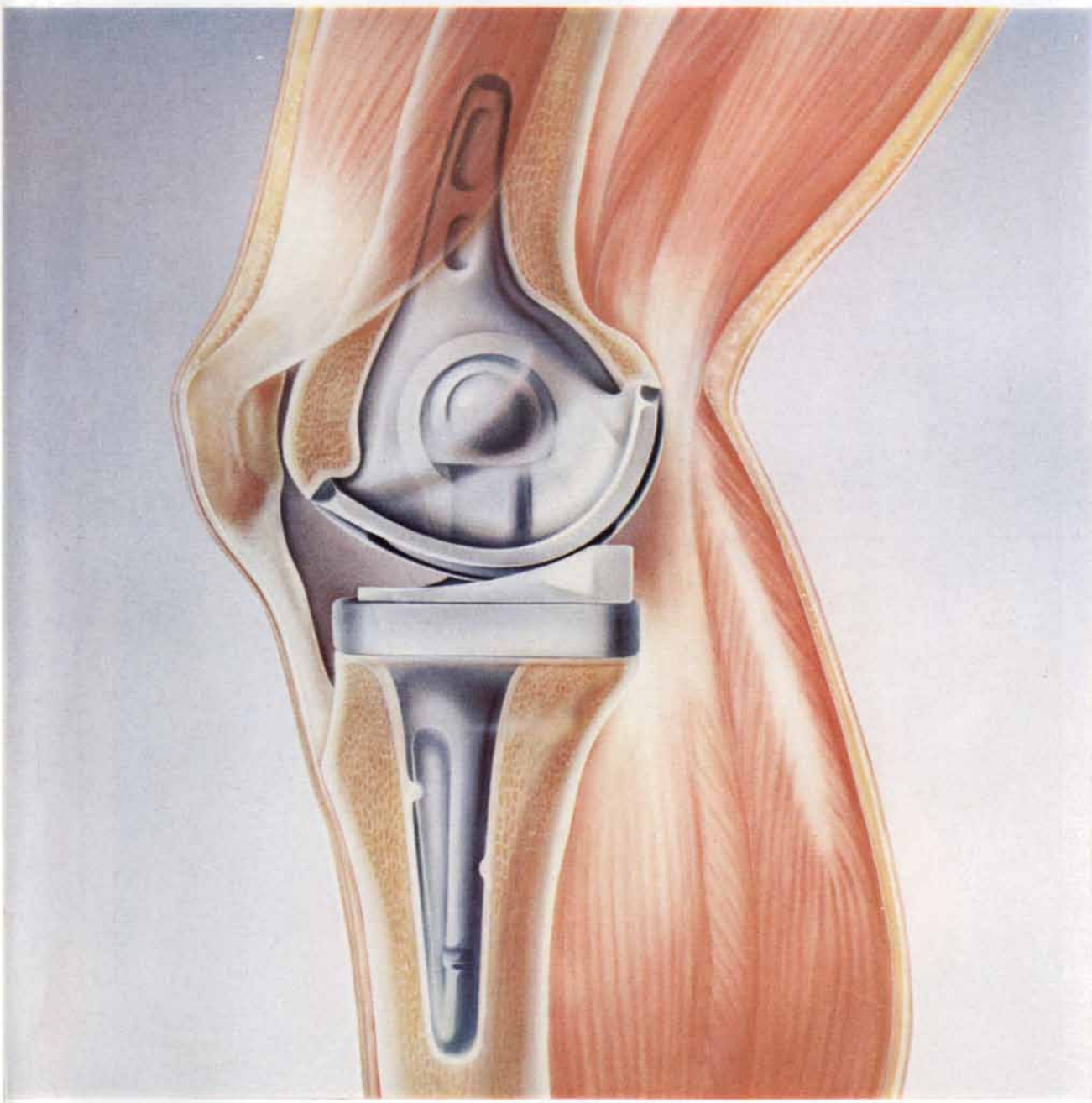
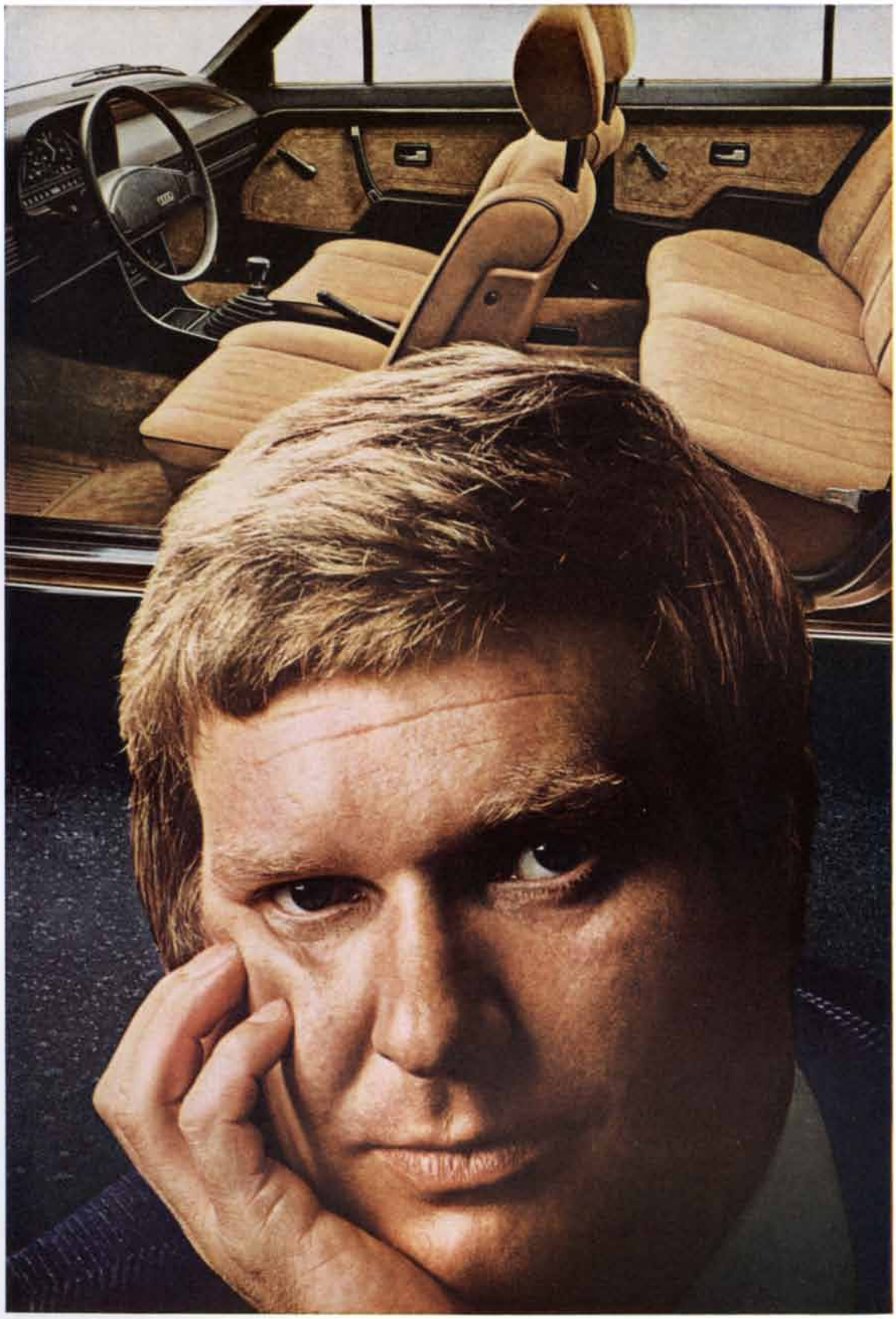
# SCIENTIFIC AMERICAN



ARTIFICIAL KNEES

*$1.50*

*January 1978*



# "I AM OPPOSED TO HAVING THE INTERIOR OF THE AUDI 5000 LOOK LIKE AN AIRPLANE COCKPIT."

AN INTERVIEW WITH HARTMUT WARKUSZ, HEAD OF STYLING

**Don't you like dials and stark interiors?**

Warkusz: It's not what I like or don't like—everything we do is tested and proved. Our psychological testing showed that dark colors and cockpit type interiors make drivers more aggressive, less relaxed. One of our principal objectives was to make the driver calmer. It is no good for you to go from a nice, calm, cheerful house right into an aggressive car.

**Should a car's interior be like a house?**

Warkusz: In some ways, yes. The old kind of interiors with stark black leather and polished chrome everywhere is no longer appropriate for today's cars. We wanted to create a more modern living environment inside the car. A car should be a pleasant, bright and comfortable place to be in. You should feel at home. Your living room at home isn't black all over, is it?

**No. But my living room doesn't have wheels.**

Warkusz: And we didn't just put wheels on a living room and so we didn't design the car only for the comfort of the driver. The Audi 5000 holds all 5 people comfortably. The trunk is huge enough to hold all their luggage.

The ventilating system is unique in that there are outlets to provide plenty of fresh air but not on anybody's neck. You can even have warm air on your feet and cool air on your face to keep you alert. For the sake of comfort and safety, the rear passengers have headrests, too. We have even gone so far as to be sure that the upholstery doesn't tear the fur out of a mink coat. Some fabrics we tested did just that.

**Will this type of interior appeal to Americans?**

Warkusz: We believe so, because we studied the American market very carefully. Not to copy, but to take American tastes into consideration. The Audi 5000 is altogether European, but there is no reason why a European car can't be bright and roomy and comfortable and quiet. A high performance European car does not have to look or sound like the cockpit of a plane.

**The lines do look European.**

Warkusz: Lines are not European or American. Lines can be ugly or beautiful, but most importantly, lines are functional or not functional. The aerodynamic values of the Audi 5000 make the air resistance very low. The drag coefficient is better than most other cars in our size category. Also, the aerodynamics made us lift up the tail of the car. That's one reason why the trunk came out so big.

For the same reason, we have complete instrumentation without two dozen unnecessary switches to distract the driver with playthings. I am glad you like the lines, but they are more than lines to me.

**What would you want an American buyer to understand about it?**

Warkusz: I would want him to understand the mature, functional design of the car. The pleasant, comfortable and quiet atmosphere of the interior. The high quality of the finish. And, the fact that it's the largest German car you can buy for the money—about $8,500*. And I would want the American buyer to understand that nothing about it is by chance.

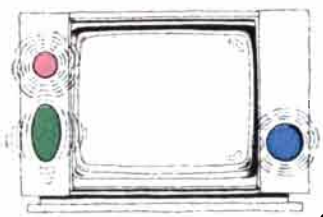
*Suggested 1978 retail price $8450, P.O.E., transp., local taxes, and dealer delivery charges, additional.

# How to buy a television set with your eyes closed.

Ever since the early days of television, manufacturers have stressed picture, picture, picture.

As a result, television sound has changed little since the introduction of the first sets over 50 years ago.

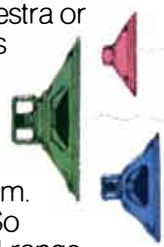
**Hear as well as you see.** But now to go along with Quasar's sharpest, clearest picture yet, we add Audio Spectrum Sound.™



A Quasar exclusive, it finally makes television sound as good as it looks.

**Not one speaker but three.** You see, whether it's a 100 piece orchestra or a simple conversation, most televisions cram everything through one speaker.

But our new Audio Spectrum Sound sets have three speakers. Each technically tailored to reproduce a precise portion of the sound spectrum.




**Higher highs, lower lows.** So there's a conventional speaker for mid-range sounds. A bass speaker specifically designed to deliver lower lows, plus another speaker for higher highs.

Together they add a whole new exciting dimension of reality to every show you watch.

Why it's *so* real, that when you watch a concert you'll think you're in the front row instead of your living room.

**Some features we'd never change.** Besides a great new sound system, Audio Spectrum Sound sets also feature traditional Quasar quality and reliability. Because the show must always go on. And on. And on.

So before buying your next TV, see your Quasar dealer and hear the difference Audio Spectrum Sound makes. And while you're there, see Quasar's Great Time Machine. It's a home video tape recorder that lets you record your favorite programs to watch at your favorite times. (That's it under the TV set below.)

But remember, if you don't hear Audio Spectrum Sound, you might spend the next few years missing out on a lot.

*Simulated Picture*

# Quasar
## makes television special again

Established 1845 **SCIENTIFIC AMERICAN** January 1978 Volume 238 Number 1

ARTICLES



DEPARTMENTS





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: $18 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $22 PER YEAR, ALL OTHER COUNTRIES.

# HP measurement and computer advances

The HP 2635 printing terminal (right) has a keyboard with functionally grouped keys for interactive applications.

## *New "smart" bidirectional printers make hard copy easy.*

With the help of HP's innovative SOS microprocessors, these new printers and printing terminals combine exceptional performance, convenience, and mechanical simplicity. They can print at 180 characters per second left to right and right to left, shift into high speed (45 inches per second) when 10 or more blank spaces in a row are detected, and skip over blanks at the beginnings and endings of lines.

Both the HP 2631 printer and the HP 2635 printing terminal are controlled by a 16-bit parallel microprocessor on a 34 $mm^2$ chip produced by silicon-on-sapphire (SOS) technology. Pioneered by HP and first announced on these pages several months ago, the SOS chip has speed, density, and power consumption characteristics that allow us to design new high-performance products at affordable prices.

The HP 2631 and HP 2635 can print in any of three character modes (normal, compressed, or expanded) under either program or operator control. Seven different line spacings can be selected (1, 2, 3, 4, 6, 8, and 12 lines per inch), or can be changed mid-text to permit the printing of legible subscripts and superscripts.

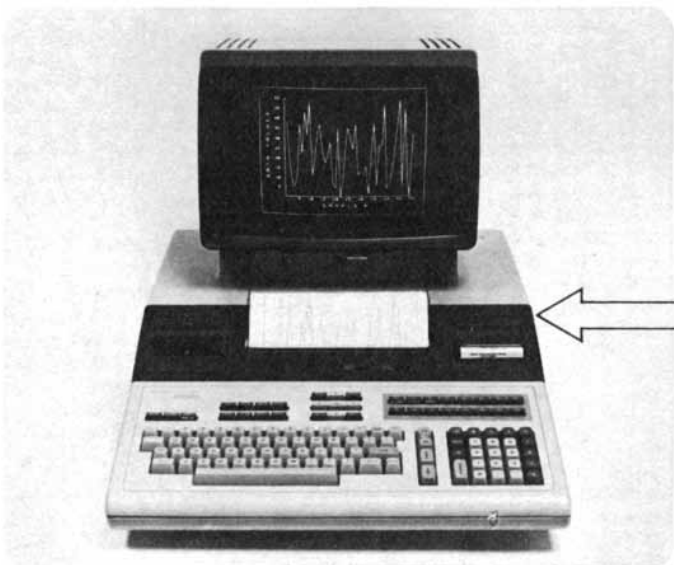
The 9-wire print head prints clear, crisp characters, including commas and semicolons, in $7 \times 9$ dot matrix on up to six-part forms. True continuous underlining is available under program control.

Our specially developed quick-change ribbon cartridge makes ribbon changes clean and easy. Its Mobius loop design results in a life span of at least 10 million characters per cartridge.

Both printers support the full US ASCII 128-character set; optional secondary character sets are available. A variety of interchangeable interface options can link the printers to a wide range of computers, terminals, and instrumentation system controllers.

Prices start at $3150* for the HP 2631 printer and $3450* for the HP 2635 printing terminal.

# *extend your possibilities.*



NMOS II chips like this one, shown approximately actual size, enhance the performance of HP's desk-top computers in a growing number of ways.

## *NMOS II: Another evolving LSI technology that supports measurable product advances.*

In parallel with HP's growing silicon-on-sapphire technology (which makes possible such products as the high-speed printers described at left) is the on-going development of NMOS—the n-channel metal-oxide semiconductor—that lies at the heart of our desk-top computers.

While our involvement with NMOS started a decade ago, when we established our first in-house NMOS facility, the present NMOS II process increases component density by a factor of two and speed by 10 over the first generation.

Using NMOS II, we developed a 16-bit ROM for the HP 9815 desk-top computer, quadrupling its firmware memory per component over that of previous desk-tops. Later, we introduced the HP 9825 desk-top computer with an NMOS II hybrid processor capable of executing 86 discrete instructions and operating at a speed in excess of 6 MHz.

Our latest product to incorporate NMOS II is the System 45 computer introduced in these pages in November. Seven NMOS II processor chips control the computer—e.g., powering the language and peripheral processes, accessing firmware memory, and controlling operation of the minicartridge tape driver.

System 45 is a clear indication that NMOS II is not limited to memory and microprocessor applications, but is branching out to encompass the whole array of desk-top computer circuits, giving our computer designers more capability and greater flexibility in less space.

By developing and manufacturing our own LSI components, our desk-top computers deliver performance advantages that enable our customers to extend their possibilities cost-effectively. (These proprietary circuits are not for sale as components.)

A comprehensive discussion of the NMOS II process appeared in the November issue of the *Hewlett-Packard Journal*. Mail the coupon for your copy.

HEWLETT hp PACKARD

1504 Page Mill Road, Palo Alto, California 94304

For assistance call: Washington (301) 948-6370, Chicago (312) 255-9800, Atlanta (404) 955-1500, Los Angeles (213) 877-1282.

Mail to: Hewlett-Packard, 1504 Page Mill Road, Palo Alto, CA 94304.
Please send me further information on

( ) HP 2631 printer & HP 2635 printing terminal
( ) HP's NMOS II technology

Name _____

Company _____

Address _____

City _____ State _____ Zip _____

*Domestic U.S. prices only.

00841

## THE COVER

The painting on the cover depicts an artificial knee joint, the Spherocentric knee, which was developed at the University of Michigan Medical Center as a prosthesis for people suffering from arthritis that has progressed to the stage of severe disability (see "The Surgical Replacement of the Human Knee Joint," by David A. Sonstegard, Larry S. Matthews and Herbert Kaufer, page 44). The component in the femur (the thighbone) is essentially a socket, in which rides a ball rising from the component in the tibia (the larger of the two long bones in the lower part of the leg). At the bottom of the femoral component is one of a pair of tracks that ride in runners in the tibial component and help to control the knee's motion. The components are cemented into the cancellous (spongy) parts of their respective bones with polymethyl methacrylate.

## THE ILLUSTRATIONS

Cover painting by Carol Donner

| Page | Source |
|---|---|
| 14–15 | Ben Rose |
| 16 | Ilil Arbel |
| 18–25 | Ben Rose |
| 26 | Ilil Arbel |
| 35–36 | National Aeronautics and Space Administration |
| 37–43 | Jerome Kuhl |
| 45 | University of Michigan Medical Center |
| 46–47 | Carol Donner |
| 48 | James Young |
| 49–50 | Carol Donner |
| 51 | James Young |
| 52 | Allen Beechel |
| 53 | John Wiley & Sons, Inc. (*top*), Allen Beechel (*bottom*) |
| 54–55 | Allen Beechel |
| 56–57 | Sung Hou Kim, Duke University School of Medicine |
| 58–62 | Allen Beechel |
| 74–75 | Carl Heiles, University of California at Berkeley, and Edward B. Jenkins, Princeton University |
| 76–77 | Allen Beechel |
| 78–79 | Carl Heiles, University of California at Berkeley, and Edward B. Jenkins, Princeton University (*top*); Allen Beechel (*bottom*) |
| 80 | Carl Heiles, University of California at Berkeley, and Edward B. Jenkins, Princeton University |
| 81 | Allen Beechel |
| 82 | © National Geographic Society–Palomar Observatory Sky Survey |
| 84 | Allen Beechel |
| 86–87 | J. W. Costerton, University of Calgary |
| 88 | Ilil Arbel |
| 89 | J. W. Costerton, University of Calgary |
| 90 | Ilil Arbel |
| 91 | J. W. Costerton, University of Calgary |
| 92–93 | Ilil Arbel |
| 94–95 | J. W. Costerton, University of Calgary |
| 97–108 | Tom Prentiss |
| 110–112 | Francis W. Kelsey Museum of Ancient and Mediaeval Archaeology, University of Michigan |
| 113–117 | Lorelle M. Raboni |
| 123–129 | Gabor Kiss |
| 133–136 | Michael Goodman |

# LETTERS

Sirs:

The fascinating article "The Search for Life on Mars," by Norman H. Horowitz [SCIENTIFIC AMERICAN, November, 1977], includes a short discussion on the chemistry of carbon, in which it is stated: "Among the elements carbon is unique in the number, variety and complexity of the compounds it can form.... No other element, including that favorite of science-fiction writers, silicon, has the capacity carbon has to form large and complex structures that are so stable."

