sandy valley at its foot.—The washing of the sea round its base affords a good section & I thought by studying this island attentively, I might find a good keystone for the neighbouring country.—I will begin by the lowest beds & describe the whole succession with minuteness."

Accompanying the detailed description of the beds were rock samples. In examining his specimens Darwin relied on simple devices: a lens for magnification; an acid bottle for testing alkalinity; a blowpipe for testing the reaction of a specimen to heat; a goniometer for measuring angles, and a magnet. He was careful to label each specimen with a number keyed both to his specimen book and to his notes. As a result virtually his entire collection was and still is usable, unlike many collections rendered fruitless by poor accompanying data.

Besides describing beds and collecting rocks, Darwin speculated on the origin of what he saw. Of Quail Island, for example, he concluded: "Upon the whole I should think it probable [he crossed out *clear*] that after the marine beds had been quietly deposited on the inferior Volcanic rocks, a sheet of melted matter was spread over them: that the whole mass was then raised, since which or at the time there has been a *partial sinking*."

Notwithstanding Darwin's skills as a collector and an observer, it is on his theorizing that his reputation as a geologist rests. What set him apart from most of his colleagues was his ability to treat abstract questions of structure and causation together with questions of detailed observation, and to move back and forth between the two levels with ease. His theoretical impulse, accompanied by a willingness to take on controversial issues, was displayed in his earliest notes on Quail Island. Moreover, his ambition as a geologist grew as the voyage stretched to five years from the two initially planned. With the contemporary geological literature as a guide, he allowed himself to become more speculative in his approach to the subject.

Darwin's growth as a theorist is best illustrated by examining the two largest themes he pursued: vertical movements of the earth's crust and transmutation, or the evolution of species. The first theme came to him from his reading. He picked an idea from his predecessors and then expanded it beyond the dimensions they had given it. The second theme, which he pursued later in the voyage and more circuitously, also came from his predecessors, but in this case he had to overcome their in-



VOYAGE OF *BEAGLE* from 1831 to 1836, with Darwin aboard as scientific investigator, is charted on the basis of records kept by Robert FitzRoy, the commander of the ship. In a book FitzRoy published in 1839 on two of his voyages he included a similar map

fluence, since most of them had viewed evolution negatively.

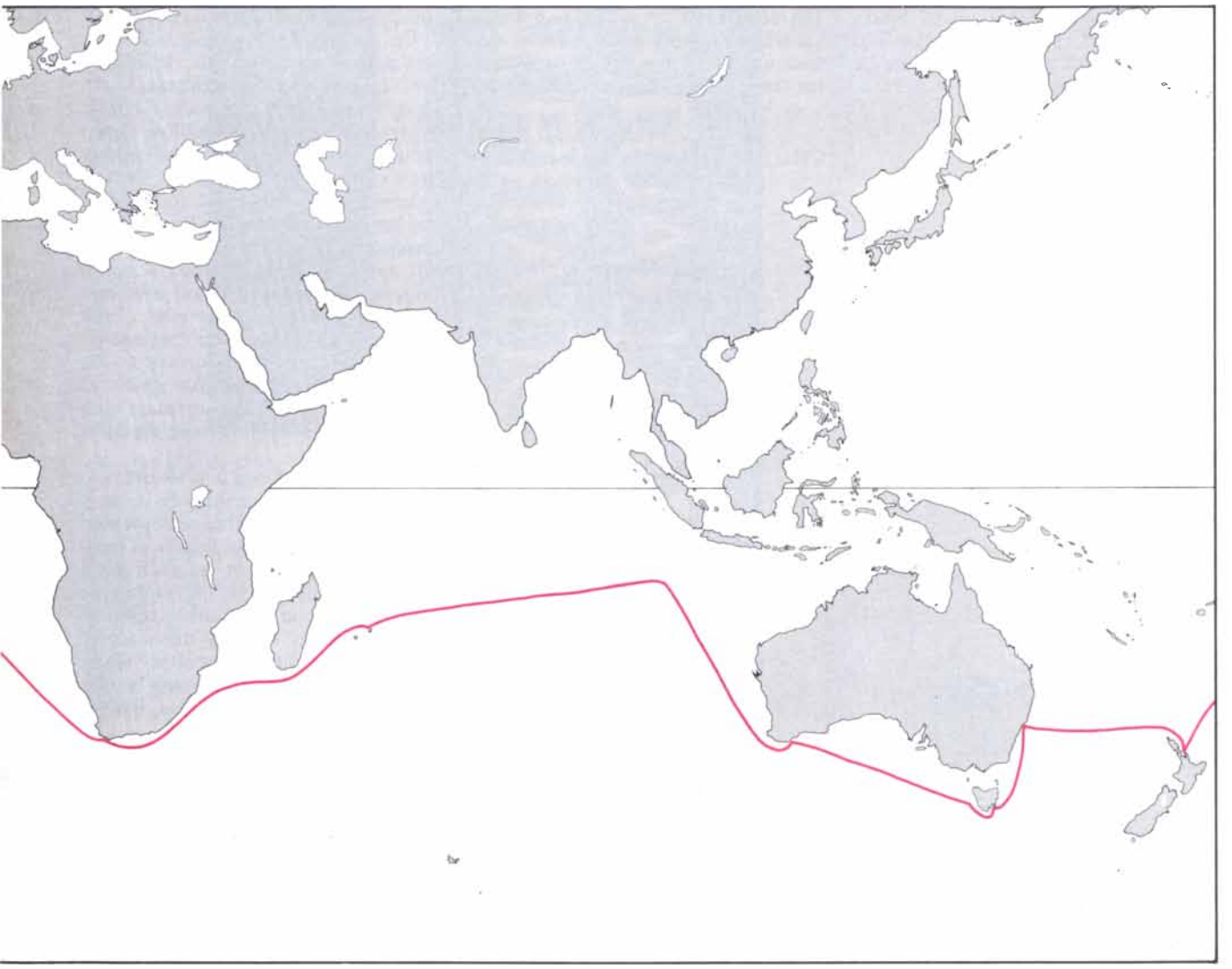
Darwin's first theme was set in his mind by the end of the voyage. In his "Red Notebook" he wrote in 1836, on the last leg of the voyage, that the "Geology of whole world will turn out simple.—" He alluded to the notion that world geology was governed by simple upward and downward motion, that is, by the elevation and subsidence of land. He believed a balance of these two forces was the key to determining the main features of the earth's surface.

In taking up this theme Darwin was continuing work begun by other geologists. The one person whose thinking influenced him most profoundly was Lyell. Following in a tradition established in the late 1700's by the Scottish geologist James Hutton, Lyell argued in *Principles of Geology* that the earth's surface is perpetually in motion. Like most of the contemporaneous geologists, he believed positing changes in the level of the land was a simpler way to account for the presence of marine remains inland than positing worldwide changes in sea level. Unlike many of his peers, he believed changes in land level took place gradually. He opposed the view held by many geologists that the earth went through periods of stability followed by periods of upheaval.