It may surprise many readers and even some chemists to know that carbon is no longer a unique element with respect to its ability to catenate, that is, to form arbitrarily large molecular frameworks of high stability by bonding to itself. Carbon's next-door neighbor, boron, is now known to rival carbon, and in some respects even exceed it, in its capacity to form stable covalently bonded molecular networks. There are the families of borane anions such as $B_{12}H_{12}^{2-}$ (an icosahedron), carboranes, metallocarboranes (for example ($C_5H_5$) $CoC_2B_5H_7$), metalloboranes, phosphacarboranes and numerous other families involving thousands of compounds. In terms of stability p-carborane (1,12-$C_2B_{10}H_{12}$) withstands temperatures in excess of 700 degrees Celsius and is impervious to boiling sulfuric acid! High polymers based on these compounds now have commercial applications.

The geometries of the boron cage systems are very different from those common to hydrocarbon chemistry, and I do not challenge Professor Horowitz' main point that life forms not based on carbon are unlikely; in this respect carbon must remain unique until it is proved otherwise. From a strictly chemical point of view, however, there are now two elements—boron and carbon—that are capable of forming extended and varied molecular frameworks. In fact, boron in its cage compounds exhibits an even greater degree of versatility than carbon; for example, boron can bond to as many as eight other atoms simultaneously with all kinds of geometries represented, whereas carbon is much more restricted in its structural roles. Moreover, the number of different elements that have been incorporated into stable boron cage frameworks now exceeds 40—nearly half the periodic table.

RUSSELL N. GRIMES

Professor of Chemistry
Department of Chemistry
University of Virginia
Charlottesville

Sirs:

"The Solution of the Four-Color-Map Problem," by Kenneth Appel and Wolfgang Haken [SCIENTIFIC AMERICAN, October, 1977], explains a "proper map" with a counterexample, the caption under the illustration on page 113 stating: "Note that the 48 contiguous states do not even make a proper map because Michigan is made up of two pieces that are not connected."

As a Michigan fan I rise to its defense. It is not falling apart. As "Goddess of the Inland Seas," Michigan is all in one piece and has been for 140 years. Its legal boundaries include portions of Lake Erie, Lake St. Clair, Lake Huron, Lake Superior and Lake Michigan. It is grossly mistaken to define Michigan as including only the land areas of its two peninsulas.

A better counterexample would have been California. Its legal description begins with the metes and bounds of a continental portion and proceeds to the following sentence: "Also all the islands, harbors and bays along and adjacent to the Pacific coast." These islands include Santa Catalina, Santa Cruz, Santa Rosa, San Miguel (20 to 25 miles off the coast), San Clemente (50 miles) and San Nicolas (60 miles). Since the intervening waters are under Federal, not state, jurisdiction, California has at least seven disconnected parts.

R. ROBINSON ROWE

Naubinway, Mich.

Sirs:

There is no question that Susan E. Wegner ["Letters," SCIENTIFIC AMERICAN, September, 1977] is right and I was wrong in exchanging Venus for Juno in my article "Poetic Responses to the Copernican Revolution" [SCIENTIFIC AMERICAN, June, 1977]. The story of Tintoretto's painting *The Origin of the Milky Way* is about Juno and Jupiter: Jupiter, wishing to give Hercules, his bastard son, immortality, put him to sleeping Juno's breast. She wakened, pushed the infant aside; he spewed out the milk; hence the Milky Way.

But there are other puzzles about the painting. Does it show a "conflation of goddesses"—that Venus is present as well as Juno? Edgar Wind, the art historian, used to emphasize the Renaissance practice of combining myths and mythological figures that may be at work in *The Origin of the Milky Way*. This might explain why Venus' net, held up by a cherub just below Jupiter's eagle, is in the picture; why a cherub is carrying a torch; why the other cherubs are shooting at Juno and Jupiter. (A torch is associated with Venus' son, the cherub Eros; a net is one of Venus' attributes.)

It is also noticeable that Tintoretto does not distinguish one goddess from another in his mythological paintings. Ariadne, Leda, Danaë resemble one another; in *Vulcan Surprises Venus and Mars* Venus looks very much like the Juno of *The Origin*.

Another reader does not want to give in about Juno. He supports Venus as the female figure of the painting. Juno was a virago, he says, a stern and jealous matron, amply robed like Rembrandt's famous *Juno*. The evidence is, however, that there was no set way of representing Juno. She is a seductive nude in an engraving by Hendrik Goltzius, and as the third figure with Minerva and Venus in Rubens' *The Judgment of Paris*.

A Tintoretto scholar maintains that Tintoretto's iconology is consistent and entirely about Juno and Jupiter. Venus' net, she maintains, is a veil, one of Juno's attributes, and the torch symbolizes marriage and need not bring Venus into the picture.

In *The Origin of the Milky Way* Juno is Juno, certainly, but perhaps it can be allowed that Venus may be there too.

MARGARET M. BYARD

New York

**Scientific American,** January, 1978; Vol. 238, No. 1. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

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

**Offprint correspondence** and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 35 cents.

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

Name

New Address

Old Address

# Choose Any Set (3 or 4 Records) $7.98


**63. MOZART Six String Quartets to Haydn** (3 LPs)
*Quartetto Italiano*


**64. BACH Six Sonatas and Partitas For Solo Violin** (3 LPs)
*Nathan Milstein*


**62. BRAHMS Four Symphonies** (4 LPs)
*Claudio Abbado: Vienna Philharmonic, Berlin Philharmonic, Dresden State and London Symphony Orchestras*

**65. LISZT Complete Hungarian Rhapsodies, Opera Paraphrases** (4 LPs)
*Michele Campanella at the piano*

## No obligation to buy any future offerings.

This remarkable $7.98 offer is brought to you by The International Preview Society—a highly selective record program that does not oblige you to buy at any time. These multi-record sets—and many, many more superb albums—will continue to be offered in an exclusive Preview magazine every eight weeks (only seven times a year). Each issue highlights a Featured Selection . . . plus an impressive variety of alternate selections. All at dramatically low prices. For example, you pay only $16.99 for a 3-record set in our regular offerings. Only $21.49 for a 4- or 5-record set. Both well below the suggested retail prices!

*The Society's Bonus Plan will save you even more.* For every album you keep, you can purchase an outstanding Bonus Selection album *at a savings of over 50%!*

If you'd like to preview the Featured Selection, you need do nothing. We'll send it automatically. But if you'd prefer an alternate selection, or none at all, just mail back the Preview Notification card—enclosed with the magazine—by the specified date. You'll have plenty of advance notice.

**Audition Any Album At Home For Ten Days . . . Free.**

At no cost or obligation, you are invited to audition any one of these multi-record albums . . . as well as any future offerings . . . at home for ten days. Only then do you decide whether or not to buy.

**As Many As Four Records For Less Than The Price Of One.**

You may choose any one of these multi-record sets for just $7.98 plus a small postage/handling charge. (Sales tax added for New York residents.) Remember, each set contains either three or four records. That's a savings of as much as $27.94...up to 78% off the suggested list price!

Send no money. We want you to judge for yourself before you decide to buy. If you are not delighted after ten days, return the set.

These are superior, silent-surfaced recordings by Deutsche Grammophon and Philips. They come in a handsome hinged storage case with an informative, illustrated brochure.

Every recording offered by the Society must meet the rigid standards of the distinguished Carnegie Hall Selection Committee—headed by Julius Bloom, Executive Director of Carnegie Hall. Only albums of the most exceptional musical interest, artistic excellence and technical quality are offered.

So do take advantage of this marvelous value and mail in the coupon today. You'll also be reserving the privilege of previewing other great international releases . . . which you may add to your collection at impressive savings.


**60. TCHAIKOVSKY Symphonies 4, 5 and 6** (4 LPs)
*Yevgeny Mravinsky: Leningrad Philharmonic*
**Sleeping Beauty & Swan Lake**
*Herbert von Karajan: Berlin Philharmonic*


**66. MOZART Don Giovanni** (4 LPs)
*Arroyo, Freni, Te Kanawa, Ganzarolli, Wixell. Colin Davis: Chorus and Orchestra of the Royal Opera House, Covent Garden*


**67. BEETHOVEN The Middle Quartets, Op. 59 Nos. 1, 2 and 3, Op. 74 No. 10 and Op. 95 No. 11** (3 LPs)
*Quartetto Italiano*

## Free Home Audition Request

**The International Preview Society**
175 Community Drive, Great Neck, N.Y. 11025

YES! Please send me, for my free audition, the multi-record set I have indicated by the number below. I may return it after ten days and owe nothing . . . or keep it and pay only $7.98 plus a small postage/handling charge (sales tax added for New York residents). This is up to $27.94 off the suggested retail price. I will also receive, every eight weeks, free preview privileges of the finest classical recordings. I may cancel this arrangement at any time.

**Print album number you have selected here:** ☐


**FREE BONUS ALBUM**
Return coupon within 10 days for valuable *free* gift—TCHAIKOVSKY'S *1812 Overture*. Bernard Haitink: Concertgebouw Orchestra, Amsterdam.

Name

Address Apt. No.

City State Zip

Offer limited to the connecting 48 states. Offer expires June 30, 1978. Limit: one membership per household. Only new members eligible. NOTE: All applications are subject to review, and we reserve the right to reject any application.

79119

# 50 AND 100 YEARS AGO

## SCIENTIFIC AMERICAN

JANUARY, 1928: "'Long-distance radio communication in the future will be handled by short waves,' predicts Mr. Marconi. 'Until a few years ago short waves were neglected. Only recently have we discovered that these waves possess most valuable and unsuspected qualities in regard to world-wide transmission and that they are capable of results unobtainable with long waves. Short-wave beam transmission has taught us many lessons. We have noticed in working long distances by means of short waves that, particularly when receiving reflectors are employed, static has been generally conspicuous by its absence, and when it is noticeable, the signal strength has mostly been well above the level of the disturbance. Fading remains one of the most serious obstacles, although it is worse on channels between 200 and 1,000 meters than on short waves. Short waves are beginning to show unhoped-for results in improving broadcasting by making it workable over great distances, even during the daylight hours. And directive methods will soon be utilized for broadcasting by enabling programs and speeches to be sent over large portions of America and to foreign countries with much more strength and freedom from interference than is possible now.'"

"In spite of the many warnings by oil men who speak with the voice of knowledge and authority, the wild and greedy scramble to suck our rich oil deposits up to the surface and turn them into cash goes on apace. Our surpassing wealth today is to be credited, more than anything else, to the lavish way in which nature has endowed the United States with valuable resources. Both the Federal Government and the wisest leaders in the oil industry are striving to find some way to check the prevailing extravagance. Secretary Work of the Department of the Interior has suggested the creation of a committee of nine to study legislative action, and both the American Bar Association and the oil industry have appointed their respective members. The Oil Conservation Board of the Federal Government now will name three representatives. Note, if you please, that this call by the Government is for the purpose of studying 'legislative action.' We have long been of the opinion that only by legislative action can the present orgy of greed and wastefulness be stopped. Appeals to the good business sense and to the patriotic feelings of the oil getters alike seem to have fallen on deaf ears. There is no indication of a willingness among the personnel of the oil industry at large to get together for the purpose of controlled drilling. It is high time for the law to step in."

"Coal, a basic material for chemical industry, will play a very important role in obtaining liquid motor fuels in those countries that have no deposits of oil, according to Dr. Hans Tropsch of the Kaiser Wilhelm Institute for Research in Coal at Mülheim-Ruhr. 'In Germany today the problem of transforming coal into oil is being intensively studied,' Dr. Tropsch said. 'The goal is to be reached by two different roads: through the direct hydrogenation of coal, as is done by Friedrich Bergius and lately by I. G. Farbenindustrie, or catalytically from water gas, a process that Franz Fischer and I have worked out successfully at the Kaiser Wilhelm Institute for Research in Coal. The decomposition of coal solely by heat at low or high temperature to make coke or smokeless domestic fuel yields a series of chemically valuable liquid substances, but their quantities are generally small in comparison with the amount of coal used. In their production we are limited first of all by the possibility of marketing the principal product, in this case coke. For economic reasons, therefore, the quantities of the by-products that are valuable in and for themselves cannot be increased at will. It is important for the economic success of distillation that the semi-coke, the principal product of the process, be obtained in dense lumps suitable for transporting and storing. A distillation of soft coal for the sake of the oil alone has little prospect of being economically successful.'"

## SCIENTIFIC AMERICAN

JANUARY, 1878: "Trials of the Bell telephone were recently conducted before the Emperor of Germany at the palace in Berlin. His Majesty manifested the liveliest interest in the invention, and deigned to inquire its name, whereupon a high Post Office functionary coined the title 'Fernsprecher,' which means 'far talker,' and which the Emperor at once approved, so that it is now a part of the German language."

"A new oil pipe, known as the Sea Board Pipe Line, is soon to be laid from Butler County, Pa., to Baltimore, a distance of 230 miles. The transporting capacity will be 6,000 barrels of oil per day. It is expected to bring into Baltimore annually about two million barrels of crude oil, about equal to the quantity now carried there by two railroads."

"One of the most remarkable chapters in the history of civilization is the persistent efforts of the Chinese government to prevent the importation and use of opium among its people. More than a century ago the Chinese government recognized the dangers inherent in the use of this drug and began to legislate against it. The opium was grown in India and smuggled into China by English ships, encouraged by the English government. The trade finally culminated in a war (1839), which Mr. Gladstone denounced as one calculated to cover his country with permanent disgrace. In 1857 another war was provoked by England, for the mercenary purpose of increasing her Indian treasury. The sagacity of the Chinese officials in recognizing the fatal effects of the introduction and use of opium is a lesson to our boasted civilization that we cannot afford to ignore. In this country we are just beginning to realize the necessity of some legislation to control the abuse of opium and its compounds, and no one dare deny that the same peril the Chinese discerned a century ago threatens us today. The opening of Parish Hall, Brooklyn, N.Y., as an asylum for opium cases alone, and the large number of patients already gathered, are significant hints of the presence of a disorder that can destroy the vitality of both community and nation."

"Ewald and Kühne describe their experiments on the regeneration of vision purple outside the body. The retina of a frog, after it is removed from the eyeball, deprived of every vestige of black pigment and bleached in direct sunlight, still has the power to regain its original color, although not in the original intensity. In a few hours, if it is kept in the dark, it will turn first yellow, then buff-colored and lastly rose-red. This succession of changes can be repeated several times in one and the same specimen. Solutions of retina purple in purified bile, free from ether, are also capable of regaining their color after they have been robbed of it by exposure to light. Similar solutions of the retinal epithelium (apart from the rods), freed by mechanical means from suspended particles of black pigment, likewise exhibit the above property. In the dark they are rose-colored; they are bleached by light; they regain their rosy hue once more when the light is shut off from them. This power of regeneration of the vision purple is most strikingly exhibited by mixed solutions of the rods and retinal epithelium."

# WHAT THE EXPERTS CALLED THE BEST LAST YEAR WASN'T GOOD ENOUGH FOR US.

"IT CANNOT BE FAULTED."
SA 9500-STEREO REVIEW
"AS NEAR TO PERFECT AS WE'VE ENCOUNTERED." TX 9500-POPULAR ELECTRONICS
"CERTAINLY ONE OF THE BEST... AT ANY PRICE." TX 9500-MODERN HI FI


SA9500II


TX9500II

Last year, the experts paid Pioneer's integrated amps and tuners some of the highest compliments ever.

But rather than sitting back and resting on our laurels, we set out to build even better amps and tuners. Amps and tuners that would not only surpass anything we'd ever built before, but anything anyone ever built before.

Our new TX 9500II tuner is a perfect example.

Its front end features three newly developed field effect transistors that work to let you pull in beautiful FM reception no matter how far you live from the transmitter.

It has separate bandwidths for weak and strong stations. Which means you can filter all the noise out of weak stations without worrying about overfiltering strong ones.

And where most tuners have multiplex circuits that accidentally cut out some of the frequencies that add depth and presence to the music, the multiplex circuit in the TX 9500II preserves them. So the music begins to sound as if it's coming live from your living room, instead of from some radio station miles away.

Naturally, after developing a tuner this advanced, we had no choice but to create an amplifier that could match it.

That amplifier is our new SA 9500II. A fully integrated amplifier that's capable of producing a continuous power output of 80 watts per channel RMS into eight ohms, from 20 to 20,000 hertz, with no more than 0.1% total harmonic distortion.

In other words, an amp that cannot only reproduce the entire audible frequency range of every conceivable instrument, but an amp that can reproduce *all* music with a total harmonic distortion level well under what the human ear can hear.

Obviously, both the SA 9500II and the TX 9500II are very sophisticated pieces of equipment. But all of the engineering skill that went into making them has gone into every tuner and amplifier in our new series II. No matter what the price, no matter what the specifications.

And that's something you don't have to be an expert to appreciate.

High Fidelity Components

PIONEER®

WE BRING IT BACK ALIVE.

# THE AUTHORS

GEORGE M. WOODWELL ("The Carbon Dioxide Question") is director of the Ecosystems Center at the Marine Biological Laboratory in Woods Hole, Mass. He attended Dartmouth College and obtained his Ph.D. in botany from Duke University in 1958. For the next three years he taught at the University of Maine at Orono, leaving in 1961 to join the scientific staff of the Brookhaven National Laboratory. In January, 1975, he took up his present position. Woodwell's research has focused on the structure and function of natural communities, particularly forests and estuaries, and the environmental effects of ionizing radiation and pesticides.

DAVID A. SONSTEGARD, LARRY S. MATTHEWS and HERBERT KAUFER ("The Surgical Replacement of the Human Knee Joint") are orthopedic surgeons and bioengineers at the University of Michigan. Sonstegard is associate professor of applied mechanics and biomechanics in surgery. He did his undergraduate work at the South Dakota School of Mines and Technology and obtained his Ph.D. in mechanical engineering from Michigan in 1965. After working for seven years as a research engineer he returned to Michigan in 1970 on a fellowship to study biomechanics. Matthews is associate professor of orthopedics. He attended Harvard College and the Harvard Medical School, where he received his M.D. in 1963, and came to Michigan for his internship and orthopedic residency. After serving two years in the Army as an orthopedic surgeon he did research in biomechanics for a year in the laboratory of Carl Hirsch at the Karolinska Institute in Stockholm, returning to Michigan in 1971. Kaufer is professor of orthopedic surgery and chief of the orthopedic outpatient clinic. He was educated at Michigan, where he received his M.D. in 1959. After doing his internship and orthopedic residency there he served in the Air Force as chief of orthopedic surgery at Scott Air Force Base in Illinois. He joined the Michigan faculty in 1966.

ALEXANDER RICH and SUNG HOU KIM ("The Three-dimensional Structure of Transfer RNA") are physical biochemists with a special interest in the mechanism of protein synthesis. Rich is William Thompson Sedgwick Professor of Biophysics at the Massachusetts Institute of Technology. He attended Harvard College and the Harvard Medical School, where he obtained his M.D. in 1949. He then spent five years as a postdoctoral fellow doing research in physical chemistry with Linus Pauling at the California Institute of Technology. From 1954 to 1958 he was chief of the physical chemistry section at the National Institutes of Health. He has been on the M.I.T. faculty since 1958. Kim is associate professor of biochemistry at the Duke University School of Medicine. A native of Korea, he did his undergraduate work at Seoul National University and obtained his Ph.D. in crystallography from the University of Pittsburgh in 1966. Over the next six years he did postdoctoral work in Rich's laboratory at M.I.T.; he moved to Duke in 1972.

CARL HEILES ("The Structure of the Interstellar Medium") is professor of astronomy at the University of California at Berkeley. He did his undergraduate work at Cornell University and obtained his Ph.D. in astronomy from Princeton University in 1966. That year he joined the Berkeley faculty, where he has since remained except for the year 1969–70, which he spent studying pulsars at the 1,000-foot radio telescope at Arecibo in Puerto Rico.

J. W. COSTERTON, G. G. GEESEY and K.-J. CHENG ("How Bacteria Stick") are microbiologists studying the bacterial cell wall. Costerton is professor of biology at the University of Calgary in Canada. He did his undergraduate work at the University of British Columbia and received his Ph.D. in bacteriology from the University of Western Ontario in 1960. The following year he helped to organize the department of premedical sciences at Baring Union Christian College in India. In 1966, after a two-year research fellowship at the school of botany of the University of Cambridge, he joined the faculty of McGill University in Montreal; he moved to the University of Calgary in 1970. Geesey is a sessional lecturer in the department of biochemistry and microbiology at the University of Victoria in British Columbia. He attended the University of California at San Diego and received his Ph.D. in marine microbiology from Oregon State University in 1975. He then went to Costerton's laboratory as a postdoctoral research associate. Cheng is a ruminant microbiologist with the Canadian Department of Agriculture. Born in Taiwan, he was educated at the National Taiwan University and the University of Saskatchewan, where he obtained his Ph.D. in microbiology in 1969. He went to Costerton's laboratory at McGill in 1968 as a Canadian National Research Council fellow in microbiology and took up his present job in 1971.

HARRY R. LEWIS and CHRISTOS H. PAPADIMITRIOU ("The Efficiency of Algorithms") are assistant professors of computer science at Harvard University's Center for Research in Computing Technology. Lewis attended Harvard College, where he studied applied mathematics and pursued an interest in computer graphics. After being graduated summa cum laude in 1968 he served as a commissioned officer in the Public Health Service, designing systems and application programs for a large computer at the National Institutes of Health. He returned to the university for graduate study in applied mathematics, receiving his Ph.D. in 1974. Papadimitriou, a native of Athens, did his undergraduate work in electrical engineering at the National Technical University of Greece. He obtained his Ph.D. in applied mathematics from Princeton University in 1976 and joined the Harvard faculty that same year.

JOHN H. HUMPHREY and JOHN GRIFFITHS PEDLEY ("Roman Carthage") are classical archaeologists at the University of Michigan. Humphrey is assistant professor in the department of classical studies and assistant curator of the university's Francis W. Kelsey Museum of Ancient and Mediaeval Archaeology. He studied classical archaeology at the University of Cambridge, where he received a starred first honors degree, and went on to obtain his Ph.D. in archaeology from Bryn Mawr College with a dissertation on Roman circuses. Pedley is professor of classical archaeology and Greek and also director of the Kelsey Museum. Born in England, he received his master's degree from Cambridge in 1959 and spent the next four years on fellowships at Harvard University and at the American School of Classical Studies in Athens. He returned to Harvard to complete his Ph.D. in 1965 and then joined the Michigan faculty. Pedley has participated in excavations at Sardis in Turkey, at Apollonia in Libya and most recently at Carthage.

PETER DUNN-RANKIN ("The Visual Characteristics of Words") is professor and chairman of the department of educational psychology at the University of Hawaii. He attended Florida State University, where he received his bachelor's degree in secondary education and his master's in language arts in 1954. After serving in the Navy he worked for five years in the Sarasota, Fla., public schools. He then returned to graduate school to obtain his master's degree in mathematics from Louisiana State University and his Ed.D. in education research from Florida State in 1965. Dunn-Rankin joined the University of Hawaii faculty in 1970, where he has remained except for a year's leave of absence in 1973 at the Psychophysics and Perception Laboratory of the University of Stockholm.

# CHOOSE ANY ONE OF THESE VALUABLE COMPLETE WORKS

## as a trial member of the BOOK-OF-THE-MONTH CLUB®

The Suggested Trial: You simply agree to buy four Club choices within a year at substantial savings on most books you choose

## CARL SANDBURG's *Pulitzer Prize-Winning* ABRAHAM LINCOLN

**The entire six-volume set**

More than a portrait of a great American, this monumental 6-volume biography is a tapestry of America in crisis and resolution. Winner of a Pulitzer Prize for history...a work of timeless beauty and humanity.

for only $18.50

PUBLISHER'S LIST PRICE: $120

## THE STORY OF CIVILIZATION

by WILL and ARIEL DURANT

**Eleven matched volumes**

This multi-volume history of all mankind—ancient and modern, Occidental and Oriental, is one of the great literary achievements of the twentieth century.

for only $20.00

PUBLISHER'S LIST PRICE: $248.55



## THE NEW COLUMBIA ENCYCLOPEDIA

**New fourth edition... completely revised after a lapse of twelve years**

Cited in earlier editions as the best single-volume reference work, this latest illustrated edition has over 7000 new articles, numerous useful features and lists: tables of rulers of many countries...Third World persons, events, art forms and places.

for only $10

PUBLISHER'S LIST PRICE: $79.50


## THE COMPACT EDITION OF THE OXFORD ENGLISH DICTIONARY

Contains the complete contents of the 13-volume *Oxford English Dictionary*, generally considered the final arbiter of the meaning, origin, history and usage of words in the English language. Bausch & Lomb magnifying glass included.

for only $17.50

PUBLISHER'S LIST PRICE: $95



IF YOU are not aware of the many clear advantages, both economic and intellectual, of shopping in America's Bookstore,® the dramatic opportunity outlined here is well worth your attention. Until now, high retail prices have put the magnificent works shown here beyond the reach of many reading families. However, as a trial member of the Book-of-the-Month Club, you can choose any one of these works at an extraordinary saving—as much as $228.55

This bounty for booklovers is not a one-time affair but is a demonstration of a continuing benefit. Not only will you be able to keep abreast of the important new books you are eager to read, but if you continue your membership you will enjoy similar savings on expensive library sets and high-priced reference works. It's a wonderful way to build a well-rounded library and save money, too.

### FACTS ABOUT MEMBERSHIP

- You will receive the *Book-of-the-Month Club News*,® a literary magazine published by the Club fifteen times a year. The *News* describes the coming Selection and scores of Alternates, and will be sent to you approximately every three and a half weeks.
- If you wish to purchase the Selection, do nothing and it will be shipped to you automatically.
- If you do not want the Selection—or you would like one of the Alternates or no book at all—simply indicate your decision on the reply form always enclosed with the *News* and mail it so we receive it by the date specified.
- If, because of late mail delivery of the *News*, you should receive a Selection without having had 10 days to decide whether you want it, that Selection may be returned at Club expense.
- All books distributed by the Book-of-the-Month Club are identical to the publishers' editions in content, format, size and quality.

**BOOK-OF-THE-MONTH CLUB, INC.**
**Camp Hill, Pennsylvania 17012**

8-A170-1

Please enroll me as a member of the Book-of-the-Month Club and send me the work I have checked in the box at right, for which you will bill me. I agree to buy four Selections or Alternates during the first year I am a member, paying in most cases special members' prices. My membership is cancelable any time after I buy these four books. A shipping charge is added to all shipments.

- ☐ THE COMPACT EDITION OF THE OED $17.50
- ☐ CARL SANDBURG'S ABRAHAM LINCOLN $18.50
- ☐ THE NEW COLUMBIA ENCYCLOPEDIA $10
- ☐ THE STORY OF CIVILIZATION $20.00

MR. / MRS. / MISS .....63

(Please print plainly)

*Address* .....*Apt*.....

*City & State* .....*Zip*.....

# MATHEMATICAL GAMES

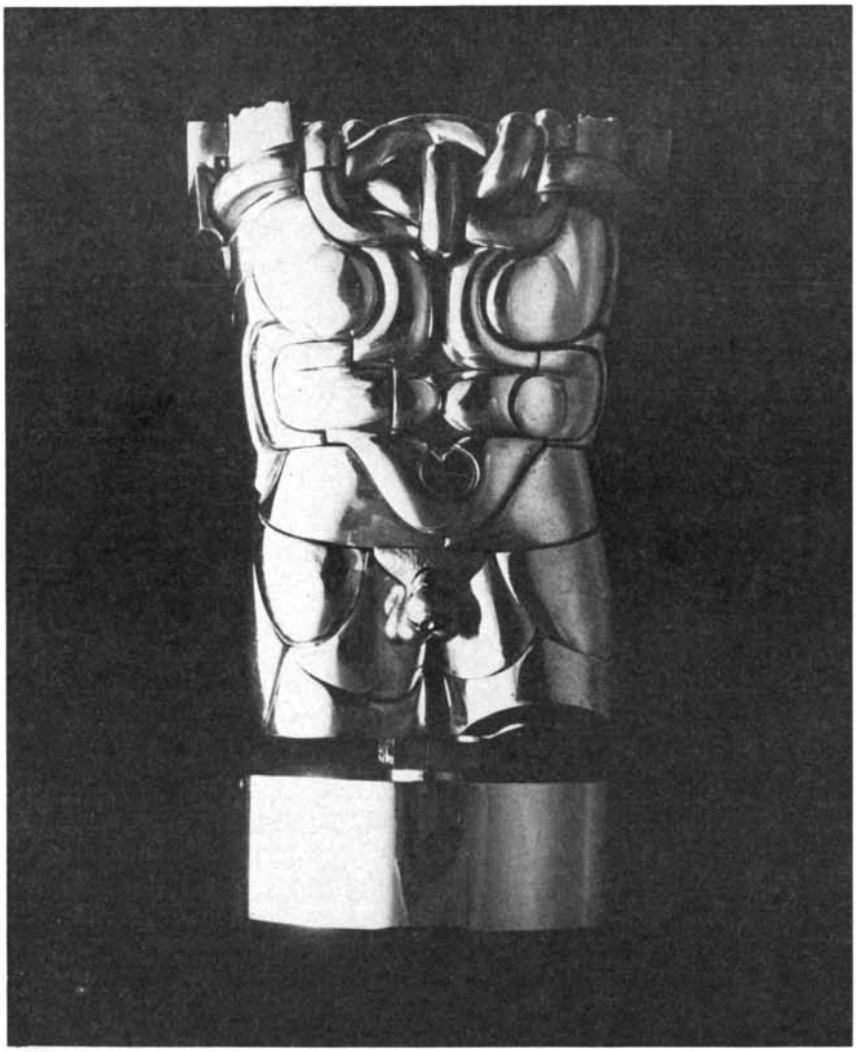
## *The sculpture of Miguel Berrocal can be taken apart like an interlocking mechanical puzzle*

by Martin Gardner

"We then fell into a disquisition whether there is any beauty independent of utility. The General maintained there was not. Dr. Johnson maintained that there was; and he instanced a coffee-cup which he held in his hand, the painting of which was of no real use, as the cup would hold the coffee equally well if plain; yet the painting was beautiful."

—JAMES BOSWELL,
*Life of Samuel Johnson*

The purpose of this month's column is to introduce readers to Miguel Berrocal, the foremost living Spanish sculptor. Berrocal is already a cult figure in Europe with a steadily growing following, but he is surprisingly little known in the U.S. Before detailing Berrocal's remarkable achievement, however, and explaining why his sculpture should be discussed in a column on recreational mathematics, we must consider one of the oldest categories of mechanical puzzles.



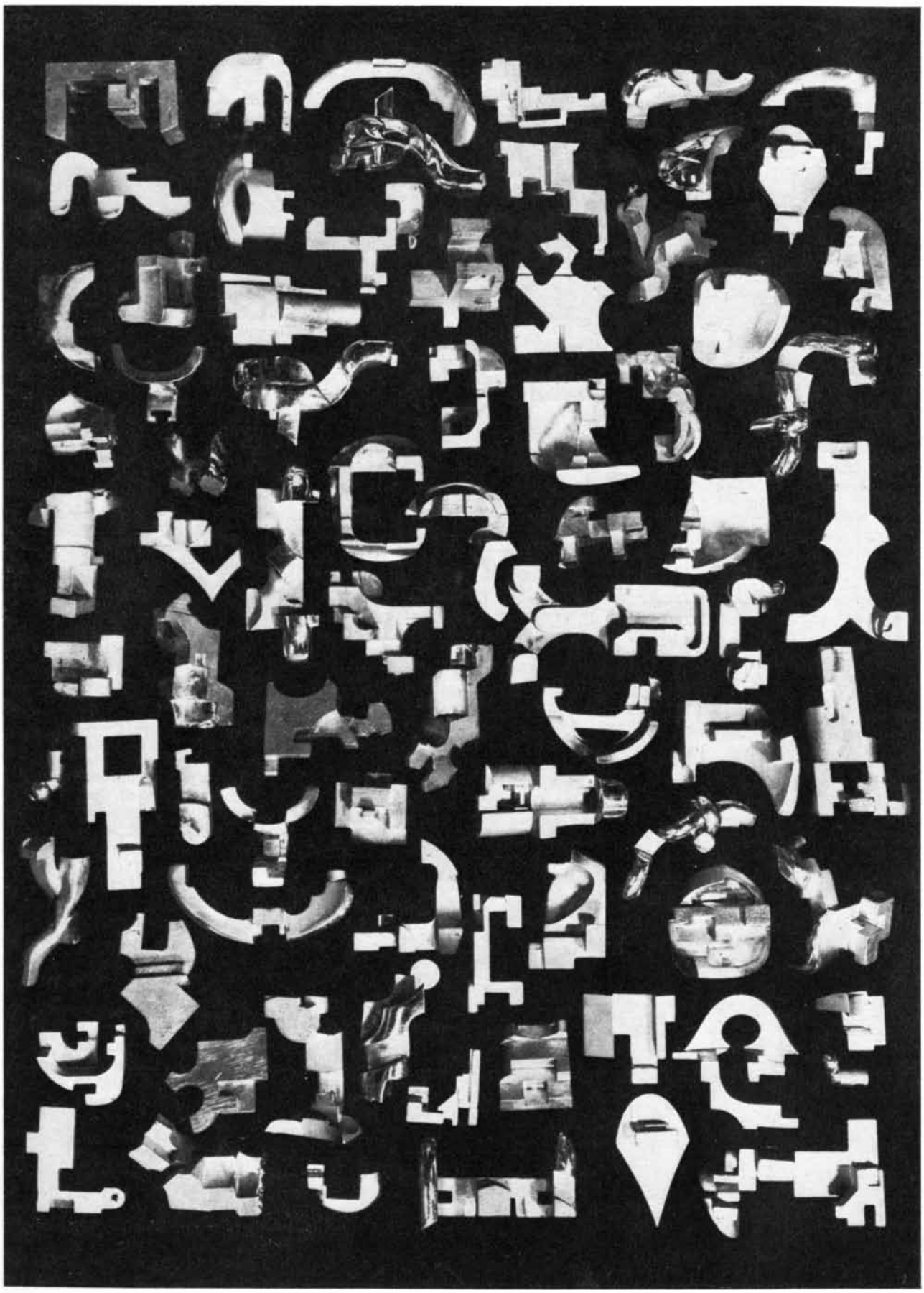
***Miguel Berrocal's* Goliath**

Is there any reader who has not at some time held one of those puzzles made of wood pieces so cleverly interlocked that they are quite difficult to separate? These objects are usually called Chinese puzzles, and once they have been taken apart putting them back together again can be even more challenging. There is usually a single piece, called the key, that must be removed first in order to disassemble the puzzle. In many such puzzles the pieces must be replaced in a certain sequence in which the key is always inserted last to lock the other pieces firmly in place. Hundreds of differently structured puzzles of this type have been sold around the world for centuries, most of them invented anonymously (even though hundreds of patents for unmarketed variations on the puzzles have been issued).

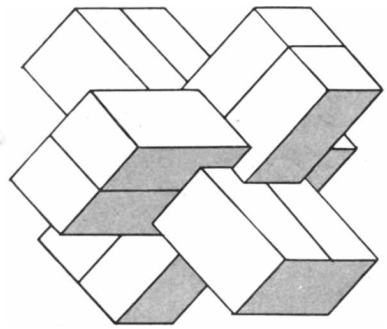
Almost nothing is known of the early history of these puzzles, but "Chinese puzzles" is probably a misnomer. East Asian countries were certainly producing puzzles of this kind as early as the 18th century, but so were European countries. Moreover, other kinds of mechanical puzzles of unquestionably Western origin were also called Chinese puzzles. As Joseph Needham writes in Volume III of *Science and Civilization in China:* "Perhaps Europeans were inclined to ascribe to puzzles the name of what was, to them, a puzzling civilization."

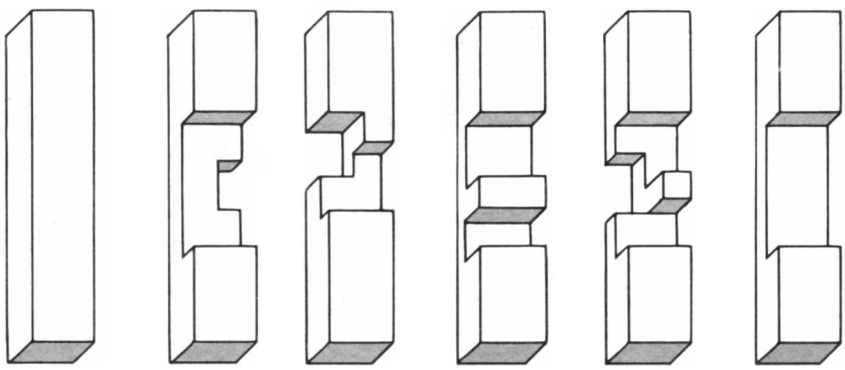
Chinese puzzles can have as few as three pieces, but the simplest nontrivial model is the popular six-piece puzzle shown in the top illustration on page 16. The six pieces are shown below the assembled puzzle as they were depicted by "Professor Hoffmann" (the pseudonym of Angelo Lewis) in his 1893 book on mechanical puzzles, *Puzzles Old and New*. The unnotched bar at the far left is the key. A different set of six pieces appeared earlier in *The Magician's Own Book,* an anonymous work first published in the U.S. in 1857. Anthony S. Filipiak, in *100 Puzzles: How to Make and Solve Them* (A. S. Barnes and Co., 1942), calls such an assembly a "six-piece burr puzzle," presumably because it looks like a seed burr. The name "burr" is now commonly applied to all puzzles of this kind.

Manufactured versions of the six-piece burr are seldom alike. This fact suggested to Filipiak a surprisingly difficult problem in geometric combinatorics. Imagine that the middle part of the puzzle's unnotched key piece is divided



*The 80 pieces of* **Goliath**

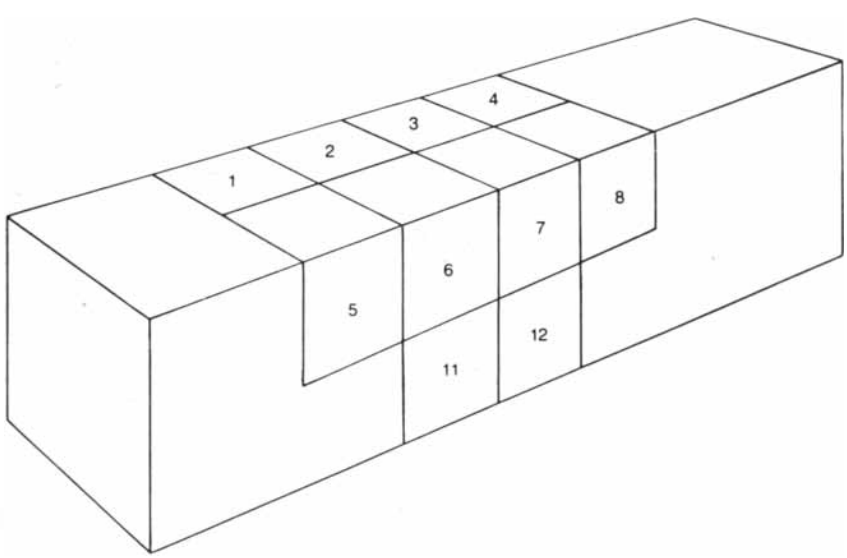




*A six-piece "burr" puzzle*

into 12 unit cubes as is shown in the illustration below. The sides of each cube are equal to half the depth of the key. There are four more cubes under the cubes numbered 2, 3, 6 and 7, but only two of them, 11 and 12, can be seen in the illustration. Each piece of any six-piece burr can now be described by stating which of the 12 unit cubes have been removed.

Since each unit cube is either present or absent in a piece of the puzzle, there are $2^{12}$, or 4,096, possible patterns in all. Those that divide the bar into separate pieces must of course be eliminated. Among the remaining pieces simple rotations and end-for-end turns account for eight repetitions of any single pattern. Therefore for each pattern seven duplicates can be eliminated. (Mirror reflections of asymmetric configurations are not excluded.) Finally, there are some patterns that for mechanical reasons will not combine with any other pattern to form an interlocking structure. These patterns can also be eliminated. After the above exclusions have been made, how many configurations remain? In other words, how many of the 4,096 patterns can be utilized in a six-piece burr puzzle? Filipiak believed there were 432.




*Pattern for the pieces of a burr puzzle*

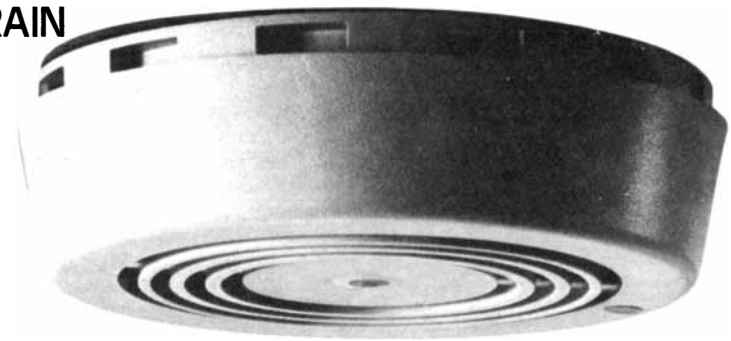
In recent years three mathematicians have employed computer programs to attack the problem. They are William H. Cutler of Wartburg College in Waverly, Iowa, Robert H. Mackay of London and C. Arthur Cross of Cheshire in England. Cutler and Cross now agree that the number of usable pieces is 369. Sixty-seven of those pieces, however, can be utilized only with duplicates and two can be utilized only with triplicates, making a set of 440 pieces in all. A second question, still unsolved, now arises: How many different six-piece burr puzzles can be made from this set? Interlockings that are impossible to put together or take apart must be excluded. It is also advisable to omit constructions with internal holes, because they usually do not hold together well. Given these constraints, Cutler's computer program found the number of distinct six-piece burrs to be 119,963, but Cutler is not satisfied that the program is free from errors. Interested readers can write to him for details.