Darwin approached the phenomena he saw on the voyage with Lyell's views in mind. Whereas Lyell's fieldwork was done in fairly confined areas of Europe, Darwin did his over several oceans and continents. His different experiences affected the way he applied Lyell's ideas. The major change he made in Lyell's notion of elevation and subsidence was to work with it on a larger scale, that of continents rather than that of a confined area.

When he explored what is now Argentina, Darwin was impressed by the immensity of the plains and by the persistence for great distances along the coast of raised beaches of nearly identical height. He inferred from these observations the operation of an elevatory force acting over a large geographic area and working so slowly that it did not disturb the continuity in the height of the plains and the raised beaches. When he got to the Andes, he was impressed by their scale compared with the mountains in Europe. Near Valparaíso in Chile he found evidence of a rising coastline: the presence of recent marine shells well inland and above the high-water mark on the coast. On February 20, 1835, he witnessed an earthquake at Valparaíso accompanied by a land rise of several feet.

All these observations combined to lead him to the conclusion that the entire continent of South America was



that he described as a "general chart shewing the principal tracks of H.M.S. *Beagle*." Because FitzRoy's concern was with the "principal tracks," the map does not show all the ship's movements in such cases as the numerous trips to and from the Falkland Islands.

rising. He saw the creation of the Andes and the occurrence of earthquakes and volcanic eruptions in South America as side effects of the motion.

In his next application of Lyell's notion of elevation and subsidence Darwin made his most lasting contribution to geology outside his work on evolution. This was his theory of the origin of coral reefs, published fully in 1842 as *The Structure and Distribution of Coral Reefs*.

It may seem puzzling that, as the geologist H. W. Menard once put it, "at times, Darwin seems to be concerned with the origin of coral reefs only so far as it gives evidence for regional subsidence of the sea floor." There is no puzzle, however, if one realizes that Darwin formulated his theory as a corollary to his understanding of crustal motion. For him as for Hutton and Lyell, the rise of one area of the earth's surface had to be balanced by the fall of another area. Once he had arrived at the conclusion that all of South America, and even presumably all of North America as well, was rising he looked for a corresponding area of subsidence elsewhere. "Does not the great extent of the Northern & Southern Pacifick include this corresponding Area?" he asked.

Once again Lyell had given Darwin the fundamental points of his argument. In *Principles of Geology* Lyell suggested the Pacific Ocean as an area of subsidence. He also offered an explanation for the shapes of various kinds of coral reef. His argument was that since corals do not grow above the surface of the water and reef-forming corals grow only at shallow depths, reefs must form on preexisting submarine platforms such as subsided mountaintops or volcanic craters. Coral islands must rest on the former, coral atolls on the latter.

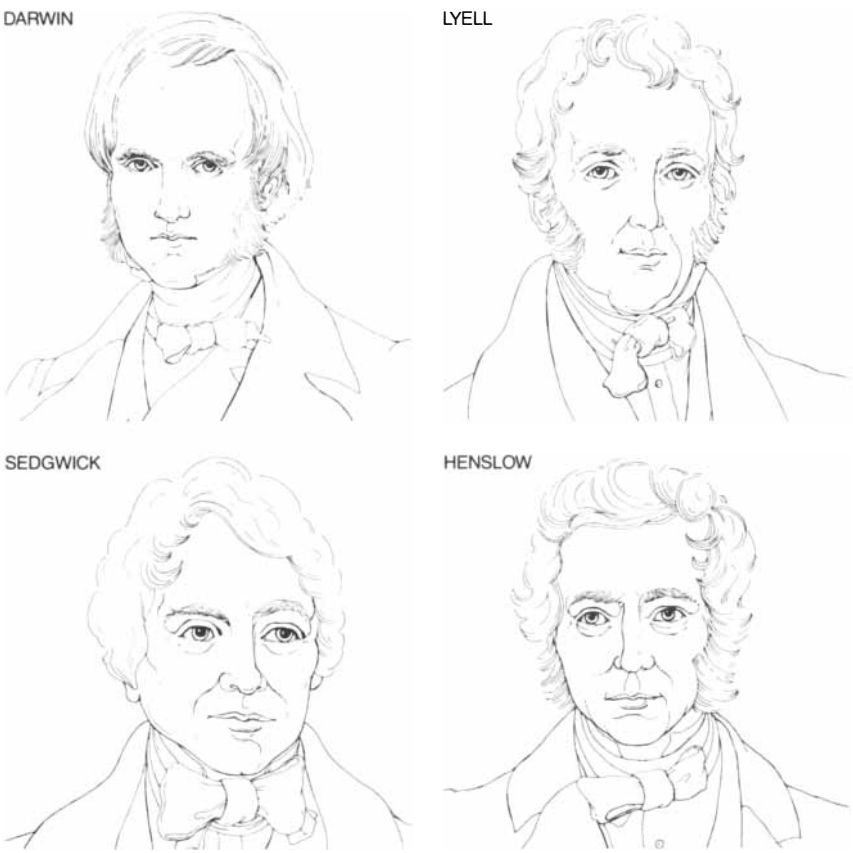
Darwin thought Lyell's explanation of the origin of coral reefs was ingenious but believed Lyell had taken insufficient account of the vastness of the Pacific Ocean. In his view such a large area must have experienced correspondingly extensive subsidence. Therefore it was erroneous to posit the existence of barely submerged mountains and volcanoes simply to account for the presence of coral reefs in mid-ocean. It was better, Darwin thought, to imagine reefs forming on platforms they themselves had built as the floor of the ocean subsided below them. In Darwin's view—which was not finally confirmed until the 1950's—coral reefs might be of great thickness.

What of the various forms of reef? Here Darwin's earliest sighting of a coral reef was instructive. The *Beagle* stopped at Tahiti in November, 1835. Darwin climbed some 2,000 to 3,000 feet up a mountain and gazed toward the beautiful reef-encircled island of Eimeo (now called Mooréa). As he looked at it rising out of its "glassy lake" (as he wrote later) he thought, "Remove the central group of mountains, and there remains a Lagoon Isd." But how could a subsiding mountain leave a ring-shaped reef rather than a disk-shaped island? Because, Darwin maintained, the lagoon that formed where the mountain had once stood held still water and was too shallow to provide a good environment for the growth of reef-forming corals.

On the way back to England, Darwin generalized his insight. He divided reefs into three classes: fringing, barrier and atoll. Fringing reefs (slightly separated from land by shallow water) would occur where the reef simply grew outward from the shore. Darwin believed such reefs must form when the land is stable or rising. Barrier reefs (separated from the land by a deep channel) would be formed when a fringing reef subsided. Coral grew best in full surf, and the outer edge of the fringing reef would form the barrier. The channel would be formed as the interior landmass sank. An atoll reef would be formed if the landmass sank entirely but slowly; otherwise the coral would drown.