I know of only two places where more elaborate kinds of burr puzzles can be obtained. In England, Pentangle (Over Wallop, Hants, England S020 8JA) will send a catalogue of their mechanical-puzzle line, which includes several handsome wood puzzles of the burr type. The one called "Grandpa Chuck" has 96 pieces. In the U.S. if readers send a stamped and self-addressed business envelope to Stewart T. Coffin (Old Sudbury Road, R.F.D. 1, Lincoln, Mass. 01773), he will send back a brochure for his unusual line of original burr puzzles. His prices are high because all the puzzles are handcrafted from hardwoods. For several years Coffin issued an occasional newsletter, *Puzzle Craft,* that dealt with the history and construction of mechanical puzzles. He tells me he will send a 24-page reprint of the six issues that have appeared, along with the brochure, for $2 postpaid.

From time to time Chinese puzzles have been manufactured to resemble familiar objects: a car, a pistol, a battleship, an airplane, a pagoda, a barrel, an egg and various animals. Most interlocking wood puzzles are not designed to represent anything, and yet because of their symmetries they are always pleasing to look at. It is the representational models that are usually low in aesthetic appeal. This brings me to Berrocal. As far as I know he is the first

# SMOKE DETECTOR WITH A BRAIN

# The Nose



*The JS&A fire and smoke detector acts as a space-age nose in the early detection of a fire.*

## *A new computerized fire and smoke detector continually sniffs the air, measures danger and sounds a new alarm.*

It's proven. A fire and smoke detector saves lives, and if you want to own the very best detector with computer circuitry and solidstate alarm, this advertisement is important.

The new JS&A computerized fire and smoke detector is a breakthrough. The unit's integrated circuit takes the dozens of other conventional detector components and reduces them to only a few—a potentiometer, one capacitor and one LED warning light. With fewer components there's less chance of failure. And the integrated circuit also measures the smoke and triggers the solid-state alarm.

### NEW SOLID-STATE ALARM

The solid-state alarm replaces the electromechanical alarms on other fire and smoke detectors. The big advantage here is its reliability, the very low power drain, and the loud sound. Your alarm will sound louder longer, use less battery power, be more efficient, and your single 9-volt battery will last much longer.

The JS&A alarm produces a loud 85db sound at a distance of ten feet with only 12 milliamps of current. With elecromechanical alarms, it takes 75 milliamps to produce an 85db sound or six times the power of the JS&A unit to do the same job.

The most dramatic difference is the start-up power required to trigger the alarm. In a standard smoke detector, the alarm requires approximately 300 milliamps during start-up. In the JS&A unit only 12 milliamps—or 25 times less power is required. Even if your batteries are practically expired, the JS&A unit can still operate efficiently.

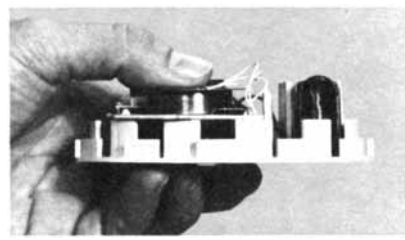
### AUTOMATIC TEST PROGRAM

The unit also has a testing system that does more than just warn you when the battery gets low. The unit will chirp every minute every day for one month to indicate that the battery needs changing. Even in an emergency the JS&A will still function perfectly during this chirping period. You can then easily remove the cover and replace the 9-volt battery. There is also a test circuit to test the alarm. Just touch the side of the unit with a broom handle and the alarm will sound—also perfect for fire drills.

### SENSITIVE SENSING SYSTEM

The nose or sensing system of the unit has two chambers. The first chamber is sealed and is used as a reference. The second is open and continually tests the air in the room.

Each chamber contains a space-age element called Americium. This element ejects small alpha particles of harmless ionized helium atoms which charge the air in both chambers. If smoke enters the open chamber, the charged particles decrease. The small computer circuit then measures the decrease of change in the charged particles and compares the change with the sealed reference chamber. Even if the charge drops one millionth of a percent, the computer will sense the change and trigger the alarm. The system's computer will still differentiate between a real fire and cigarette smoke unless the cigarette smoke is blown directly into the unit.



*Only 4½ inches in diameter, the new JS&A unit has almost all of its components condensed on one integrated circuit. Shown above is the unit from a side view showing the solid-state alarm and the battery holder.*

*The encapsulated integrated circuit is shown above. This is the key to the breakthrough that makes the new JS&A detector one of the most reliable products on the market.*

### HIGH RELIABILITY

The JS&A smoke detector is so reliable that it passes the military performance specification 880. Each unit is thoroughly tested and "burned in" before it is packaged. We designed the JS&A unit for quality. It is not the cheapest unit on the market—but then we felt that reliability and confidence in an early warning security system were more important than skimping to save a few dollars.

### DON'T BE CONFUSED

**The JS&A detector gives you a new and better choice—the totally solid-state detector. It is the latest development in the computerization of previously conventional systems—a great boost for dependability and efficiency. Don't be confused. You can't buy a better fire and smoke detector.**

### NO INSTALLATION REQUIRED

There is no installation required. Just attach the JS&A detector on your ceiling with pressure sensitive tape supplied with each unit.

The JS&A detector is small due to its few components and small integrated circuit. It measures only 4½ inches in diameter and 1 inch deep and weighs only 5½ ounces so it is less obvious than any other detector and it doesn't require the expensive alkaline batteries that other detectors require. An ordinary 9-volt battery will last over a year, whereas an alkaline battery will last almost two years.

JS&A is America's largest single source of space-age products—further assurance that your modest investment is well protected. We have been in business for over a decade providing advanced electronic products directly to consumers. Service should never be a problem. If, however, your tests indicate a defect, JS&A will replace your unit with a brand new one during our one year limited warranty.

To order your JS&A fire and smoke detector, credit card buyers may call our toll-free number below or send a check for **$29.95** plus $2.50 for postage and handling (Illinois residents add 5% sales tax) to the address below.

Let the latest space-age computer technology protect your home with confidence. Order your JS&A fire and smoke detector today.

JS&A® NATIONAL SALES GROUP

Dept. SA One JS&A Plaza
Northbrook, Ill. 60062 (312) 564-9000

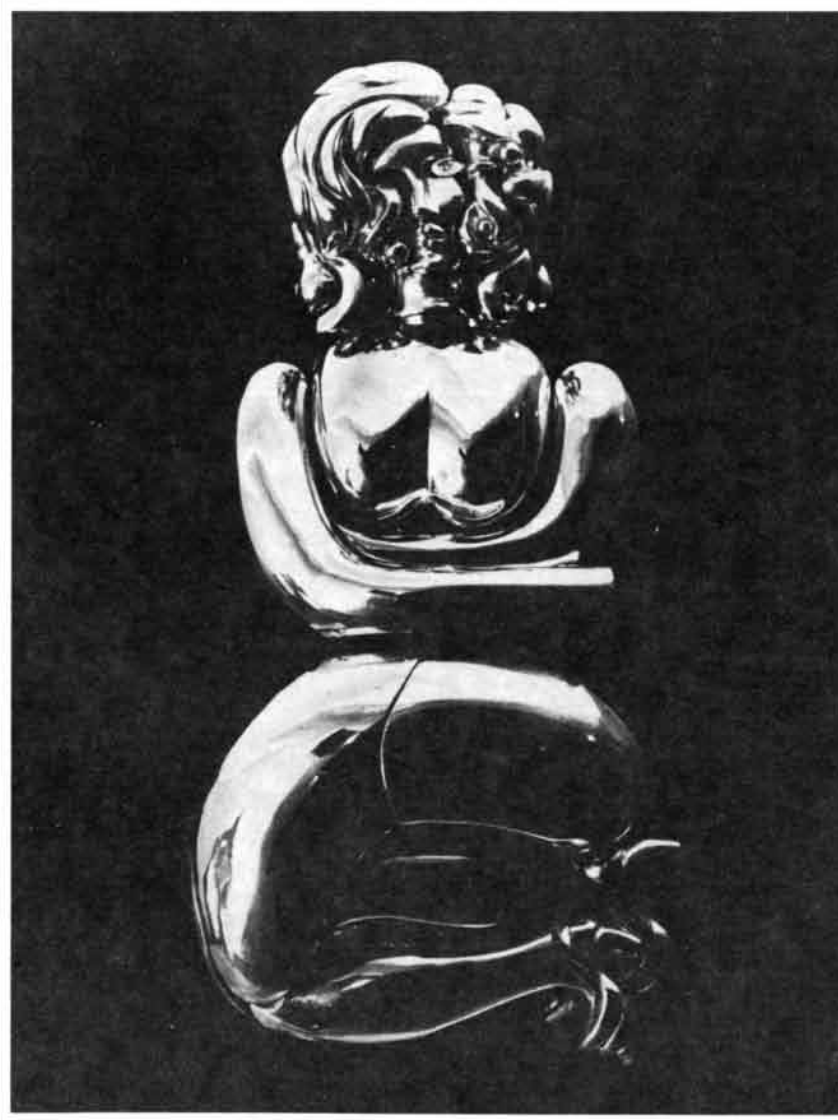
**CALL TOLL-FREE.... 800 323-6400**
**In Illinois call ..... (312) 498-6900**

person to combine the interlocking Chinese puzzle with high art.

Berrocal was born in 1933 into a middle-class Spanish family in Málaga. He studied mathematics and architecture at the University of Madrid, then art in Paris and Rome before settling in Negrar, a suburb of Verona. He lives there today in a palatial villa with his second wife, Princess Cristina. (She is the granddaughter of the last king of Portugal.) Berrocal presides over a 200-employee foundry in Negrar that casts not only his own work but also much work of other European sculptors. "I am the boss of the sculptors' Mafia," he once said. Like Pablo Picasso and Salvador Dali, the most notable Spanish painters of the 20th century, Berrocal is a virtuoso who combines a prodigious output with skillful public relations and a dazzling, immodest personality.

It is impossible to appreciate the unique combination of values in Berrocal's work—visual beauty, tactile pleasure, humor and the intellectual stimulation of a three-dimensional combinatorial puzzle—until one has taken a Berrocal apart and put it back together several times. For example, consider the two Berrocals shown stacked together in the illustration below. The head, called *Portrait de Michèle,* consists of 17 separate, curiously shaped elements, each one designed to be an individual piece of abstract sculpture and also pleasing to the fingers. The body, called *La Totoche* (The Plump Lady), is one of several bodies onto which the head can be fitted. It can be broken down into 12 pieces.

Berrocal introduced the term "multiples" for the mass-produced copies that he casts of each of his works. His edition of *Portrait de Michèle* is typical: six multiples in solid gold, 500 in sterling silver and 9,500 in nickel-plated bronze. Each copy is numbered and signed and flawlessly crafted with engineering precision. In every Berrocal the pieces must



**Portrait de Michèle (head)** ***attached to*** **La Totoche (body)**

# Where superior technology makes the musical difference: Sansui's new DC integrated amplifier and matching tuner.

Sansui is proud to introduce the new AU-717 DC integrated amplifier and matching TU-717 tuner, designed for your greatest listening pleasure. We are proud of the superlative specifications that our sophisticated research has achieved. The finest available at any price.

But the best specs alone don't always mean the finest music reproduction. And so we are proud that our precision engineering and superior circuitry design create pure and brilliantly clean tonal quality that's distinctly superior.

Listen to what we offer: Frequency response of the AU-717 from main in, 0Hz to 200kHz (+0dB, −3dB), (the widest of any DC integrated amplifier available), gives you sharp, clean transients and greatly reduced phase shift problems. Total harmonic distortion is astoundingly low, less than 0.025%, from 10-20,000Hz. 85 watts/channel min. RMS, both channels driven into 8 ohms.

Dual independent power supplies provide truest stereo separation and a large power reservoir. For uncolored phono reproduction equalization is within ±0.2dB (20-20,000Hz, extended RIAA curve). And the calibrated-attenuator level control guarantees volume precision.

The matching TU-717 tuner features dual IF bandwidth to let you select for lowest distortion (0.07% mono, 0.07% stereo) or maximum selectivity (80dB). S/N is excellent: 80dB mono, 77dB stereo.

In addition, the AU/TU 717's are elegantly styled, offer rack mounting adaptors and are most attractively priced. Less than $450* for the AU-717 and less than $320* for the TU-717.

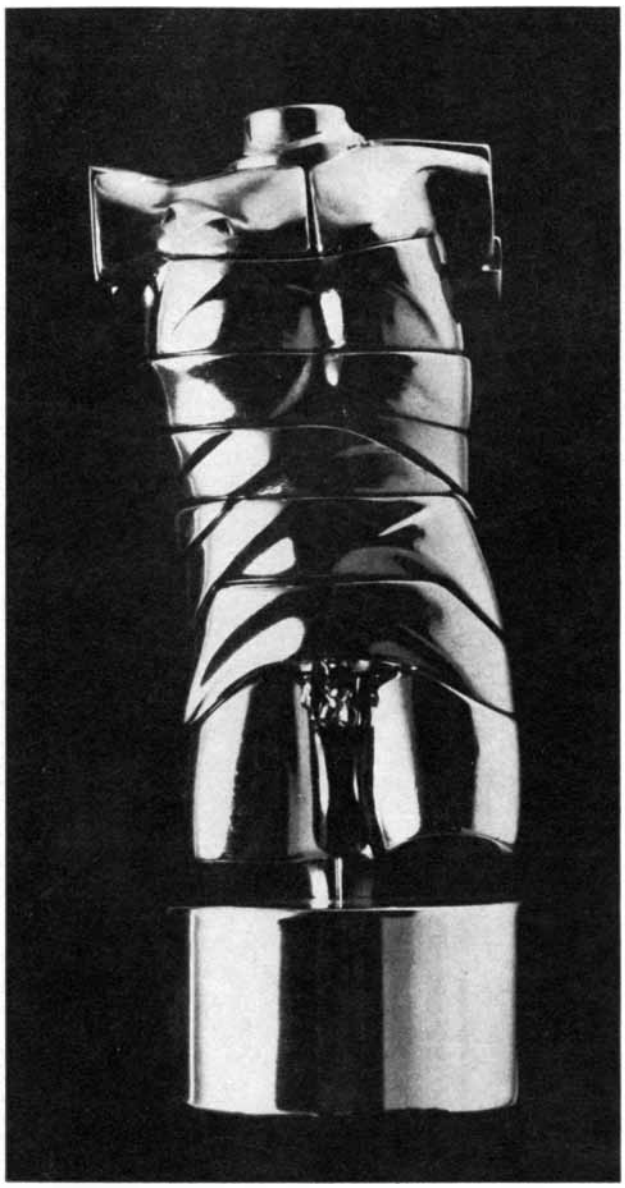
Listen to these brilliant new components at your franchised Sansui dealer today. When you hear the new Sansui AU/TU-717's, you will never again want to settle for less than the best.

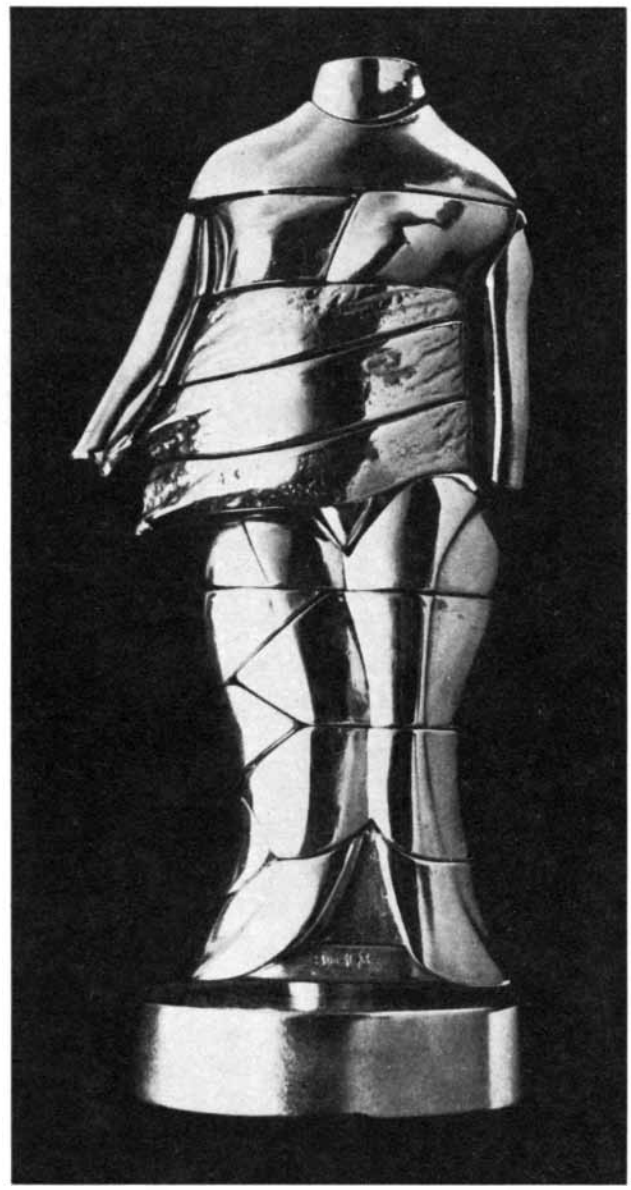
Sansui. A whole new world of musical pleasure.

*Approximate nationally advertised value. The actual retail price will be set by the individual dealer at his option.

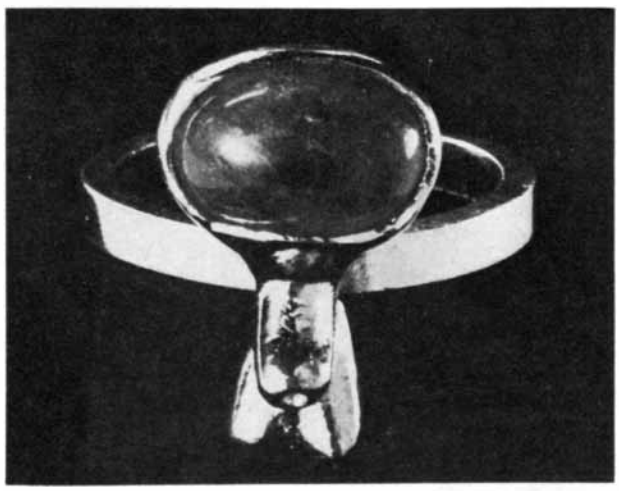

Sansui

**SANSUI ELECTRONICS CORP.**
Woodside, New York 11377 • Gardena, California 90247
SANSUI ELECTRIC CO., LTD., Tokyo, Japan
SANSUI AUDIO EUROPE S.A., Antwerp, Belgium • In Canada Electronic Distributors





**Mini David (left)** ***and*** **Mini Cariatide (right)**



***Rings from*** **Mini David (left)** ***and*** **Mini Cariatide (right)**

be separated and reassembled in a certain order. To disassemble *Portrait de Michèle* one must first remove an element in the neck. Chinese burr puzzles tend to fall apart as soon as a few pieces are removed, but a Berrocal holds firmly together until the last two pieces are separated. In many cases the assembled model is completely solid—that is, there are no interior holes—and until piece $n$ is removed it is impossible to remove piece $n + 1$. The interlockings are so ingenious that sometimes a piece cannot be taken out until the positions of other pieces have been slightly altered. Every multiple comes with a hardcover instruction book illustrated by Berrocal. Each stage of assembly is depicted on a separate page with the piece to be moved shown in color. At the end an isometric drawing displays the outline of all the pieces (in transparent form) after they are in place. Even when the instruction book is consulted, however, it may take several days to master the technique of taking apart and reassembling a Berrocal.

The number of pieces in a Berrocal varies from three to almost 100. Exquisite finger rings and bracelets, all wearable, are elements in many of the sculptures. For example, the pupil of Michèle's eye is the aquamarine stone of a ring. *Mini David,* the torso shown at the left in the top illustration on the opposite page, is another popular Berrocal. One of its 22 pieces is the ring shown at the left in the bottom illustration on the opposite page. The genitals of the torso hang below the gem of the ring. The entire edition of this work has been sold. It consists of six solid gold, 500 gold-plated bronze and 9,500 nickel-plated bronze multiples. In the gold-plated set the gem of the ring is green jade; in the nickel-plated set it is sapphire. Berrocal often produces "micro" versions of his works designed as pendants. There is a *Micro David* that includes a ring with a mesh band and a blue heart of lapis lazuli. The internal structure of a micro work is always completely different from that of its mini counterpart.

*Mini Maria,* shown in the upper illustration on the next page, can be disassembled only when a ball bearing on one leg is pressed. This sculpture is made of 23 pieces. One of them is a ring with a moonstone that forms one of Maria's breasts. Inside the figure there is a male sex organ, appearing at the left in the lower illustration on the next page, that comes apart in five pieces, two of which are steel balls. The corresponding element in the pendant *Micro Maria* has a tiny aquamarine on the tip of the organ. The gem is attached to a ring with a mesh band.

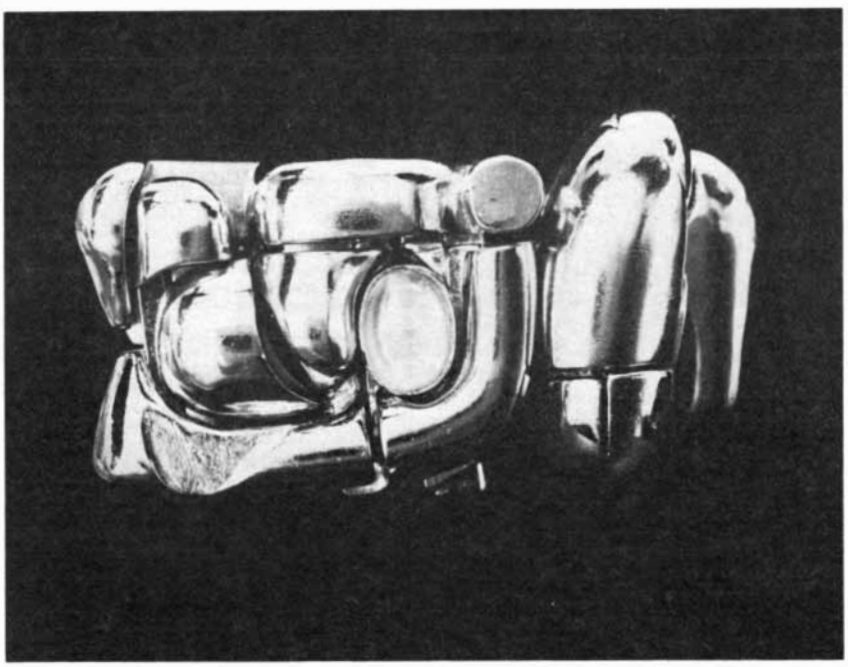
Another reclining figure, *Mini Zoriade,* opens when one shoe is rotated. Zoriade's breasts are the moonstones of

a ring. *Mini Cariatide* appears at the right in the top illustration on page 20. The mons veneris of the figure is on the gold ring appearing at the right in the bottom illustration on page 20.

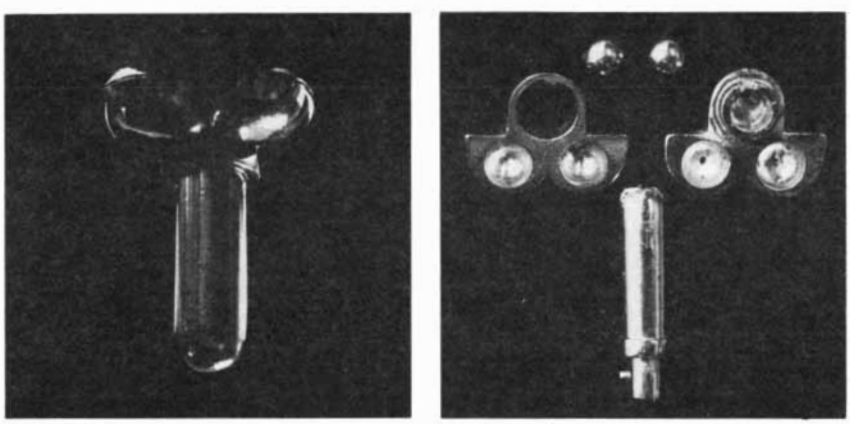
*Goliath,* shown in the illustration on page 14, is Berrocal's most complicated work. All 80 of its elements appear in the illustration on page 15. The sculpture is shown partially disassembled in the top illustration on page 24 to display how a Berrocal is a striking work of abstract sculpture at any stage of its assembly. When the torso of *Goliath* is completely assembled, its fig leaf can be rotated to expose the genitalia. Actually there are two pieces representing genitalia, one circumcised and one not, and the torso can be assembled with either one exposed.

Berrocal's coffin for Romeo and Juliet appears in the bottom illustration on page 24 together with an earlier 16-piece work that depicts the ill-fated lovers intertwined. Inside the lovers is a surprise best not described in a family magazine. The coffin is even more surprising. It disassembles into 84 pieces that can be refitted to make complete place settings for two: 23 pieces of silverware, four wine goblets, four candlesticks, four ashtrays, a man's dinner ring, a woman's dinner ring and a chafing dish that is larger than the coffin.

*Columbia Jet,* shown in the upper illustration on page 25, was commissioned by Iberia Airlines as a gift for its executives. The bird is Picasso's dove of peace. Its body is a water pitcher that pours from the dove's beak when a handle is raised. The ornate pedestal is a drinking glass. Unfold the dove's wings and landing gears descend from the wing tips. Another useful Berrocal is

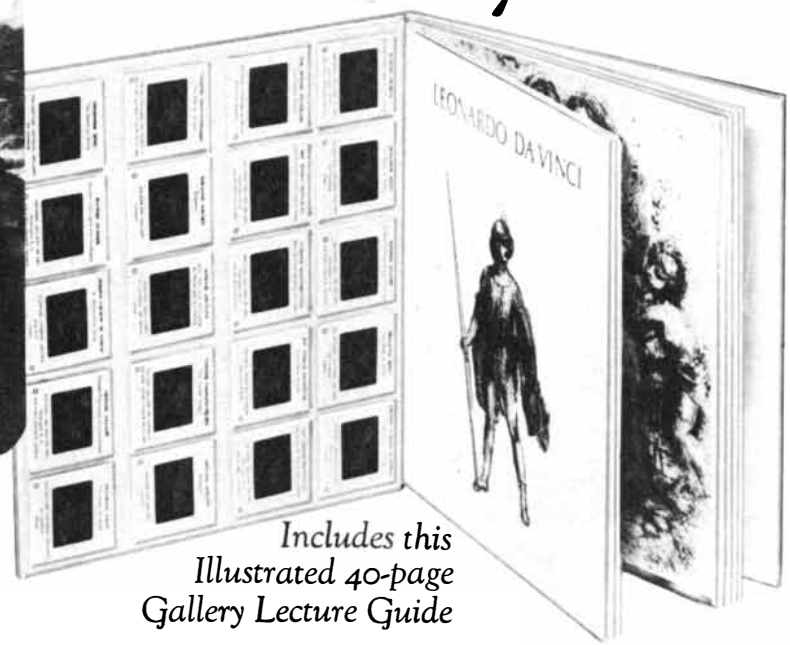


*Berrocal's* **Mini Maria**

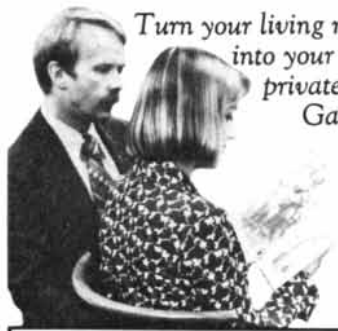


***An interior element of* Mini Maria (left) *and its five pieces* (right)**

# 20 Color Slides of da Vinci Masterpieces for only $1 (comparable value $20)




*Includes this Illustrated 40-page Gallery Lecture Guide*



*Turn your living room into your own private Art Gallery*

**HERE'S WHAT YOU GET FOR ONLY $1**

LEONARDO DA VINCI, a handsome Album with 20 beautiful color slides, including the beautiful and enigmatic "Mona Lisa" (shown above), radiant in her true colors and actual size... the tenderly evocative "Madonna and Child with St. Anne"..."St. John the Baptist"... "The Adoration of the Magi"... stunning details of the legendary "Last Supper"... and other priceless art treasures by Leonardo. The Album also contains an illustrated 40-page guided lecture by Renaissance authority Helmut Wohl, who gives you revealing background on the paintings and the enigmatic genius who created them.


**The Color Slide Program of the Great Masters**

Come with us. To meet the legendary Leonardo. And share in the exciting discovery of the most celebrated masterpieces of all time.

Come to Paris, and behold the enigmatic smile which has fascinated men for centuries . . .

Visit Florence, and discover young Leonardo's first spark of brilliance in a painting that shamed a great teacher . . .

Travel to a refectory in Milan . . . to stand before the most unforgettable fresco in the world and sense the impending martyrdom of Christ.

LEONARDO DA VINCI—the complete Art Album with 20 full-color slides and handsomely illustrated Gallery Lecture Guide—is yours for only $1 as your introduction to the *Color Slide Program of the Great Masters.* There is no obligation to buy anything more.

### View the priceless works of art in the full glory of the originals

Each lavish Album contains twenty color slides devoted to the works of a famous master. Created by art photography specialists in Florence, Italy, these exquisite 35mm. slides achieve what small print and art reproductions cannot—for they capture the authentic colors of the original works . . . with every subtle line and detail brilliantly revealed. And they permit you to *view each priceless painting on your home screen in the full glory of the original.*

In addition, bound into each Album is an illustrated, 40-page Guide in which a noted authority points out important aspects of each painting and shares with you fascinating glimpses of the artist himself.

As you view the paintings projected on your wall and read the absorbing commentary, you and your family will discover a whole new dimension in the enjoyment of art. Indeed, it is much like taking your own "private tour" with an art expert at your side.

### The Program is flexible—the cost is low

Approximately every six weeks, you are offered a complete new Color Slide Album of 20 distinctive slides and an illustrated lecture. Forthcoming Albums will be devoted to the lives and works of such great masters as Michelangelo, Winslow Homer, Rembrandt, Gauguin, and Picasso, Chagall, Van Gogh, Lautrec, etc.

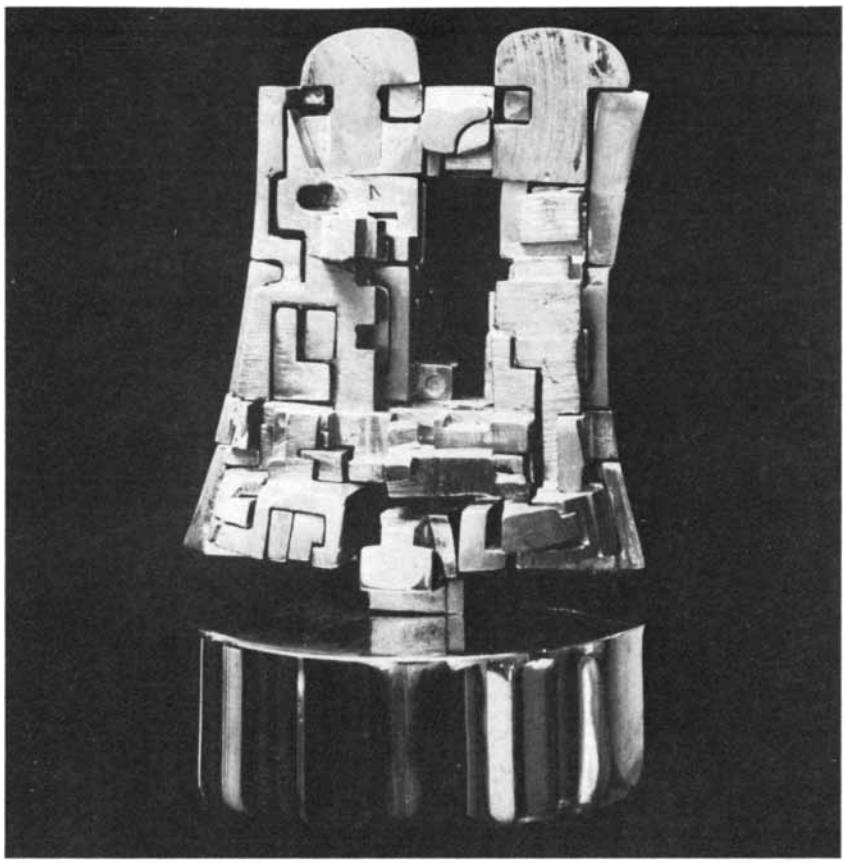
If the slides were available from museums, you would expect to pay as much as $20 a set. Yet, as a member of this new Program, you pay only $7.95 for each complete Album—including 20 color slides and 40-page illustrated lecture. And you may take as few or as many as you like.

### Accept this $1 demonstration offer

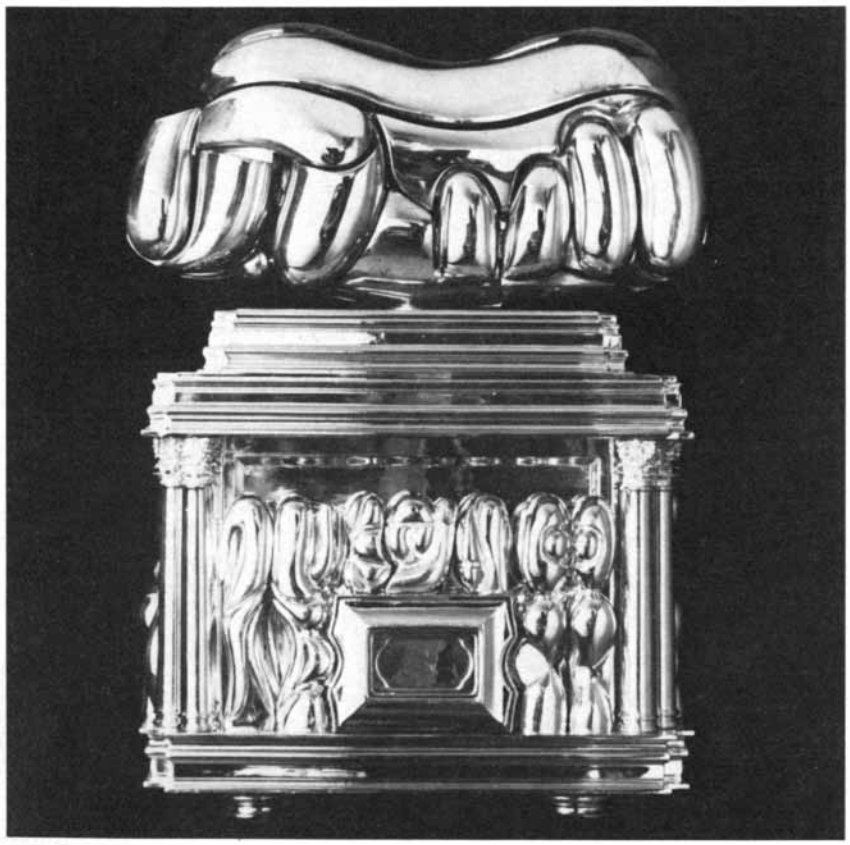
To introduce you to the *Color Slide Program of the Great Masters,* we will send you *Leonardo da Vinci,* complete, for 15 days trial. If you are delighted with the demonstration, send only $1, plus a few cents for tax and shipping. If not thrilled with the introductory Album, simply return it within 15 days; and you will be under no further obligation. Mail the attached card today.

**If you do not have a projector:** here's an unusual opportunity. We can offer you the sensational Kodak Carousel BC 604 Projector supplied with a 500-watt lamp, and 4-inch f/2.8 Ektanar lens. This top-rated projector shows all regular 2" x 2" 35mm. slides up to 1/16-inch thick. It has dependable gravity feed and you have instant access to any one of its 140-slide capacity at any time. List price $114.50—subscribers' price only $85.95, plus tax. Remittance must accompany trial subscription form. If you are not completely delighted with the projector, you may return it in 15 days for a full refund.

**If Reply Card has been removed, write Color Slide Program of the Great Masters, Dept. BC-109, 1221 Avenue of the Americas, New York, N.Y. 10020.**



**Goliath** *partially disassembled*



*A sculpture of Romeo and Juliet on top of another sculpture depicting their coffin*

*Paloma Box,* a sculpture that is about a foot high and wide and opens into a jewel case with 16 felt-lined drawers. A circular mirror can be raised and then opened to reveal a sculpture of the head of Picasso's daughter Paloma. The pieces making up the head include two bracelets and two belts. Berrocal's *Il Cavallo,* shown in the lower illustration on the opposite page, has a different kind of flexibility. The 14-piece horse has legs articulated so that it can assume different positions.

I have described only a small portion of Berrocal's work. His largest sculpture is his *Homage to Picasso,* which is 18 feet long and weighs 18 tons. It is now on permanent display in the Picasso Gardens in Málaga. A miniature bronze version, *Siextasis,* consists of 20 pieces locked together by 20 small magnets. Another large sculpture, Berrocal's tribute to his good friend Dali, is in Madrid. A miniature bronze version, called *Dalirium Tremens,* will soon be available.

Berrocal's exclusive agent in the U.S. is his longtime friend and admirer James Smith Rudolph. Rudolph owns the two Centicore bookshops in Ann Arbor and Centicore Arts, International (336 Maynard Street, Ann Arbor, Mich. 48108). For many years Rudolph has supplied galleries and stores with Berrocals and arranged for about a dozen one-man shows in various cities each year. There is a show in Miami this month, and one coming up in Cleveland in February. The prices of Berrocal's works begin at $180 for the micro (pendant) multiples and rise rapidly into many thousands of dollars for the mini multiples.

There are no screws or bolts in any Berrocals except the pendants. The pendants hang by a key piece that screws in to prevent the pendant from accidentally dropping off its chain. All gold and sterling-silver mini multiples have key pieces that lock into place with an ordinary key so that the sculpture cannot be disassembled without the owner's cooperation.

Let me anticipate an objection. What, you may ask, does the fun of taking sculpture apart and putting it together again like a Chinese puzzle have to do with art? In a sense the answer is "Nothing," but that is not the whole story. The visual beauty of art has always been combined in countless ways with other values: the sexual emotions aroused by nudes; the sentiments evoked by landscapes, seascapes and family portraits; the rhetorical function of political and religious art; the didactic value of textbook illustration; the humor of comic art; the physical comfort of chairs, beds and sofas that are designed to be beautiful; the utility of tables, vases, bottles, cups, dishes, silverware, cars, houses, quilts, ships, watches, tools and so on. Berrocal's unique achievement is the combining of visual and tactile plea-

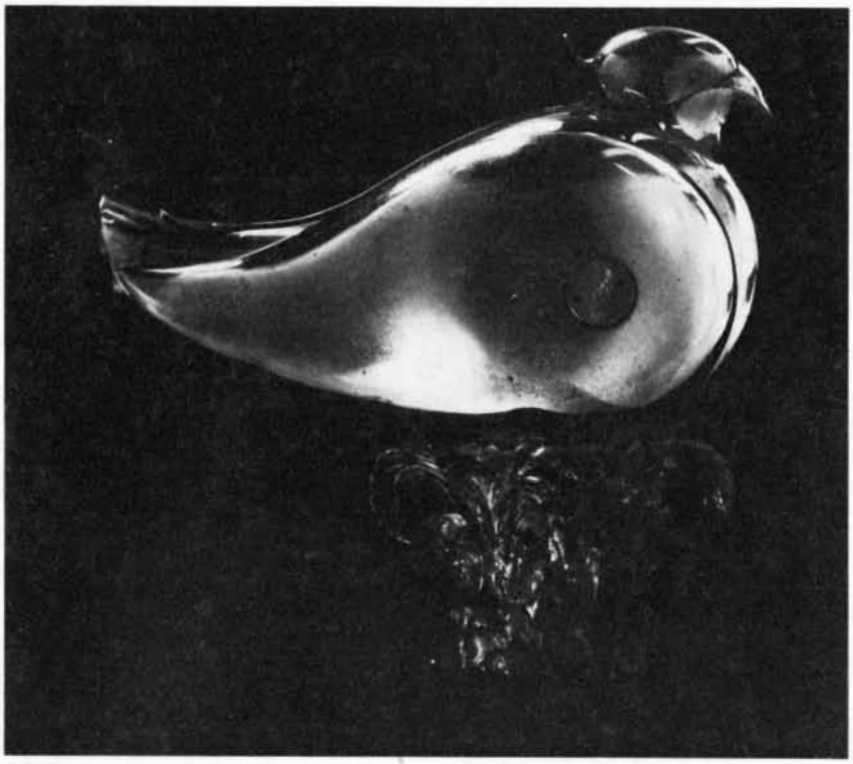
sures with the intellectual play of a mechanical puzzle. If you do not enjoy that particular combination, then a Berrocal is not for you.