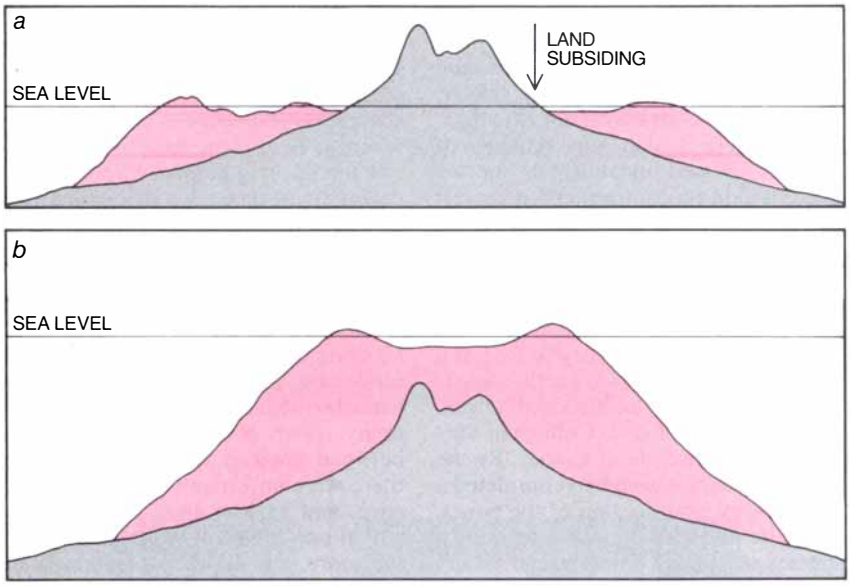
Interestingly, Darwin's theory entailed the conclusions that one form of reef evolved into another, that the rate of motion of the ocean floor was sufficiently gradual to not drown a good number of reefs and that by noting the presence of one or another class of reefs one might identify the areas of the ocean floor that were either rising or subsiding. (Fringing reefs would indicate stability or elevation, barrier and atoll reefs subsidence.)

Although Darwin's theory of the structure and distribution of coral reefs has subsequently been challenged at specific points, it is nonetheless held in high regard. Judged as a theory, it remains a model of simplicity and explanatory power.



DARWIN AND HIS GEOLOGICAL MENTORS are shown as of about 1840. Charles Lyell's *Principles of Geology* influenced Darwin strongly. Adam Sedgwick, professor of geology at the University of Cambridge, gave Darwin experience in geological fieldwork in the summer of 1831. John Stevens Henslow was professor of botany at Cambridge, but he was also active in geology and put Darwin in touch with recent developments in the field.

Darwin came to his second grand ambition in geology more circuitously, but once he was there the work led ultimately to his most famous pub-



DARWIN'S THEORY on the formation of coral reefs rested on the concept that corals grow on land that is slowly sinking. Here corals (*color*) build up in shallow water over land that was once the above-water perimeter of a sinking island (*a*). Eventually the entire island subsides and the corals form an atoll: a reef around a lagoon that lies over what was the highest part of the island (*b*). Darwin explained correctly that corals do not grow in a lagoon because they need moving water in order to flourish, and lagoon water is too still.

lication: *On the Origin of Species.* He did not begin the *Beagle* voyage with the intention of writing on that subject. Nor did his peers in science expect it of him.

In hindsight one can identify certain circumstances in his background that prepared him to pursue such a goal, notably the tradition of free thought that existed in his family and the example of his grandfather, Erasmus Darwin, who had written a developmental cosmology. Nevertheless, for Darwin to espouse a transmutationist hypothesis (holding that new species arise from preexisting ones) in the 1830's required him to break with the majority of his peers. His new ambition required a rupture with tradition.

Ironically Lyell was once again the major influence on Darwin. Lyell had devoted most of the second volume of *Principles of Geology* to an attack on the transmutationist hypothesis of the French writer Jean Baptiste de Lamarck. In the course of attacking the hypothesis, however, Lyell gave his readers a clear statement of it. When Darwin received the volume in November, 1832, while the *Beagle* was at Montevideo, he had at hand an up-to-date and well-informed presentation of both sides of the controversy. Hence early in the voyage he knew what the tradition of interpretation regarding species was and what breaking with it would require. Moreover, in arguing against transmutation Lyell had presented an enormous amount of information on the geographic distribution of species and on nature's economy, or what would now be called ecology. Later Darwin would draw on this information for his own purposes.

The central issue that had brought the question of transmutation to the fore in geology was extinction. In 1796 the French naturalist Georges Cuvier had stated that the bones of a large fossil found in Paraguay belonged to an unknown animal. He invented a name for the new form: *Megatherium* (large beast). Because the continents had been fairly well explored by his day and no such animal had ever been sighted, Cuvier concluded that the animal had become extinct. Before long the remains of other large animals without living representatives (*Mastodon* and *Palaeotherium*) were identified. By the early 19th century the case for extinction was so strong that it was regarded as fact.

Darwin confronted the extinction issue as both collector and interpreter. As collector he took pains on the voyage to pick up specimens of the great extinct land animals. Eventually he had partial specimens of about a dozen different forms.

As interpreter Darwin's main concern with fossils during the voyage was to determine their geologic association, that is, their position in a particular stratum and the relation of that

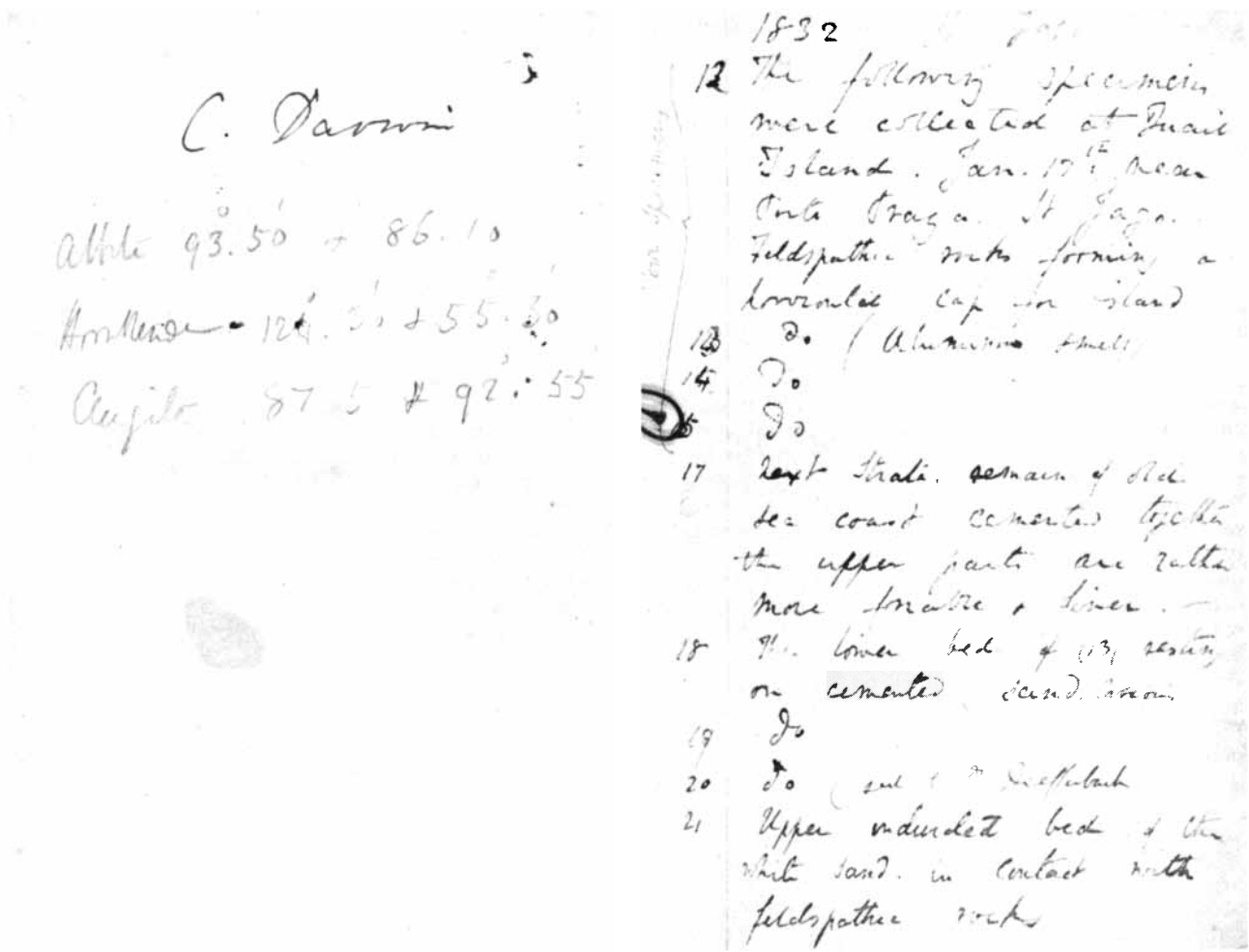
bed to the surrounding beds. His aim was to determine the conditions under which the bones had been deposited and from that knowledge to gain some understanding of the causes of extinction. He was aware of the split in English geology between those (including Lyell) who believed species became extinct gradually and those (including William Buckland, reader in geology at the University of Oxford) who believed species died out rather suddenly, presumably as a result of dramatic changes in living conditions arising from, say, a flood or an abrupt change in temperature.

Darwin decided in favor of the gradualist view, partly because of the gradual nature of the land elevations he thought were taking place in South America. As he wrote while he was in Chile in 1835: "With respect to the *death* of species of Terrestrial Mammalia in the S. part of S. America I am strongly inclined to reject the action of any sudden debacle.—"

Then, like Lyell, he went on to assert that extinct species must have been replaced by new species. His reason may surprise the present-day reader, because it was drawn from theology. "The fitness which the Author of Nature has now established," he argued, would be contradicted if extinct species were not replaced. In 1835, then, he believed he had theological grounds for considering how replacements might arise.

When Darwin returned to England in October, 1836, he turned his fossil bones over to Richard Owen, anatomist at the Royal College of Surgeons and a disciple of Cuvier. By the end of January, Owen had completed a preliminary examination of the bones. Some he declared belonged to known species of extinct animals, but others were completely new. What particularly impressed Darwin were the connections Owen pointed out between the fossil bones and various living animals peculiar to South America. One specimen, later named *Toxodon Platensis,* had affinities to the modern capybara. Another, later named *Macrauchenia Patachonica,* had affinities (eventually questioned) to the modern llama, or guanaco. Lyell himself called attention to this striking pattern of geologic succession in the forms of South American animals soon after Owen had pronounced on Darwin's specimens.

To a present-day reader such a pattern might seem to demand a transmutationist explanation; in 1837 an insuperable difficulty intervened. However many points of similarity there were between ancient and modern forms, there were still many points of difference, and various lines existed along which one might draw affinities. Furthermore, no series of intermediate forms connecting ancient and modern forms were known. Hence the fossil evidence, although it was suggestive, was insufficient to support a transmu-



FIRST TWO PAGES of Darwin's notebook recording the geological specimens he collected suggest the care he took in recording data. Quail Island (now Ilhéu de Santa Maria) in the Cape Verde Islands is where Darwin began collecting geological specimens and where, he wrote in his autobiography, it "first dawned on me that I might perhaps write a book on the geology of the various countries visited." He was meticulous about numbering his specimens and putting the number alongside his comments about the specimen.

tationist hypothesis. Another element was needed.

That element was supplied by facts drawn from the present geographic distribution of species. The advantage of living forms was that they had fine gradations. Darwin, relating how he adopted a transmutationist hypothesis, wrote in his autobiography he was struck "by the manner in which closely allied animals replace one another in proceeding southwards over the Continent" and also "by the South American character of most of the Galapagos archipelago, and more especially by the manner in which they differ slightly on each island of the group."

One of Darwin's key examples of the replacement of a species by a closely allied species involved the two rheas: the greater or common rhea, found from northeastern Brazil to the Río Negro in central Argentina, and the lesser or Darwin's rhea, found in the Patagonian lowlands. His key example of island-by-island variation in form was provided by the Galápagos mockingbirds. Three islands, Charles, Albemarle and Chatham, each had its own mockingbird, barely distinguishable from one another by eye but nonetheless separate species according to the ornithologist John Gould.

Rheas, mockingbirds and similar examples made transmutation seem possible to Darwin. Clearly it was easier to imagine a living form transmuting into a slightly different living form than it was to imagine the passage of an extinct animal into a related but still markedly different living animal. Once he had grasped the transmutation potential of living species, Darwin quickly read it back into the past. As he observed in the "Red Notebook" in March, 1837: "The same kind of relation that common ostrich [the large rhea] bears to [the small one]: extinct Guanaco to recent: in former case position, in latter time... being the relation." In the notebook he went on to consider the possibility that one species can change into another. He had become a transmutationist.

Thus did Darwin's second grand ambition in geology crystallize in March, 1837, when he drew an analogy comparing the distribution of species in space to their distribution in time. This same ambition was fully realized in 1859, when he published *On the Origin of Species*. In the intervening 22 years he remained faithful to his vocation as a geologist, but he worked primarily as an author, writing up material for which the fieldwork had already been done. Both as field-worker and author, however, his legacy as a geologist was assured.

THE AMATEUR SCIENTIST

Wire that "remembers" its shape is put to work running an engine

by Jearl Walker

A piece of metal that has been slightly deformed by a blow or a strain will return to its original shape when the pressure is released. Most metals will not do that, however, if the deformation is severe. An exception is found in certain unusual alloys that retain a "memory" of their original shape even after severe deformation. When an object made of a "shape memory" alloy is deformed and is later heated above a certain temperature, it returns to its original shape. Its returning motion may be strong enough to push or pull something else. In other words, the re-forming object can do work.

Can a useful engine be based on this principle? Several workers have tried to build thermal engines with working parts made of shape-memory alloys. A few engines of this type, based on simple designs, are now on the market. David Johnson of Berkeley, Calif., studies such engines and their alloys. Recently he sent me a manuscript describing how to construct two of the engines. He has also offered to supply sample wires made of the shape-memory alloy called Nitinol. (The name comes from the nickel and titanium of the alloy and from the U.S. Naval Ordnance Laboratory, where the alloy was first studied.)