In the Quadling region of Oz (as described in *The Emerald City of Oz*) the city of Fuddlecumjig is inhabited by a whimsical race of people called the Fuddles. Each Fuddle is made of hundreds of fantastically shaped pieces of painted wood that fit together like a three-dimensional jigsaw puzzle. Whenever a visitor approaches, a Fuddle clatters into a heap of disconnected parts so that the visitor will have the fun of fitting him or her together again.

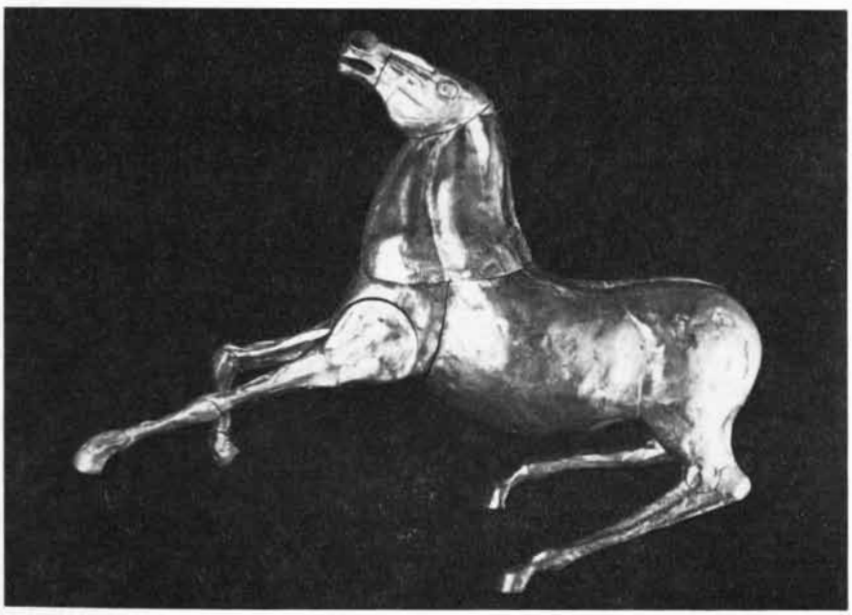
"Those are certainly strange people," Dorothy's Aunt Em said when she met the Fuddles, "but I really can't see what use they are, at all."

"Why, they amused us all for several hours," responded the Wizard. "That is being of use to us, I'm sure."

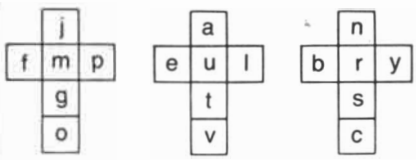
"I think they're more fun than playing



*Berrocal's* **Columbia Jet**



*Berrocal's* **Il Cavallo**




*Solution to last month's calendar-cubes problem*

solitaire or mumbletypeg," added Uncle Henry. "For my part, I'm glad we visited the Fuddles."

Here are the answers to Dr. Matrix' number puzzles in this department last month.

1. The problem is to prove that no number starting with 9 and followed by digits in cyclic descending consecutive order (with or without 0) can be prime. All primes except 2 must end in 1, 3, 7 or 9 because if the last digit is even, the number is divisible by 2, and if the last digit is 5, the number is divisible by 5. It is easily shown, by adding digits and reducing to the digital root, that if the sequence terminates with 1, 3, 5 or 7, the digital root must be a multiple of 3, proving that the number also is a multiple of 3.

2. In an alphabetical list of the English spellings for the integers 0 through 1,000 the next to last entry is two hundred two.

3. In an alphabetical list of the Roman numerals from 1 through 1,000 the last entry is XXXVIII, or 38.

4. The smallest number whose name contains all five vowels plus *y* is one thousand twenty-five.

5. The next term in the sequence $10^3$, $10^9$, $10^{27}$, $10^2$, $10^0$... is $10^{.60206...}$, or 4. The first term, "one thousand," is the smallest positive integer whose English name contains *a;* the second term, "one billion," is the smallest to contain *b;* the third, "one octillion," is the smallest to contain *c;* the fourth, "one hundred," is the smallest to contain *d;* the fifth, "one," is the smallest to contain *e,* and the sixth, "four," is the smallest to contain *f.*

6. The "unfolded" cubes shown above demonstrate how three cubes can be given lowercase letters so that the cubes can be arranged in a row to spell the first three letters of any month. Note that this is possible because *u* and *n* and *p* and *d* are inverses of each other. In fact, it is possible to leave two faces unlettered, since *b* is a recognizable *g* when it is turned upside down, and *u* can double for *c* when given a quarter turn.

Many readers quite properly chided me for mentioning in my September column on hyperbolas that tossed objects follow paths that are parabolas. The paths are very close to parabolas, but strictly speaking (and ignoring air resistance) a tossed object follows an elliptical orbit around the earth's center of gravity.

*(No. 3 in a series on history-making "easy performances")*


Legend has it this expression for "easy performance" was given birth during the Napoleonic wars. Seems, for the want of manpower, apprentice chef Jacque Brissoe was relegated to aiding the engineering corps in erecting a troop bridge of logs over the Loire River. A complete novice in these activities — Jacque continuously tumbled from the logs into the river. After several dippings and upon removing his army issue hip boots, the sopping chef discovered he had "netted" three nice trout. Elated, he roasted the little devils over the campfire and ceremoniously prepared a sauce of goat's milk, watercress, shallots and horseradish. His proudly displayed gourmet creation caught the eye of the "little Colonel"— who took and ate them. While picking his teeth "Nappie" asked of Jacque, just how during his bridge building did he have time to go fishing? Jacque replied, *C'est aussi facilekque de tomber d'une bûche!* (English translation) T'was easy as falling off a log! Napoleon rewarded Brissoe with a larger pair of boots.

We, at LOMAC®, can't let time-honored "easy performance" expressions like this one just lie there. They're all too apropo when referring to how easy it is to *teach, work* or *use* our ADAM computer (the first *program-it-yourself in English* office computer). To bear this out, ADAM won't operate on costly intangible outside "software." Won't operate to confusing computer languages. ADAM does work very easily to your business words, to words from the English dictionary, or whatever words you wish to "tell" ADAM. That is, with ADAM you continue to run *your* business *your way*—minus any log jams.

*NOTE: ADAM is a* self-organizing in English *computer product, not a computer system. This is a direct opposite of "systems in English," which are words others "use" to sell their established system computers (ones that function only to pre-programmed computer packages).*

Our quest for renowned "easy-performance" utterances has produced some surprising historical facts on the beginnings of (1) No Sweat (2) It's a piece of cake (3) It's a snap, etc. Our ADAM literature espouses why we're gathering all these performance similies. So, as they say, keep your *facts* coming— with all that the big advertisers are spending, we can use them.

Send your findings to Ms. Omokawa at our address below; if we select yours to help tell the ADAM story you could receive the 30″ wide hand-carved log wall-hanging pictured above (or some other item listed in prior ads). Naturally, times being what they are, we have to say all entries become our property. O.K.?

(ADAM)

**LOMAC**

**Logical Machine Corporation**
**1294 Hammerwood Ave.**
**Sunnyvale, CA 94086**
**(408) 744-1290**

**Ms. Omokawa:**

☐ That *easy bit* appeals to me. I'm tired of doing things the hard way. Please send your ADAM literature.

☐ I've thoroughly researched this "easy performance" quote:_____. Despite the subject, it wasn't easy. See attached.

Name_____

Company_____

Address_____

City_____State_____Zip_____

# BOOKS

## *Equality, historical supernovas, the acoustic environment and guides to the stars and wine*

by Philip Morrison

THE SCIENCE AND ETHICS OF EQUALITY, by David Hawkins. Basic Books, Inc., Publishers ($7.95). "The belief in fixed hereditary potentialities is about on a plane with Aristotle's version of the laws of motion. Both generalize from a range of experience that is limited and uncontrolled. Rocks do fall and fire rises, while what is up above goes round in circles." In this clear, quiet voice, never so earnest as to mask the sound of reason, never so analytic as to lose the depth of engagement, the brief set of lectures by a Colorado philosopher of science, experimental educator and mathematician aims squarely at the issue of human equality. That deep question "appears and disappears like a comet, to return again ... altered ... by every perihelion passage."

The question is blazing today, sporting a double-helical tail, demanding attention. We cannot forget the last terrible passage, the tail then a hooked cross, with poor men and princes alike dead in heaps. Hawkins grounds his wise and luminous argument on the classical philosophers who built the values of our culture and on the demands of a moral policy for public education. But the power of his lectures—which were given in the name of John Dewey—comes from their weighty content of science, while their felicity of style comes from their inductive structure. Biological equality is an axiom of first approximation, as much for old Hobbes as for Marx or Dewey. Given the deep equality of human beings, their reason and the value they set on self-preservation, Hobbes saw two solutions of the game of strategy we call the civil order—three centuries before von Neumann. One is the "war of each against all"; the other is the social contract, some species of Leviathan. We never see natural rulers, lions to pacify forever us clever apes; all our history offers demonstration. Kings or communes, a fasting Gandhi or a marching Sherman, are varied forms of the one solution. Our innate capabilities encompass all those alternatives and more.

The canary chick begins singing with a tuneless warble whose basic finch style is inborn; then radio, piano, typewriter and voices all mold the song. It remains canary song, yet it is individual. Our human self-molding capability is so broad that no mapping of conflict or education or politics can be found in instincts. Human beings are a single well-defined reproductive species "unique in its capacity for education and invention."

Of course, genotypes vary. But the nature-nurture relation is one of strong interaction, far from simple additive combination. The measures of heritability work out better for the curve of mortality than for I.Q. Indeed, longevity is related more plausibly to recessive lethals, say, than any facet of the higher functions is. It is not true, however, that U.S. blacks, for example, are shorter-lived than whites because of innate factors. We know enough facts to exclude that supposition, not least of them the fact that a black subpopulation of 25 percent ranks above the white average in this variable. And yet there is a large black population that has the same death rate at age $x$ that the white population has at age $x + 17$ years. The I.Q. data are not so very different, measured in standard deviations. There too the genetic conclusion is invalid; it is circular, rejecting by assumption the obvious inference, which is that the environment makes up the difference. The heritability index comes out at 90 percent for longevity and only 75–80 percent for I.Q. That the WISC test scores as wrong the answer given by a New Mexico Indian child ("Who discovered America? Wrong answer: *We* did") is not the main point, although it is noteworthy. The scalar test itself, not its items, is most faulted.

Additivity, Gaussian normality and scalar measures are simply naïve: they are crude Aristotelian regularities. The complexities of diverse human abilities require a measure in some function-space, a complementary relation of nature-nurture variables, a distribution more multiplicative (log-normal) than additive (normal). In the dexterous, planful game of billiards, scores are nowhere near a normal distribution. A tyro cannot make even one; an average player makes a handful; the experts go on for hundreds. Is billiards uniquely cumulative among human talents? One doubts it. In the second approximation human beings are, of course, unequal. What could equality mean in so many dimensions? Can we hope for a society in which all others are duplicates of ourselves? A similarly dreadful model is one where one person is superior to all others in every respect. In each case there would be nothing to learn, and that is the deepest error. We do not by nature outrank each other uniformly around the wind rose of all our talents and achievements.

The naïve hereditarian view leads straight to stratified education. The simplistic environmentalist view leads to a single "best" track for all, plus remedial classes. "The first sort of program tends to be strongly self-confirming, the second to fail." What we really need is a subtle, complementary, strategic, optimistic, investigative school, anxious to seek success somewhere, even among the damaged, a school aware of each potentiality, where a Helen Keller appears not as a miracle but as a consequence. Thermodynamics excludes perpetual motion not merely because inventors have failed but because we have succeeded in the rational design of many engines that burn fuel. Once we have drawn working and coherent general principles from the success of our educational designs, we can assert impossibilities with confidence. Until then we have a commitment to learn.

THE HISTORICAL SUPERNOVAE, by David H. Clark and F. Richard Stephenson. Pergamon Press ($15). All we know about the earliest recorded supernova appears on a single Chinese page, reproduced on the cover of this delightful compact monograph of mysteries unraveled. The text cites a date that is equivalent to December 7, A.D. 185, and in translation tells of a "guest star ... within *Nan-mên* ... large as half a mat.... According to the standard prognostication this means insurrection. When we come to the 6th year, the governor ... punished and eliminated the middle officials." (One hopes the astronomers were spared.) These observers watched from Lo-yang, and they were pretty veracious. We must of course test them if we are to believe their terse witness. So the authors do, using modern tables to confirm the eight planetary positions cited elsewhere in the treatise. That work is itself only one astronomical chapter in the official history of the Later Han dynasty. One need search no musty shelves or old manuscripts to find the history; it was compiled from official records by about A.D. 510 and printed time after time since then. The version we see (not explicitly referenced) is probably from an edition of about 100 years ago or from a 20th-century reprint. The book is to be found in every Chinese library, among the 26 dynastic

accounts of the past 20 centuries, summarized and official, to be sure, yet an "inexhaustible goldmine of astronomical observations, without equal anywhere in the world."

Where was the guest star in the sky, and what was it? Its 20-month duration makes it a sure candidate for supernova status; a comet can be ruled out, and no ordinary nova lasts so long. A careful check of Chinese star maps over the centuries establishes that the asterism cited (the Southern Gate) was a pair of stars in Centaurus that are as far south as can be seen from central China, allowing for precession. Where was it exactly? The text uses a word meaning "within," and we must know what that means. After all, a common but vague word meaning "in the vicinity of" was avoided. Back to the planetary reports; once again the modern positions confirm that when the Former Han observers said "within," they meant it! Their guest star is fixed within a few degrees.

How bright was it? There is no guesswork here either; one sees the arc the star traversed in the sky across the jagged horizon of ancient Lo-yang (made from a careful relief map of the area), noting that it just clears the hills. It first rose at dawn, and it must have been very bright. A comparison with modern supernovas leads to the conclusion that we need to look out one or two kiloparsecs in the sky to find some relic of the explosion that foretold the disloyalties of the bureaucracy so long ago.

Four candidates have been found in the lists of radio sources; all are possible and all resemble the expanding shock we expect, but rough distance measures of various kinds select one of them. A high-resolution radio contour map made at the Molonglo observatory in Australia, resembling an optical photograph showing a few circular arcs of glowing filament, is satisfyingly familiar from a couple of other cases, such as the remnant of Kepler's nova. The search has borne fruit: we are looking at a probable supernova of the Later Han, all from about 100 characters on a page.

Such is the style of this detective work, the collaboration of two talented young scholars working in England. Clark is a New Zealand radio astronomer and Stephenson a historian of science with sinological expertise. They exhume from the old books eight such identifications, four of them certain (including the famous Renaissance examples named after Tycho Brahe and Kepler), two probable and two possible. (They summarized their argument in a *Scientific American* article in June, 1976, emphasizing the conclusions more than the evidence.) The Chinese records are the mother lode, but confirming material is often to be found in the parallel records of Japan and Korea, all written in classical Chinese, and for a few cases also in Arabic, Syriac or medieval Latin.

The supernova of A.D. 1006 is the brightest on record, described in all the languages listed above. For the ingenious astrologer Chou K'o-ming, who was away from his capital when the supernova appeared, time for reflection allowed him to shift it, by citing its color and some fine old classics, from the category of the baleful to that of the auspicious. He was made "Librarian and Escort of the Crown Prince." A young Egyptian scholar, Alī ibn Ridwān, reported the same new star out of "my own personal experience"; the intensity of its light was "a quarter of that of moonlight." The authors present a good case to the effect that a viewer could make out the landscape clearly by the light of the star. It was admired at one Alpine monastery even though it grazed the local mountain horizon. Today this position also shows a radio source; within the past year or so optical filaments have been seen as well, perhaps the nearest recent supernova. There is a pulsar still closer in the constellation Vela; it is presumed to be a supernova relic some 10,000 years old. (More by wishfullness than by well-sifted evidence the Vela star has been associated elsewhere with a famous set of cuneiform tablets from Assur, dating from about 700 B.C. but believed to preserve more ancient material.)

Kepler's star is the last to be treated. The authors present the now classical work of Walter Baade in reconstructing the light curve of this object from the rich European record, a demonstration of the unity both of nature and of science across three centuries: Kepler's naked-eye curve of 1604 fits Baade's 1937 data beautifully, although the supernovas Baade saw were in distant galaxies and were recorded through the 100-inch telescope on Mount Wilson. It is striking that the Korean observers of Kepler's star provide concordant points. The photometric method of the 17th century—innocent even of the conventional definitions of our astronomy—was a direct comparison by eye between the guest star fading over the months and the stars and planets glowing near it. The accuracy is admirable; moonlight and atmosphere effects can be allowed for but are unimportant.

It is eye-opening to examine the asterisms of old China, each with its own identity and function in the celestial palace. Those meticulous astronomers did not neglect stars merely because they were of the fourth magnitude or fainter. They used a sighting tube to fix the distance of close approaches between fixed stars and the planets or new stars; it seems pretty clear that an "invasion" was spoken of only when the newcomer was seen within the field defined by a tube that must have opened about one degree at maximum. Such for the most part was the positional resolution, and it served stellar astronomers well for millenniums. By Kepler and Tycho's time (still before any lens had magnified the starry sky) the Europeans, with different goals and different standards, had pressed their measurements down to an arc minute or two, two orders of magnitude better—a harbinger of caravels, cannon and conquest, the technological dominion of the West, today plainly ended.

We approach the fourth centennial of the last naked-eye supernova seen in our own galaxy. Sir Fred Hoyle has somewhat tartly remarked that a wonder so accessible must be given us only when a great observer is living on our planet. We can be pretty sure that a hundred such light beams from yet unknown supernovas in our galaxy are already on the way; maybe one will arrive tomorrow! Until it appears some clear night we can learn about these extraordinary events, to which our own substance is kin, by seeking the witness of our astronomical forebears all over the world. Their integrity and their insights are made manifest in an ingenious and scrupulous book they would have enjoyed.

THE TUNING OF THE WORLD, by R. Murray Schafer. Alfred A. Knopf ($12.95). A landscape is easier to map than a soundscape, but the latter is just as real, just as much a part of our own sensory and aesthetic environment. We cannot shut our earlids; awake, we are always open to the keynote sounds that cannot be overlooked, to the foreground signals, to the old soundmarks we remember and cherish. Armed with original concepts such as these, with various notational schemes, with tape recorder and parties of student listeners and interrogators, this Canadian composer and acoustic investigator offers in print a summary, an analysis and a manifesto—learned, witty, crisp and definitely undamped.

Much has he traveled in the realms of gold, citing with equal aptness Hesiod, Tolstoy, Vitruvius and the Canadian painter Emily Carr. Schafer has not stayed home reading, either, but has listened to the gurgles and rumblings of hot springs in New Zealand, the hammers of happy tinsmiths in a Teheran bazaar and the brilliant slam followed by the sharp click of the latch on the doors of the old Paris Métro (now to be heard "only on the Mairie d'Issy–Port de la Chapelle line").

In this rich and protean style Schafer has made a clear case. The world of men's hearing was once ruled by the sounds of wind and water. The sounds of animals were perhaps next in salience. Then came sounds of old-time rural life, discrete and individual: the hammer blow or the post horn. The dominant church bell was joined by the

clock bell as the rule of the soundscape began to pass from the sacred to the secular. Noise has always been linked to power, and the mill, the locomotive and the "plangent voice" of the airplane became sovereign. (Note the progression from the distinguishable rhythm to the steady, unresolved line of granular intensity.) Now fully in the electrical age, we find white noise deliberately spread, together with the drone of every power source and sink, the infamous Muzak and the "schizophonia" of a billion loudspeakers that have torn music and voice out of any context.

After this indictment from history Schafer proceeds to analysis, by way of a three-dimensional notation and lots of statistics. Antinoise legislation (as old as Caesar) has been effective less for what it prevents than for the data it offers on what sounds rule. (Church bells, once the delight of Christendom, now begin to come under ban.) The question is a sociopsychological one, and legislation is capricious and oversimplified at best (although Moscow and Göteborg have actually been able to stem the steady rise of the city noise level). In Vancouver, Schafer and his colleagues measured the capabilities of fire-engine sirens, from 1912 veterans to the newest 1974 wonder. Their yield had risen on the average about half a decibel per year, a simple and clean confirmation and extension of many costly direct surveys that could not reach back over such a long base line. The inference is easy: the higher the ambient noise level, the more piercing the required alarm.