A shape is implanted in Nitinol wire by annealing it: the wire is held in the wanted shape and brought to a high temperature. Suppose the wire is formed into a U shape and heated. After it cools it is twisted into a new shape, say an S. When the wire is then heated above a transition temperature (which depends on its composition), it returns to the U shape.

Nitinol can also develop a memory of its shape when it is below the transition temperature if it is put through a "training" procedure. The wire is repeatedly heated to a point above the transition temperature, allowed to regain its annealed shape (say a U) and then cooled below the transition temperature. Each time it is cooled it is forced into a particular second shape, say the S again. After several cycles of heating and cooling the wire begins to remember the S. Thereafter it will form into an S whenever it is cooled below the transition temperature. This memory changes somewhat the memory of the annealed shape; it is also weaker than the memory of the annealed shape.

One of Johnson's designs for a Nitinol thermoturbine is shown on the opposite page. Pulleys *A* and *B* are fixed to a shaft. *A*'s radius exceeds *B*'s by 30 percent. Two other pulleys, called idlers, are immersed in water baths, one hot and one cold. Running around the array of pulleys is a continuous length of Nitinol wire that has been coiled in a tight helical spring. Pulleys *A* and *B* are grooved to keep the spring from slipping.

When Johnson's apparatus is properly made, the spring pulls itself around the pulleys, forcing the shaft on which *A* and *B* are mounted to turn rapidly. The motion continues for as long as the water baths are kept hot and cold.

The motion results from the changes in the Nitinol as it passes through the water baths. In the hot water it heats above the transition temperature and contracts, pulling strongly on *A* and *B*. In the cold water it cools below the transition temperature and expands. Now its pull on *A* and *B* is weak.

Consider the forces on *A* and *B* as the spring contracts in the hot water. Note that the spring is wrapped around the pulleys in opposite directions. Hence the force on *B* is toward a counterclockwise rotation and the force on *A* is toward a clockwise one. The competition is not a stalemate because the rotation of an object depends on the torque, rather than just the force, on the object. Torque is the product of the force and a lever arm. The lever arm for the pulley is its radius. Since *A*'s radius is larger than *B*'s, the torque on *A* dominates the rotation. The shaft turns clockwise.

The spring in the cold-water bath also pulls on *A* and *B*. Again *A* has the larger torque because of its larger radius. The result should be a counterclockwise rotation of the pulleys and the shaft. Remember, however, that the forces from the expanded spring in the cold water are weaker than the forces from the contracted spring in the hot water. The net motion of the pulleys and the shaft is therefore clockwise.

Several precautions are necessary to ensure that the device functions properly. *A* and *B* must have steep grooves (at an angle of at least 60 degrees) to prevent the spring from slipping. The shafts of all the pulleys must be mounted on ball bearings to reduce friction. The idler pulleys must be far enough from *A* and *B* to stretch the sections of spring between them by about 50 percent each.

The Nitinol wire is .02 inch in diameter. Make the spring by wrapping the wire in a tight spiral around a metal rod 3/16 inch in diameter. Heat the wire and rod to a temperature of about 520 degrees Celsius, at which the array glows a dull red. This step anneals the wire, so that whenever it is heated above its transition temperature by the hot water, it tends to regain the shape it had on the rod. When the engine is run and the spring passes through the cold-water bath several times, it becomes trained to a second shape: a looser helix. Thereafter the spring has a memory of two shapes, the tight helix when it is above the transition temperature and the loose helix when it is below it.

The memory in Nitinol arises from the types of crystals that develop in it. They are sensitive to temperature and external stress. When the wire is above its transition temperature, it is in what is called the parent phase. If it is cooled below the transition temperature, large groups of atoms rearrange themselves to form a crystal structure sometimes called quench-induced martensite. (The name is misleading because it implies a need for rapid cooling.)

If the cooled wire is put under strain by stretching, some of the martensite is transformed into what is sometimes called stress-induced martensite. Greater stress leads to more transformation. If the wire is then heated above the transition temperature, the

atoms change from the martensite formations to the parent phase. All these transformations are said to be diffusionless, meaning they involve a large-scale reordering of the atoms rather than a diffusion of individual atoms through the crystal structure.

The parent phase consists of cubic crystals and the martensite phase consists of needlelike crystals that collect in small domains. Within a domain the crystals are aligned. One explanation for the different types of martensite derives from the relative alignment of the domains. When Nitinol under no stress is cooled from the parent phase, the martensite domains are randomly oriented. This arrangement characterizes quench-induced martensite. If the wire is then stretched enough to distort its shape, some of the domains are reoriented so that they point along the direction of stress. The reorientation allows the wire to accommodate for the change in its length. Greater strain leads to more extensive alignment and thus to a greater amount of this type of martensite.

When the wire is heated above its transition temperature, the needlelike crystals of martensite change into the cubic crystals of the parent phase. Because the cubic crystals and the needlelike ones have different shapes, they do not fit into the same space. The new crystals form under strain and have strain energy stored in them. They move to change their positions and orientation in order to relieve the strain and release the stored energy. Where the martensite domains are randomly oriented this attempt to relieve strain does not result in any net movement of the wire, but where the domains are aligned the relief from the strain forces the wire to undo the effects of the previous stretching and deformation. By this action the wire regains its annealed shape.

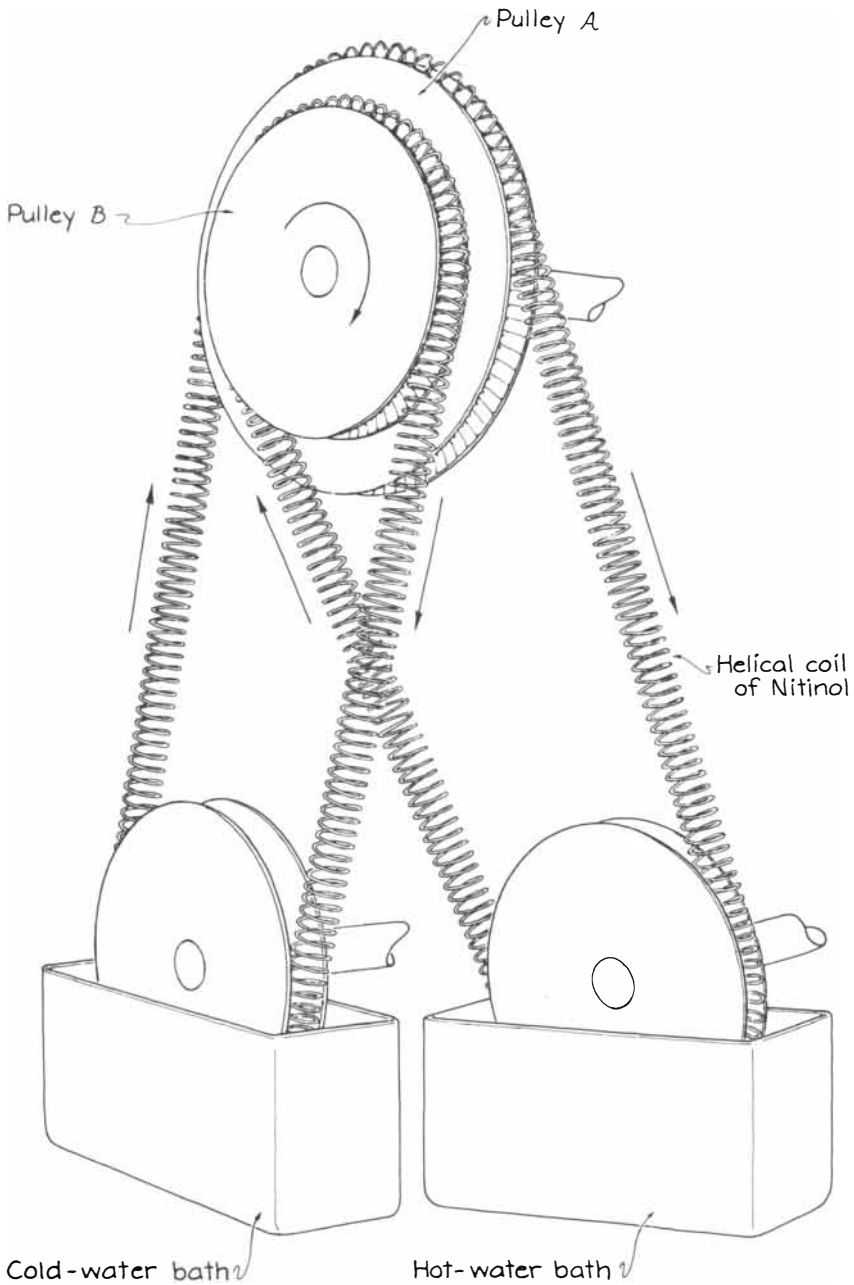
One generates work by reheating the wire while it is under a moderate strain produced by external tension. As the wire passes the transition temperature, conversion to the parent phase again yields cubic crystals under strain. As before, they tend to relieve the strain by returning the wire to its annealed shape. This time the wire must push or pull against the external tension, thereby doing work. The energy for the work the wire does comes from the strain energy stored in the cubic crystals as they form.

At present no one fully understands why Nitinol can be trained to remember a second shape when it is below the transition temperature. One hypothesis is that the repetition of heating and cooling and the related crystal transformations introduce defects into the crystal structure. Each defect creates stress in the crystal. When a trained wire is cooled below its transition temperature, the stresses arising as a result of the defects force the wire into its trained shape.

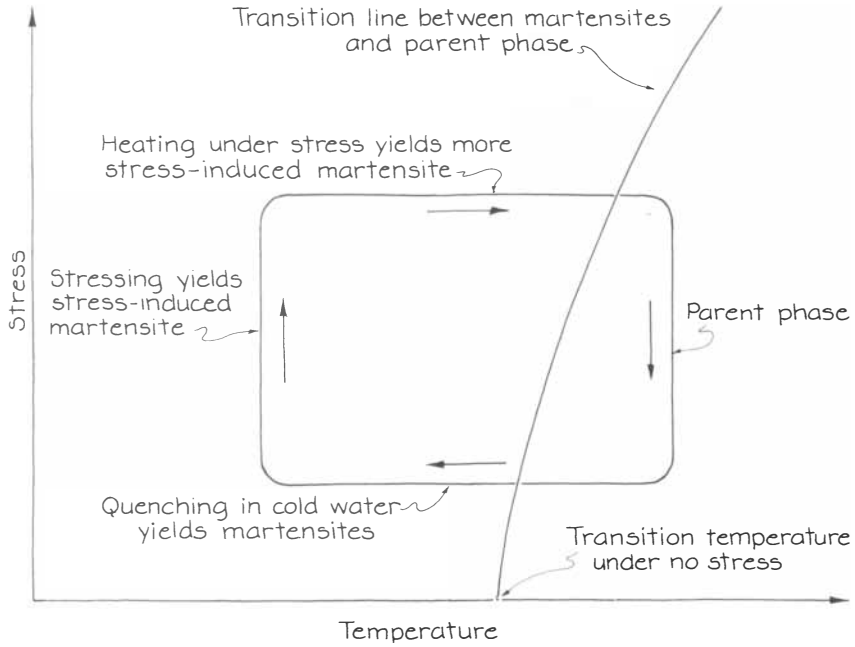
Another explanation for training assumes that martensite cooled from the parent phase can exist in several different formations. Suppose there are only two such formations and one of them grows faster if the wire is under compression when it is cooled whereas the other grows faster if the wire is under tensile stress.

If the alloy is under no stress when it is cooled, the two types of crystal grow at the same rate. Equal amounts develop in the cooled wire. They each introduce strain into the wire, but since they are equal in amount, the wire is under no net strain that would alter the shape it had before it was cooled.

If the wire is under compression when it is cooled, one type of crystal grows faster than the other; the rela-

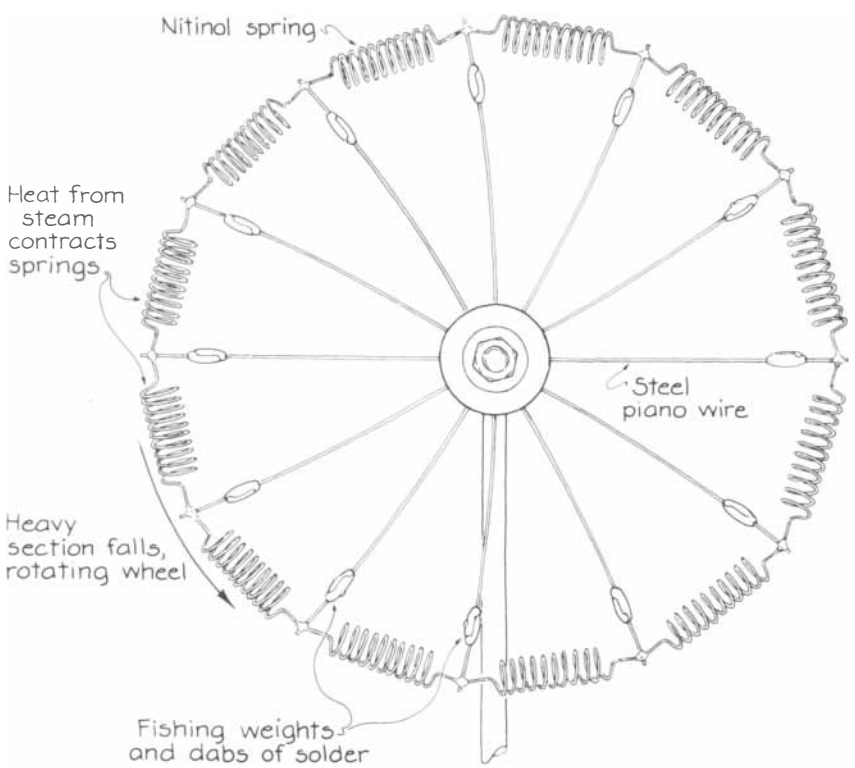


David Johnson's Nitinol-wire engine



Phase diagram for Nitinol in an engine

tion is reversed if the wire is under tensile stress when it is cooled. Suppose the wire is constrained to be bent when it is cooled. The compression type of martensite forms along the inner part of the bend and the elongation type forms along the outer part. When the wire is repeatedly heated and cooled while it is being held in a bent form, the two types begin to form automatically in their places in the bend. Eventually the process becomes so automatic that the constraints are no longer required. When the wire is cooled, one side develops the compression type of martensite and the other side develops the elongation type. The strain introduced by the crystals on each side of the wire bends the wire. It is then trained.