Profound change cannot be brought by proscription alone. Positive measures must be included: public "soniferous gardens," the active preservation of soundmarks of distinction, repairs to the damaged soundscape, from the telephone ring ("a bad pun on the name of its inventor?") to car horns, and above all a sense of acoustic design everywhere. The ears of modern architects are "stuffed with bacon." Ear-cleaning (largely attention-cleaning) exercise is possible, and through it we can restore, in a quieter world, the art of sonic invention.

In a Neolithic structure in Malta, in ancient Rome, in medieval Isfahan, in the abbey church of a small town not far from Zagreb there are hollow-sounding vases—Helmholtz resonators—installed in the walls, deliberately enhancing the sound of the structure. In the new Sydney opera house there are such devices too (among a multitude of acoustic sins). Hope is still alive, but Schafer's deeper hope for the recovery of positive silence, the restoration of "contemplation as a habit and skill," seems not soon to be fulfilled. The "tuning of the world" was a theme of the mystic Robert Fludd, in a stance that is Schafer's own occasional indulgence. (They hear harmony between the elements and the Pythagorean ratios.) Today we hold a more diverse understanding, but we can expect to make our way toward audio order, if not harmony, once we can arrange our social world closer to reason.

DEEP SPACE 3-D: A STEREO ATLAS OF THE STARS, by David Chandler. P.O. Box 309, La Verne, Calif. 91750 ($6.95). STAR EXPLORER, a planisphere by Dorothy A. Bennett. 8339 Terrace Drive, El Cerrito, Calif. 94530 ($1.50). THE NIGHT SKY, a planisphere by David Chandler. Sky Publishing Corporation, 49–50–51 Bay State Road, Cambridge, Mass. 02138 ($3). THE BEGINNER'S GUIDE TO THE SKIES: A MONTH-BY-MONTH HANDBOOK FOR STARGAZERS AND PLANET WATCHERS, by Clarence H. Cleminshaw. Thomas Y. Crowell Company ($7.95). With the long nights of the winter solstice at hand for most readers, we offer our annual selection of stargazing aids. First of all is a unique star map, a map that breaks the crystal spheres! What Chandler, an ingenious teacher of astronomy, has made for us is a whole-sky map presenting the visible stars (all those brighter than magnitude 4.5, as well as some others) on a series of 14 stiff cards. The map shows black stars on a white ground, stars joined into geometric patterns according to constellation. The better-known stars, as well as the constellations, are marked by name. The scale is moderate, the Big Dipper almost an inch and a half long.

Nothing new there. Below each index map, however, we see a black sky-square with the same stars marked by white dots of varying size, free of any lines or legends. More than that, the square is doubled. Together the two subtly different squares present to the two eyes, by way of the folding cardboard viewer with two glass lenses, an unprecedented stereoscopic view of the stars. We see the night sky as we might if we were celestial giants, a tenth of a parsec across the eyes. Bright Orion lies far away, but in the foreground Procyon, Sirius, Capella and the Twins stand close. Alpha Centauri is projected so near as to tempt one to pluck its image out of heaven.

The nearest 100 of these conspicuous stars are numbered on the map in serial order. The close sunlike stars, such as Tau Ceti and Epsilon Indi, targets of the classical searches for radio signals, are of course high on that list. The parallax data were those compiled in the well-known atlas catalogue by the Czech astronomer Antonin Becvar. This three-dimensional sky is no practical aid to star watching, but it offers an experience striking in itself to anyone with binocular vision and is a new resource for the imagination, somehow kin to the first photographs from space that showed the whole round earth: the expected but abstract rendered newly to the senses.

Vestigial astrolabes of our time are the two simple cardboard devices listed next. A disk bears the map of the starry sky (in stereographic projection); it is arranged to turn under a masking opening that reveals only the part of the sky that is above the horizon at a given date, hour and latitude. The first planisphere listed is the cheapest sky aid around, a bargain in its 14th edition, a standard available at museum shops and the like for 40 years. The graphics are clear and attractive: white stars of several dot sizes on a black sky with blue legends. The latitude chosen is 40 degrees north, a rough mean for the U.S. population.

The second planisphere, quite new, offers rather more. Its sky is white, and many more stars are marked, together with a number of sky sights for binoculars and the ecliptic and the equator. The novelty is all but hidden: the reverse side of the disk has a projection of the full sky of the Southern Hemisphere, that other pole flanked by the Clouds of Magellan. In the present version this other sky is masked except for the part just south of the equator, repeated on the reverse to make amends for the strong distortion inevitable in the conventional view of more than a hemisphere on the northern face. The opportunity to have a full-sky pair of planispheres is offered to any user with a simple scheme and a pair of scissors. The publishers now supply a choice of openings suitable for three northern-latitude belts; if they have not yet modified the device for Down Under, they are likely to do so soon.

Finally, we have a real book by a veteran planetarium director, also a Californian. Although tolerant of the beginner, it is in fact rather more of a handbook, useful for general reference. Twenty-four maps of the black night sky are contained, each month's stars drawn twice, for mid-evening and for predawn hours (for the Los Angeles latitude). Planet locations are given in tabular form with each map for the next decade or so, and the dates of the new and the full moon (but not the times) are listed for 20 years ahead. The lore of the constellations is told in brief, with a list of the meanings of the names of 40 stars. Maps locate the outer planets in orbit for many years ahead, and a decade of Mercury's and Venus' stints as morning and evening stars is also tabulated. All these data and many diagrams are conveniently arranged, if not quite automatically available to the uninitiate. What one cannot learn from such a simple long-time listing is the occasional striking case of a close approach between planets or of planets to bright stars or to the moon; otherwise the book

is an effective guide to the shifting sky for the long pull.

THE WORLD ATLAS OF WINE: A COMPLETE GUIDE TO THE WINES & SPIRITS OF THE WORLD, by Hugh Johnson. A Fireside Book. Simon and Schuster ($7.95). A topical atlas as interesting, comprehensive and handsome as any to be found, this book celebrates that double crop, the vine and the yeast. Out of the juice of the grape, yeasts make wine; chemical man distills the more dangerous spirits. France and Italy are of course the great lands of wine, but every region of vineyards, from Adelaide to Tbilisi, from Napa to Santiago, gets its due here. The famed communes and districts of the major vineyard nations are treated in detail by the score, each with a map, a color photograph of the landscape, a page of history and appraisal and even a set of decorative labels of the best-known vintners. Take Chianti, that delicious world of endless steep little hills fruitful with vine and olive ("the olives are losing") that stretches from Florence south to Siena. The map shows the circling contours and little villages of the district on the scale of a city map, a couple of miles to the inch. On the map stand the names of the producers of that famous wine, which by law must come from the district (although law does not always rule a sometimes "disgraceful" reality).

Another map traces the movement through France for 30 years of the root beetle *Phylloxera vastatrix,* which came from America a century ago to destroy the entire vineyard of Europe. Every vine in Europe today is grafted to an American rootstock that is adaptively immune to the pest. Would you have a dry wine? Merely allow the yeast to work to natural completion. To retain some sweetness, filter the living cells out of the fluid or add alcohol or sulfur to inhibit fermentation. (These are only the simplest methods.)

Wine is so expertly and lovingly studied by scientists and connoisseurs alike that the book becomes a source of microclimatology in practice. A set of four stunning, colorful small maps of the Rheingau in Hesse show distributions of soil, sunshine, frost and wind speed in fine detail. And yet the experts complained that wind speed is too crude a measure; wind direction is important because the vines are planted in rows, and rows that lie athwart the cold east wind make a difference to the crops. A clear and knowing text on all aspects of wine from the consumer's point of view, together with economic tables and statistics, fully complements the maps in a book that can add cheer to many a winter night. The work dates from 1971, but this paperback version is a more recent bargain.

# "It's 52° below zero up here, but, crazy as it sounds, we need this machine to keep the ground frozen."

"When you drill for gas north of the Arctic Circle, the way we're doing here at Canada's Mackenzie Delta, you expect a lot of problems," says Bob Toole, drilling superintendent for Gulf.

"But this one's a real fooler. It's a freezing unit. Even at 52° below zero, we need it to keep the ground around the wellhead frozen.

## Hot mud

"The problem is that when the drilling mud comes up from the bottom of the hole, maybe a mile down, it's hot enough to melt the permafrost that's holding up the whole rig.

"If the permafrost melted, the hole would get bigger and bigger, and the operation would have to come to a halt.

## Frozen pipes

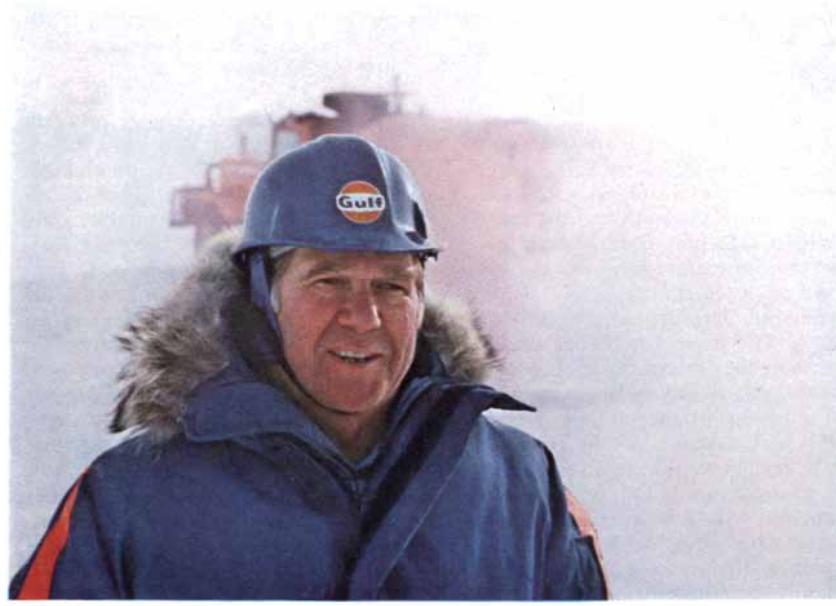
"Our freezer keeps the top thirty feet of the casing around the drill pipe at temperatures below freezing, so that doesn't happen.

"This country is probably one of the toughest spots on earth to drill for natural gas. But we're drilling the wells. We're meeting the challenge."

Gulf

**Gulf people: meeting the challenge.**

"At temperatures as low as this, you have to invent new ways to do almost everything."




Left to right: Bob Toole, roustabout Jacob Kuhoktak, and the freezer.

Gulf Oil Corporation

Established 1845 SCIENTIFIC AMERICAN January 1978 Volume 238 Number 1

# The Carbon Dioxide Question

*Human activities are clearly increasing the carbon dioxide content of the earth's atmosphere. The question is: Will enough carbon be stored in forests and the ocean to avert a major change in climate?*

by George M. Woodwell

In the century and a quarter since 1850 human activities have increased the amount of carbon dioxide in the atmosphere of the earth from 290 parts per million or less to slightly more than 330 parts per million. Perhaps a fourth of the total increase has come within the past decade. By the year 2020, if present trends continue, the amount of carbon dioxide in the atmosphere could approach twice the current value. Until recently the increase was commonly attributed to the burning of fossil fuels. Now there is evidence that it may be due in equal degree to another source: the worldwide destruction of forests.

Although carbon dioxide is only a trace gas in the atmosphere of the earth, present at a concentration of about .03 percent by volume, it plays a possibly critical role in controlling the climate of the earth because it absorbs radiant energy at infrared wavelengths. Heat trapped in this way has a large potential for altering the world climate substantially. And quite apart from possible effects on the climate, the carbon dioxide in the atmosphere also plays a critical role as the source of the carbon that is fixed in photosynthesis by green plants and therefore provides the basis for all plant and animal life.

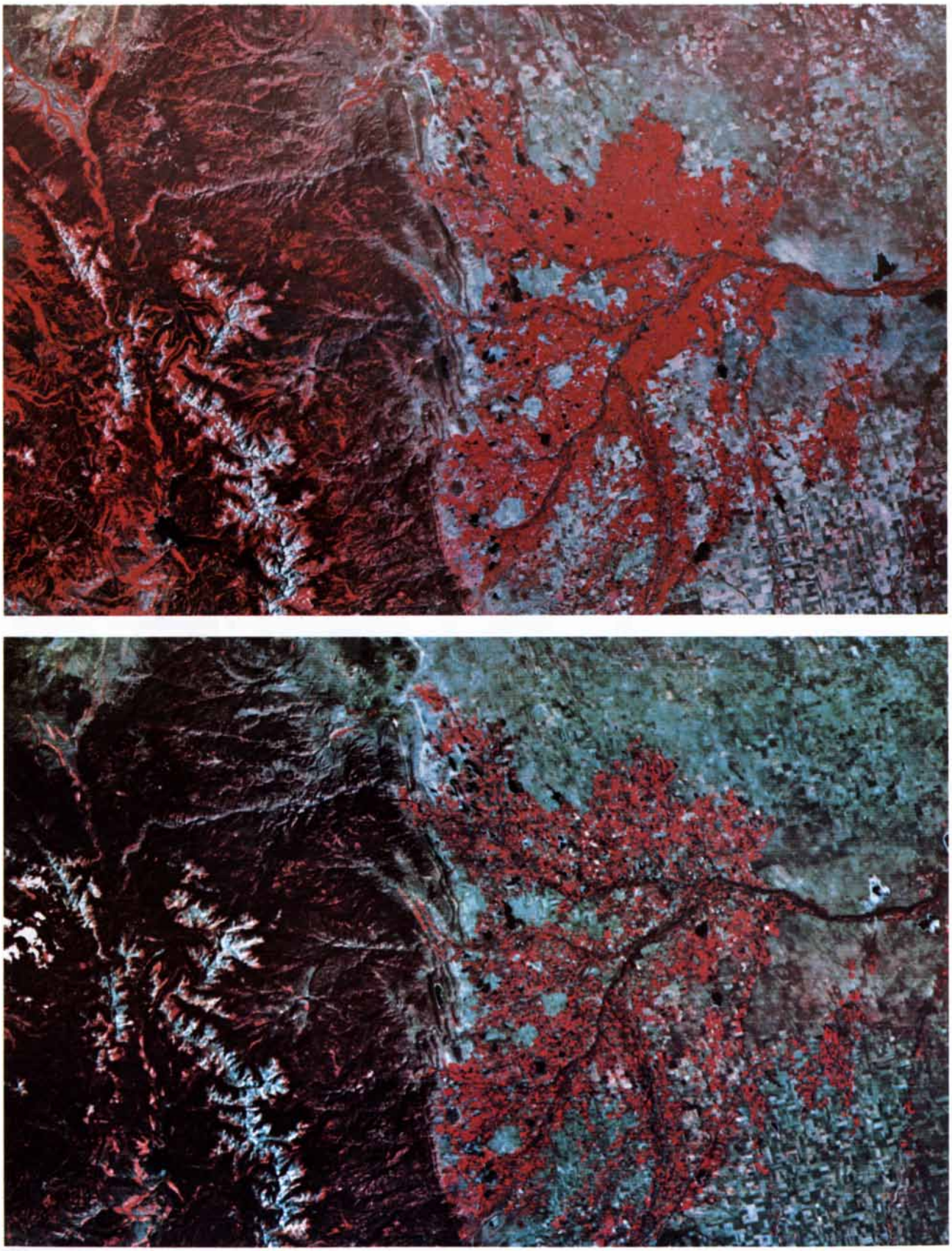
Mankind therefore faces a historic dilemma. The human activities that are increasing the carbon dioxide content of the atmosphere promise to bring a general warming of the climate over the next several decades. Although one cannot be certain of how much the climate will change, or of the precise mechanisms that will be involved, the results of a steadily rising amount of carbon dioxide in the atmosphere will almost certainly be destabilizing. An increase in the average world temperatures will probably enlarge the area of the arid zones and significantly affect agricultural production.

The other horn of the dilemma is that the kinds of corrective action that might be contemplated would surely have effects that would be equally destabilizing. The most obvious corrective action would be a major reduction in the consumption of fossil fuels. Equally important would be measures to lower the rate at which the forests of the world are being reduced or cleared by logging, by the expansion of agricultural and grazing lands, by toxification and by other consequences of industrial development. A major effort to change the balance of land use between agriculture and forest, in addition to an effort to restrict the burning of fossil fuels, would so upset established patterns of social and economic development as to be equivalent to the drastic changes in the human condition that a warming of the climate might lead to.

Although the carbon dioxide problem has been with us for more than a century, unambiguous data on changes in the carbon dioxide content of the atmosphere have been available only since 1958. In that year Charles D. Keeling of the Scripps Institution of Oceanography established a continuous carbon dioxide monitoring station on the volcano Mauna Loa on the island of Hawaii. Mauna Loa was chosen because it offered an opportunity to study the carbon dioxide content of the mixed air of the troposphere, or lower atmosphere, in the middle latitudes. The records now available from Mauna Loa and other stations show two clear patterns. First, there has been throughout the period since 1958 a regular upward trend in the carbon dioxide content of the atmosphere. The amount of the increase at Mauna Loa has been about .8 part per million per year, although there have been obvious variations in the rate of increase. Second, there is a systematic oscillation in the carbon dioxide content of the atmosphere correlated with the seasons. The carbon dioxide content rises to a peak in late winter, usually April in the Northern Hemisphere, and falls to a minimum at the end of the northern summer, in late September or October. The data from Mauna Loa are the longest and most accurate continuous record of carbon dioxide concentration ever made anywhere in the world.

Records of carbon dioxide concentration have been kept for various periods at the South Pole, in Australia, at Point Barrow in Alaska, on Long Island in New York and at other locations. Investigators in the U.S., Sweden, Australia and elsewhere have also sampled the atmosphere extensively from airplanes. The data all show a winter-summer oscillation with a minimum in late summer and a maximum in late winter. The oscillation follows the seasons of each hemisphere. The data also show a more or less continuous increase in the carbon dioxide content of the atmosphere, with the amount varying with time and place between about .5 part and 1.5 parts per million annually.

The seasonal change in atmospheric carbon dioxide reflects one of the most important factors affecting the atmosphere: the metabolism of the biota, or the totality of living matter. The seasonal change in carbon dioxide concentration in the atmosphere is correlated with the "pulse" of photosynthesis that occurs during the summer in middle latitudes of both hemispheres. Recently it has been recognized that the primary cause of the seasonal change is most probably the pulse of photosynthesis in forests of the middle latitudes. The emphasis is on the forests because they are extensive



**SUMMER AND FALL** Landsat satellite images of an area along the eastern edge of the Rockies near Boulder, Colo., show the annual "pulse" of carbon dioxide fixation characteristic of the temperate zones. In this type of imagery the green of vegetation comes out red. In the top image, made in August, the intense red of the area to the right of the mountains at left reflects a peak of photosynthetic activity, in which plants take up carbon dioxide from the atmosphere to build themselves. In the bottom image, made in October, red has substantially faded, reflecting a diminution of the photosynthetic activity. In this area the activity is that of both forest and crop plants.

**RAIN-FOREST AREA of the Amazon Basin of northwestern Brazil that appears in this Landsat image is almost uniformly red, reflecting the intense year-round photosynthetic activity that is characteristic of tropical rain forests. Natural forest communities fix more carbon per unit area than most agricultural cropland. The area of such forests, however, is being reduced.**

in area, conduct more photosynthesis worldwide than any other type of vegetation and have the potential for storing carbon in quantities that are sufficiently large to affect the carbon dioxide content of the atmosphere.

The variation in the amplitude of the difference between the winter concentration of carbon dioxide and the late-summer concentration is consistent with this hypothesis. The difference ranges from about five parts per million at Mauna Loa to more than 15 parts per million in central Long Island. The difference drops toward the Tropics, where the seasonal pulse of metabolism is either less pronounced or absent; the difference is also less at higher elevations at all latitudes. The amplitude is substantially less in the Southern Hemisphere, apparently because the smaller continental land mass limits the area of forests. The clear conclusion is that the forests of the earth have a pronounced influence on the short-term carbon dioxide content of the atmosphere.

The cause of the long-term increase in the carbon dioxide content of the air, an increase of 10 to 15 percent since 1850, has usually been ascribed to the accelerating release of carbon dioxide through the combustion of fossil fuels. Recent analyses have thrown this assumption into doubt. My colleagues and I at the Ecosystems Center of the Marine Biological Laboratory in Woods Hole, Mass., in collaboration with R. H. Whittaker and Gene E. Likens of Cornell University, W. A. Reiners of Dartmouth College and C. C. Delwiche of the University of California at Davis, have shown that there is probably a substantial additional release from the biota. Similar findings have been published by other workers, notably Bert Bolin of the University of Stockholm and J. R. Adams and his colleagues at Rice University. The release from the biota has been chiefly through the destruction of forests and the oxidation of humus. The assumption that the increase in the carbon dioxide content of the atmosphere has been a consequence of burning fossil fuels, without regard to possible changes in the biota, has led to what now appears to be a serious miscalculation of the world carbon budget.