Nitinol and the unbalanced wheel

To follow the transformations of crystal structure in Johnson's engine I mentally labeled a small section of the wire spring as *x*. I also drew a phase diagram of stress v. temperature [*see top illustration at left*]. As *x* travels around the engine, it also moves along the path shown in the phase diagram. When it changes to or from the parent phase, it crosses the line separating the parent phase from the martensite formations. A complete cycle of *x* around the engine corresponds to a complete trip along the path in the phase diagram.

When *x* moves from the cold-water bath to pulley *A*, it is at a low temperature because it has just passed through the cold water. It is under only a moderate strain because the segment of the spring currently in the cold water is expanding. Hence *x* consists of a mixture of quench- and stress-induced martensite. This state corresponds to the part of the path at the lower left in the diagram.

When *x* passes over *A*, the tension on *x* increases because the segment of the spring then in the hot-water bath is contracting. The resulting increase in strain orients more of the martensite domains in *x*. This state corresponds to the part of the path at the upper left in the diagram.

On reaching the hot-water bath *x* is heated above the transition temperature. The martensite transforms to the parent phase. The cubic crystals form under a severe strain and store much strain energy. They relieve the strain and reduce the stored energy by making *x* contract. Because *x* is connected to the rest of the spring, it pulls on the segments of Nitinol wire between itself and the grooved pulleys. The unbalanced torques on the pulleys give rise to rotation.

The success of the engine depends on the ability of *x* and the other sections of the wire to do work when they are in the hot water. If just prior to the bath the tension on *x* is low, few of the martensite domains are oriented. When the needle crystals convert into cubic crystals, the attempt to contract *x* is small. Hence the force on the grooved pulleys is small. You might be tempted to move the idler pulley farther from *A* and *B* to increase the stretch of the spring and the stress on *x*. This step entails a disadvantage, which is that the tension in the spring is then so great that *x* cannot regain its

annealed shape when it is in the hot water. It gradually begins to lose that memory.

When x passes around pulley B, the tension on x relaxes because the segment of the spring then in the cold-water bath is expanding. This state corresponds to the lower-right part of the phase diagram. (I am ignoring the fact that the wire is already cooling when it reaches B because of the surrounding air.) When x reaches the cold water, it is cooled and the parent phase is largely transformed to quench-induced martensite. Some of the parent phase is also transformed to stress-induced martensite because x is still under some tension even though it has expanded. This state corresponds to the lower-left part of the phase diagram. Thereafter x repeats the cycle.

Another engine employing Nitinol consists of a hub to which spokes of steel piano wire are attached [*see bottom illustration on opposite page*]. The hub can be a plastic wheel from a toy. The sections of piano wire are about .03 inch in diameter and seven inches long. Attach them to the hub with epoxy glue. Helically wound sections of Nitinol wire run between the outer points of the spokes. Tie the Nitinol wire to the spokes with short lengths of copper wire. When the sections of the spring are in place, each one is stretched so that its coiled length is increased by about 20 percent. Attach quarter-ounce fishing weights near the outer ends of the spokes. Balance the wheel by adding dabs of solder to the weights. An unbalanced wheel will not turn freely on a horizontal shaft.

To set the engine in motion direct steam from a teapot onto a spring that is on each side of the wheel near the top. Once the spring has been heated above its transition temperature it contracts and pulls on the adjacent spokes. The motion shifts weight into the region being heated, thereby upsetting the balance of the wheel. As the heavy section falls, the wheel rotates. When the heated spring moves out of the steam, it cools and expands, restoring the shape of the spokes. Meanwhile another spring moves into the steam, contracts and pulls weight toward it. The cycle continues in this way. The wheel goes on rotating as long as steam is directed onto the springs. The speed of rotation is determined by the cooling rate of the springs after they leave the steam.

A sample of Nitinol wire can be obtained from Johnson at TiNi Sales, P.O. Box 1431, Lafayette, Calif., 94549–1431; send $3 and a self-addressed, stamped envelope.

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

STOCHASTIC MODELS OF PHYLOGENY AND THE EVOLUTION OF DIVERSITY. David M. Raup, Stephen J. Gould, Thomas J. M. Schopf and Daniel S. Simberloff in *The Journal of Geology,* Vol. 81, No. 5, pages 525–542; September, 1973.

PROBABILISTIC MODELS IN EVOLUTIONARY PALEOBIOLOGY. David M. Raup in *Paleontology and Paleoenvironments: Readings from* American Scientist, edited by Brian J. Skinner. William Kaufmann, Inc., 1981.

SUDDEN CARDIAC DEATH

A DECADE'S EXPERIENCE WITH OUT-OF-HOSPITAL RESUSCITATION. L. A. Cobb, J. Werner and G. Trobough in *Modern Concepts of Cardiovascular Disease,* Vol. 49, pages 31–36; 1980.

PREVENTION OF SUDDEN ARRHYTHMIC DEATH WITH IMPLANTED AUTOMATIC DEFIBRILLATORS. M. MIROWSKI IN *Annals of Internal Medicine,* Vol. 97, No. 4, pages 606–608; October, 1982.

OUT-OF-HOSPITAL RESUSCITATION USING THE "HEART-AID," AN AUTOMATED EXTERNAL DEFIBRILLATOR-PACEMAKER. Mary Heber in *International Journal of Cardiology,* Vol. 3, No. 4, pages 456–458; July, 1983.

SUDDEN CARDIAC DEATH IN THE COMMUNITY. Mickey S. Eisenberg, Lawrence Bergner and Alfred P. Hallstrom. Praeger Publishers, 1984.

PREDICTORS AND PREVENTION OF SUDDEN CARDIAC DEATH. Leonard A. Cobb and Jeffrey A. Werner in *The Heart, Arteries and Veins,* edited by J. Willis Hurst. McGraw-Hill Book Company, 1986.

GLYCOSPHINGOLIPIDS

GLYCOSPHINGOLIPIDS IN CELLULAR INTERACTION, DIFFERENTIATION, AND ONCOGENESIS. Sen-itiroh Hakomori in *Annual Review of Biochemistry,* Vol. 50, pages 733–764; 1981.

HANDBOOK OF LIPID RESEARCH, VOL. 3: SPHINGOLIPID BIOCHEMISTRY. Edited by J. N. Kanfer and S. Hakomori. Plenum Press, 1983.

A REVIEW OF THE IMMUNOGENIC AND IMMUNO-MODULATORY PROPERTIES OF GLYCOSPHINGOLIPIDS. D. M. Marcus in *Molecular Immunology,* Vol. 21, No. 11, pages 1083–1091; November, 1984.

TUMOR-ASSOCIATED CARBOHYDRATE ANTIGENS. Sen-itiroh Hakomori in *Annual Review of Immunology,* Vol. 2, pages 103–126; 1984.

THE CLIMATE OF MARS

MARTIAN METEOROLOGY. Conway B. Leovy in *Annual Review of Astronomy and Astrophysics,* Vol. 17, pages 387–413; 1979.

CLIMATIC CHANGE ON THE TERRESTRIAL PLANETS. James B. Pollack in *Icarus,* Vol. 37, No. 3, pages 479–553; March, 1979.

THE ASTRONOMICAL THEORY OF CLIMATE CHANGE ON MARS. Owen B. Toon, James B. Pollack, William Ward, Joseph A. Burns and Kenneth Bilski in *Icarus,* Vol. 44, No. 3, pages 552–607; December, 1980.

THE SURFACE OF MARS. Michael H. Carr. Yale University Press, 1981.

THE HISTORY OF WATER ON MARS. Steven W. Squyres in *Annual Review of Earth and Planetary Sciences,* Vol. 12, pages 83–106; 1984.

AQUEOUS FOAMS

FOAMS. J. J. Bikerman. Springer-Verlag, 1973.

PHYSICAL CHEMISTRY OF SURFACES. Arthur W. Adamson. John Wiley & Sons, Inc., 1976.

RHEOLOGY OF FOAMS AND HIGHLY CONCENTRATED EMULSIONS, I: ELASTIC PROPERTIES AND YIELD STRESS OF A CYLINDRICAL MODEL SYSTEM. H. M. Princen in *Journal of Colloid and Interface Science,* Vol. 91, No. 1, pages 160–175; January, 1983.

THE CHEETAH IN GENETIC PERIL

THE CHEETAH IS DEPAUPERATE IN GENETIC VARIATION. Stephen J. O'Brien, David E. Wildt, David Goldman, Carl R. Merril and Mitchell Bush in *Science,* Vol. 221, No. 4609, pages 459–462; July 29, 1983.

GENETIC BASIS FOR SPECIES VULNERABILITY IN THE CHEETAH. S. J. O'Brien, M. E. Roelke, L. Marker, A. Newman, C. A. Winkler, D. Meltzer, L. Colly, J. F. Evermann, M. Bush and D. E. Wildt in *Science,* Vol. 227, No. 4693, pages 1428–1434; March 22, 1985.

MORPHOLOGICAL VARIABILITY AND ASYMMETRY IN THE CHEETAH (*ACINONYX JUBATUS*), A GENETICALLY UNIFORM SPECIES. Robert K. Wayne, William S. Modi and Stephen J. O'Brien in *Evolution,* Vol. 40, No. 1, pages 78–85; 1986.

SUPERCONDUCTING TUNNEL DETECTORS IN RADIO ASTRONOMY

MILLIMETER- AND SUBMILLIMETER-WAVE RECEIVERS. T. G. Phillips and D. P. Woody in *Annual Review of Astronomy and Astrophysics,* Vol. 20, pages 285–321; 1982.

INTEGRATED-CIRCUIT ANTENNAS. David Rutledge, Dean Neikirk and Dayalan Kasilingam in *Infrared and Millimeter Waves, Vol. 10: Millimeter Components and Techniques, Part 2,* edited by Kenneth J. Button. Academic Press, 1983.

POSTGLACIAL FORAGING IN THE FORESTS OF EUROPE

RESEARCH AND DEVELOPMENT IN THE STONE AGE: TECHNOLOGICAL TRANSITIONS AMONG HUNTER-GATHERERS. Brian Hayden in *Current Anthropology,* Vol. 22, No. 5, pages 519–548; October, 1981.

THE SIGNIFICANCE OF FOOD STORAGE AMONG HUNTER-GATHERERS: RESIDENCE PATTERNS, POPULATION DENSITIES, AND SOCIAL INEQUALITIES. Alain Testart in *Current Anthropology,* Vol. 23, No. 5, pages 523–537; October, 1982.

TIME BUDGETING AND HUNTER-GATHERER TECHNOLOGY. Robin Torrence in *Hunter-Gatherer Economy in Prehistory: A European Perspective,* edited by Geoff Bailey. Cambridge University Press, 1983.

DARWIN AS A GEOLOGIST

DARWIN, LYELL, AND THE GEOLOGICAL SIGNIFICANCE OF CORAL REEFS. D. R. Stoddart in *The British Journal for the History of Science,* Vol. 9, Part 2, No. 32, pages 199–218; July, 1976.

THE RED NOTEBOOK OF CHARLES DARWIN, edited by Sandra Herbert. British Museum (Natural History) and Cornell University Press, 1980.

DARWIN THE YOUNG GEOLOGIST. Sandra Herbert in *The Darwinian Heritage,* edited by David Kohn. Princeton University Press, 1985.

THE AMATEUR SCIENTIST

ON THE THERMODYNAMICS OF THERMOELASTIC MARTENSITIC TRANSFORMATIONS. R. J. Salzbrenner and Morris Cohen in *Acta Metallurgica,* Vol. 27, No. 5, pages 739–748; May, 1979.

SHAPE-MEMORY ALLOYS. L. McDonald Schetky in *Scientific American,* Vol. 241, No. 5, pages 74–82; November, 1979.

TRUE TESTS OF 21ST CENTURY TECHNOLOGY: TODAY'S BUMPS, BANGS AND BURGLARS.

At General Motors we're putting technology to the test, pushing it as far into the future as it can go, designing transportation to take you into the 21st century. But first we have to help you over the obstacles you run into now. Like bumpy roads and potholes, shopping carts and car thieves. Just a few of the simple realities of everyday driving. So we've come up with ways to use tomorrow's technology to meet some of your needs today. We're using advanced computer systems, revolutionary materials, sophisticated testing and state-of-the-art electronics to make your new GM cars even safer, more durable and more reliable. At GM we're on a course for the future, but we're still dedicated to getting you down today's roads.

1.

2.

THE GM ODYSSEY: SCIENCE NOT FICTION

GM

3.

4.

5.

1. ELECTROMAGNETIC COMPATIBILITY TESTING. By simulating electromagnetic interference from broadcast signals, power transmission lines and lightning, these tests improve the reliability of vehicle electronic systems.
2. VEHICLE ANTI-THEFT SYSTEM. A coded resistor in the Corvette ignition key is the heart of a system that slows down a would-be car thief 20 to 30 minutes—time a thief doesn't have.
3. LEAN MACHINE. A 125-m.p.g. concept vehicle with an aerodynamic aluminum body and an innovative three-wheel suspension that allows the driver to lean it into turns.
4. PLASTIC BODY PANELS. The lightweight, non-rusting skin on some cars "gives" slightly on impact, helping to protect cars from dents, dings, stones and other realities of everyday driving.
5. COMPUTER-SIMULATED CRASH TESTING. Computer modeling in the earliest stages of passenger-compartment design helps gather information on crashworthiness, so safety is built in, not just added on.
6. LITEFLEX® SPRINGS. Fiberglass springs produced by General Motors are stronger, lighter, rustproof and up to five times more durable than steel.

6.