The difficulty has arisen from the assumption that the biota has been a sink for atmospheric carbon dioxide when in reality it has probably been a source of carbon dioxide released to the atmosphere. This probable error means that at the moment it is not possible to resolve major questions about the world carbon budget. If the biota has not been a sink for atmospheric carbon dioxide, and if the absorption of carbon dioxide by the oceans of the world is no greater than we have thought, then the amount of carbon dioxide in the atmosphere should be increasing even faster than the observations show. Obviously the esti-

mates are wrong. But where does the error lie?

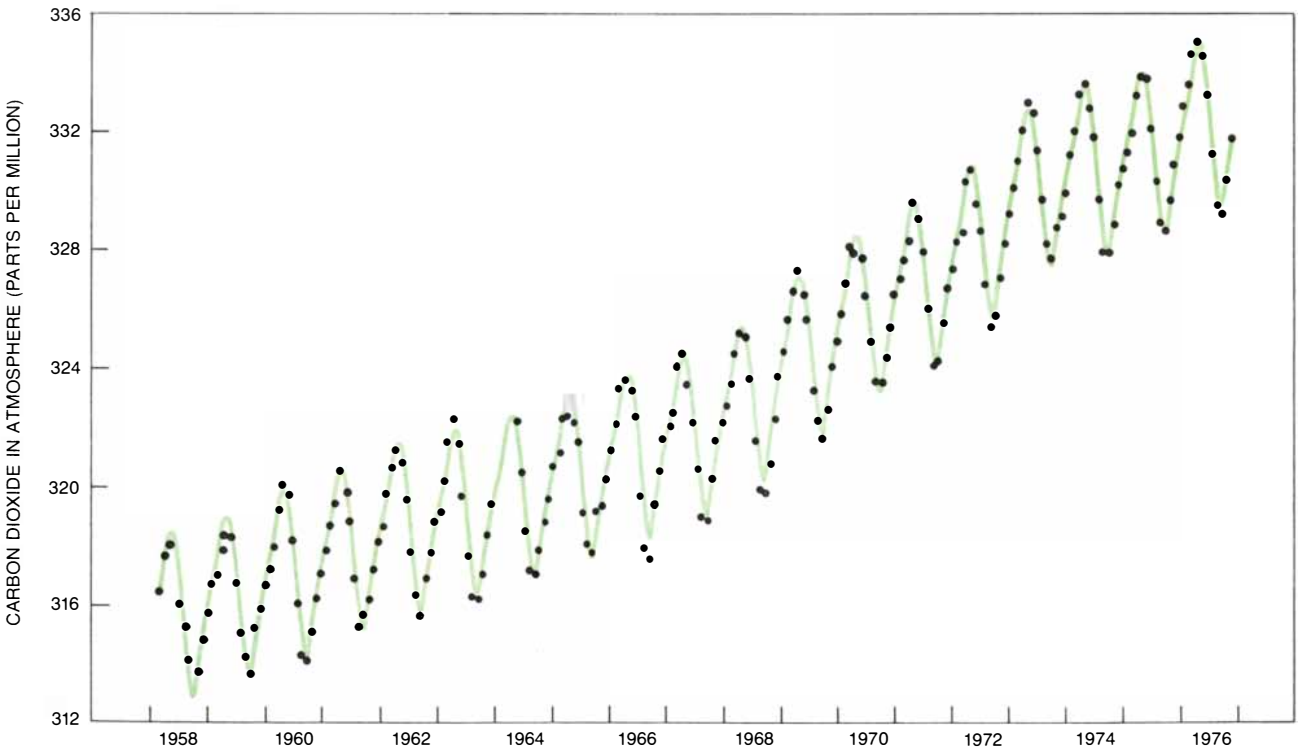
The issue can be seen more clearly by comparing the magnitudes of the pools that are more or less continuously exchanging carbon with one another. The atmosphere holds at present about $700 \times 10^{15}$ grams of carbon in the form of carbon dioxide, which is continuously being exchanged with the biota and with the surface waters of the ocean. The amount of carbon held in the total worldwide biota is about $800 \times 10^{15}$ grams, or somewhat more than is held in the atmosphere. A still larger pool of carbon, variously estimated at between $1{,}000 \times 10^{15}$ and $3{,}000 \times 10^{15}$ grams, is held in the organic matter of the soil, mainly humus and peat. The harvest of the forests, the extension of agriculture onto soils that contain large amounts of organic matter and the destruction of wetlands all speed the decay of humus, which is transformed into carbon dioxide, water and heat. The carbon dioxide released enters the atmospheric pool.

Although these three pools of carbon in continuous interaction are all roughly the same size, the total amount of carbon held in the oceans is much larger. The largest part of the carbon is in the form of dissolved carbon dioxide, which is a part of the carbonate-bicarbonate system. The total in this pool approaches $40{,}000 \times 10^{15}$ grams when the waters of the great oceanic basins are included. On a time scale measured in thousands of years the carbon dioxide content of the atmosphere may well be determined by the equilibrium established with the inorganic carbon of the deep ocean. The rate of exchange between the atmosphere and the ocean as a whole, however, is low. The most rapid exchanges are between the atmosphere and the mixed surface layer, roughly the top 100 meters above the colder abyssal waters. The surface layer contains about $600 \times 10^{15}$ grams of inorganic carbon. The second-largest reservoir in the ocean is in the form of dissolved organic matter (the "humus" of the ocean), which seems to be everywhere about one part per million and may total as much as $3{,}000 \times 10^{15}$ grams in the oceans as a whole.

The deep waters of the ocean vastly exceed in volume the waters of the mixed layer and hold by far the largest pool of carbon that is in exchange with the atmosphere: from $35{,}000 \times 10^{15}$ to $38{,}000 \times 10^{15}$ grams. (Excluded from this pool is the carbon in the carbonaceous sediments, which contain a far larger amount.) The capacity of the abyssal regions for absorbing carbon is virtually unlimited. The problem is that the carbon appears to move from the atmosphere through the mixed layer of the ocean into the oceanic depths very slowly.

When one tries to construct a flow chart showing the net transport of carbon from one pool to another, one finds that the available estimates vary widely in quality. The most accurately known figures are for the carbon dioxide released by the worldwide combustion of fossil fuels, currently about $5 \times 10^{15}$ grams of carbon per year, and the increase in the carbon dioxide content of the air, equivalent to about $2.3 \times 10^{15}$ grams of carbon per year. That leaves $2.7 \times 10^{15}$ grams of fossil-fuel carbon to be removed by some combination of terrestrial and oceanic processes. Let us assume for the moment that the terrestrial biota represents a stable pool of carbon, neither augmenting nor reducing the amount of carbon dioxide in the atmosphere. In that case (which I shall argue is unlikely) the ocean must provide the sink for $2.7 \times 10^{15}$ grams of carbon per year. Is that rate of removal supported by the evidence?

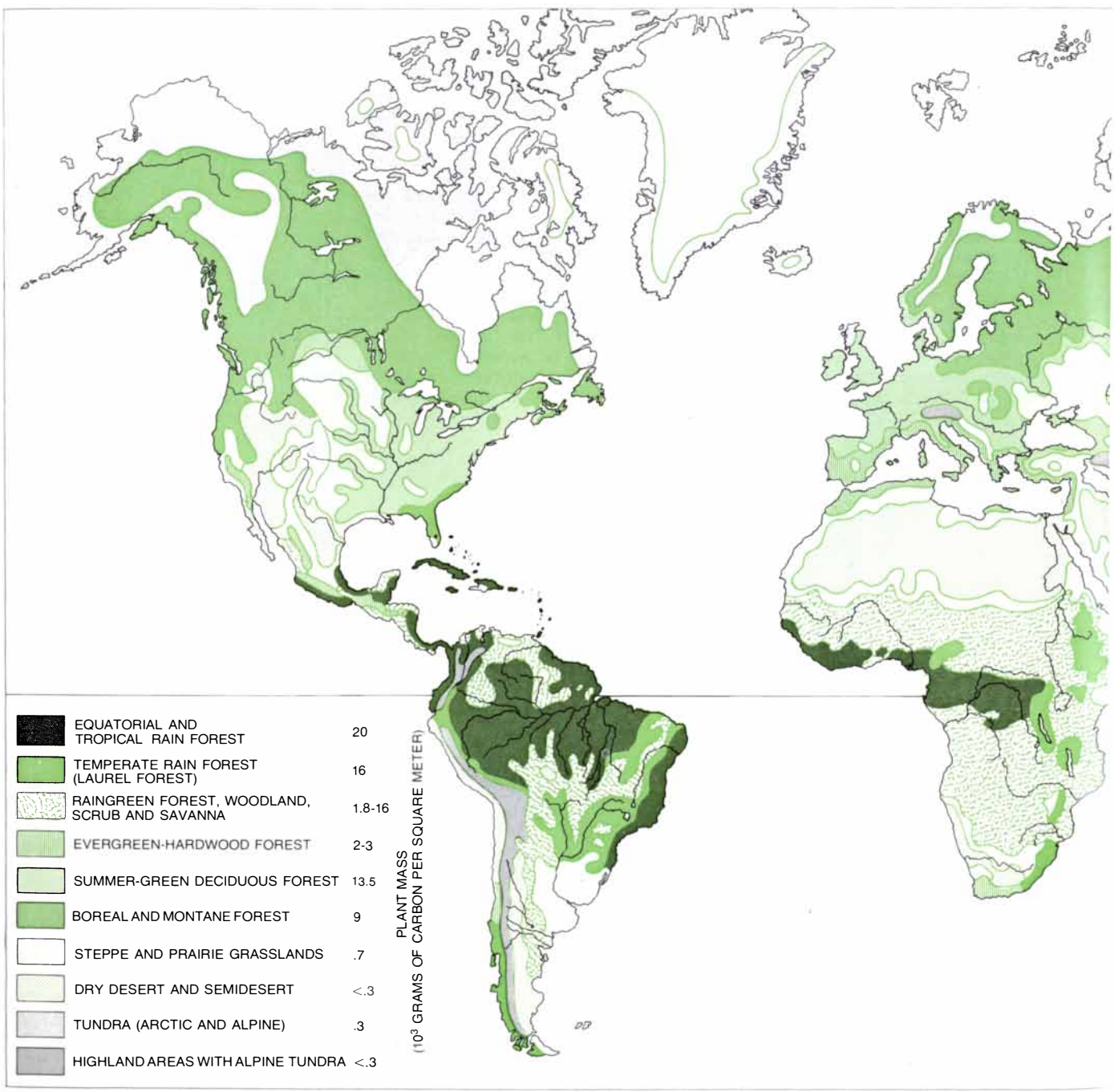
According to the best estimates of chemical oceanographers, it is difficult to explain how the ocean could absorb that much annually. Their analyses are based on a detailed knowledge of the amount of carbon held in the carbonate-bicarbonate system of the surface layer




**TREND IN ATMOSPHERIC CARBON DIOXIDE has been measured since 1958 at the Mauna Loa Observatory on the island of Hawaii by Charles D. Keeling of the Scripps Institution of Oceanography. The dots indicate the monthly average concentration of carbon dioxide. The seasonal oscillations are caused by the removal of carbon dioxide by photosynthesis during the growing season in the Northern Hemisphere and the subsequent release during the fall and winter months. The Mauna Loa measurements and those made elsewhere show that the average carbon dioxide content of the atmosphere has risen more than 5 percent since 1958. Rate of increase has varied from year to year from causes not yet known. Current rate is one part per million per year, equivalent to $2.3 \times 10^{15}$ grams of carbon.**

combined with carefully constructed models of the mechanisms of mixing in the oceans. The radioactive isotopes carbon 14 and tritium (hydrogen 3), both produced in large quantities by the atomic bomb tests of the 1950's and early 1960's, have been exploited as tracers to examine the rates at which water of the mixed layer is exchanged with the water of the oceanic depths. The studies seem to show that the mixing rate is very low indeed. The transfer of carbon from the atmosphere through the surface layer and into abyssal waters, according to some calculations, is unlikely to exceed $2.5 \times 10^{15}$ grams per year. In sum, according to these studies, the oceans seem to be barely adequate as a sink for the difference between the $5 \times 10^{15}$ grams per year of carbon currently being released into the atmosphere by the burning of fossil fuels and the $2.3 \times 10^{15}$ grams that the atmosphere actually retains.

This balance of flows has to be completely reexamined if the biotic pool of carbon, instead of either expanding or being in net balance with the carbon in the atmosphere, is actually a net source of atmospheric carbon dioxide. Whittaker and Likens have recently provided a tabulation of current information about the size of various segments of the biota. Their work shows that the largest pools of carbon in the biota are in forests. It also shows that the largest amount of net photosynthesis is on land, not in the ocean as had been assumed previously on the basis of earlier guesses about the rates of net primary production in the oceans. (Net primary production is the net amount of fixed carbon, or

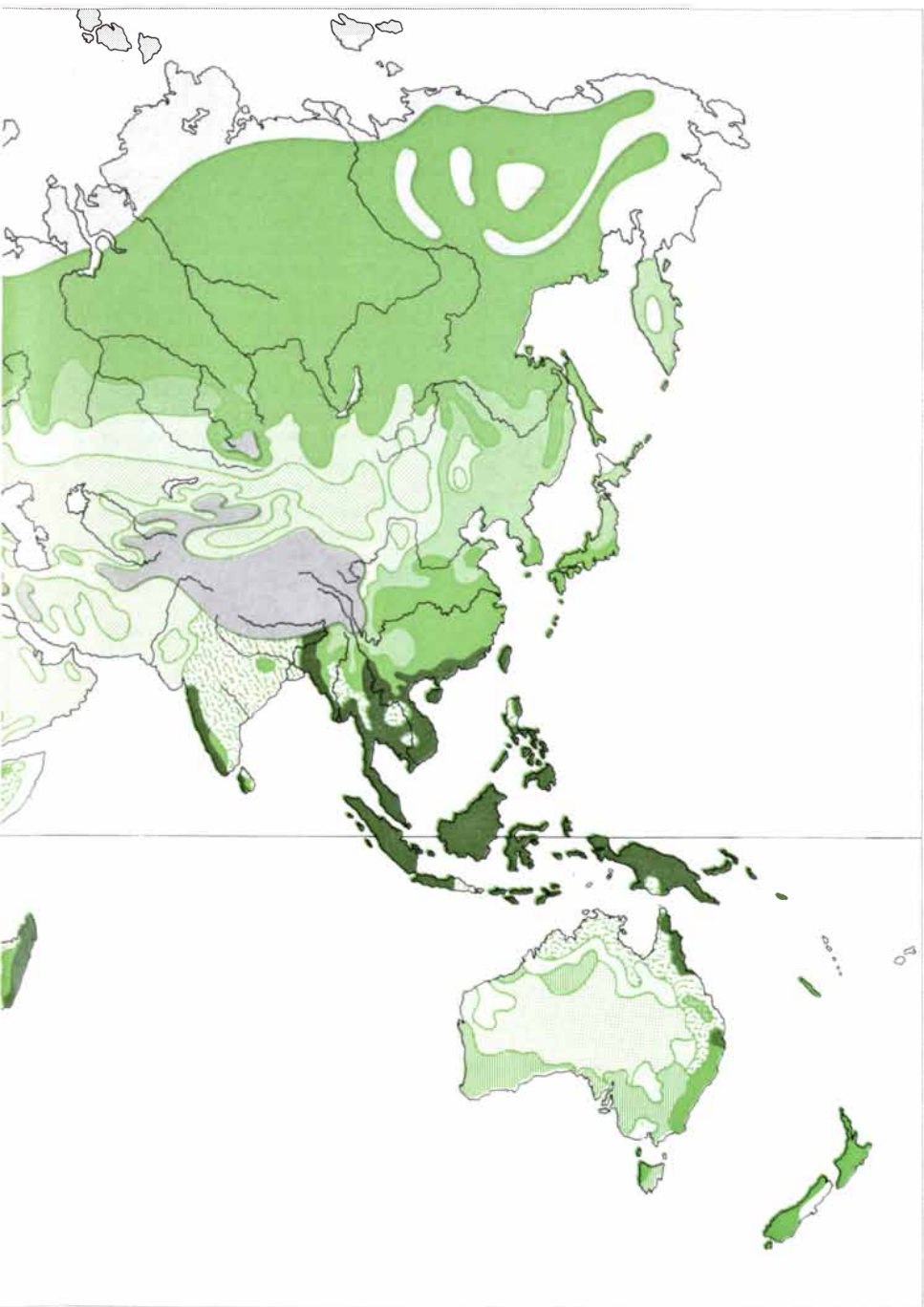



**CARBON STORED IN PLANTS** is distributed as is shown in this world map based on the work of H. Brockmann Jerosch. The total amount of carbon stored in terrestrial biomass is about $830 \times 10^{15}$ grams. In comparison the carbon in the total oceanic biomass is negligible: less than $2 \times 10^{15}$ grams. About 40 percent of all plant carbon is stored in the tropical rain forests. Another 14 percent is held

organic matter, left from photosynthesis after the needs of the plant for respiration have been met. It is the organic matter available for the growth of the plant and is ultimately available for storage or for consumer organisms such as animals or organisms of decay.)

Perhaps the most significant finding of the Whittaker-Likens study is that the tropical rain forests, with their big trees, represent the largest single pool of carbon in the biota and also have the highest total net primary production. This observation emphasizes the importance of the tropical forests in the earth's carbon budget. If these and other forests are harvested rapidly and the stored carbon is released, they have the potential for contributing significantly to the amount of carbon dioxide in the atmosphere. Conversely, if deforested lands are allowed to become forested again, the forests will absorb some of the carbon dioxide from the atmosphere, retarding the rate of increase.

Curiously, the question of whether the biotic pool is getting larger or smaller has only recently become controversial. For example, at a 1970 conference titled "Study of Critical Environmental Problems," organized by the Massachusetts Institute of Technology and held in Williamstown, Mass., the conferees assumed that the biotic pool must be getting larger and encouraged oceanographers to believe their models of oceanic circulation and carbon dioxide absorption were adequate. A subsequent series of conferences appeared to reinforce this assumption, although in 1972 doubts were raised at a Brookhaven National Laboratory conference titled "Carbon and the Biosphere."

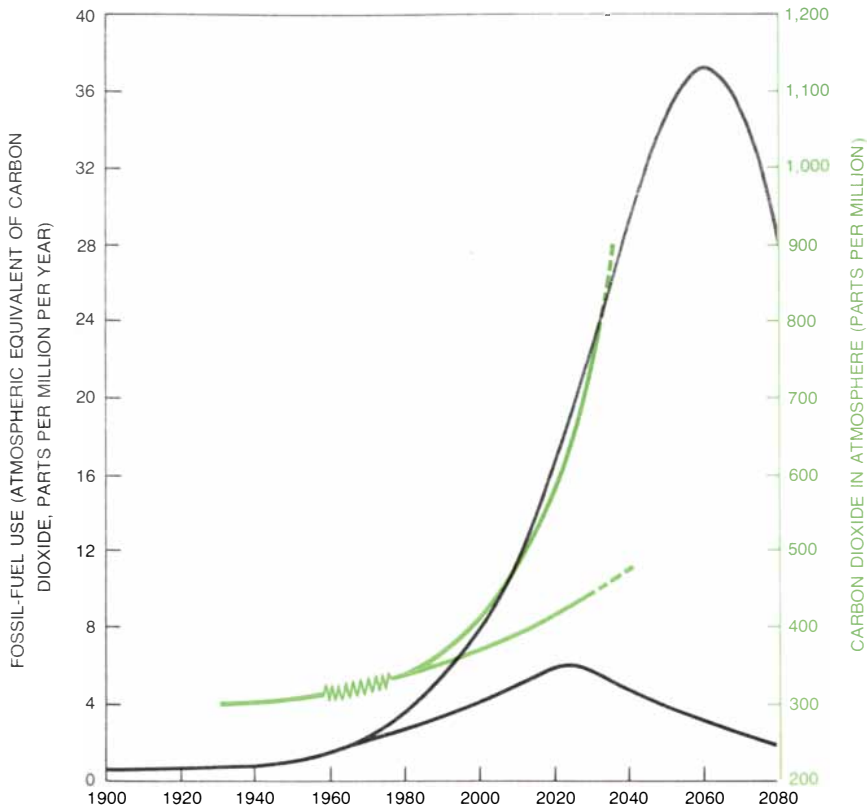


**in tropical seasonal forests. Forests at all latitudes hold nearly 90 percent of all the carbon stored in the world's ecosystems, both terrestrial and marine. Author believes reduction in the area of forests has contributed significantly to increase of carbon dioxide in the atmosphere.**

The first serious challenge to the oceanographers' model was contained in two papers presented at the Dahlem Conference on Biogeochemistry, held in West Berlin in November, 1976. In one paper R. A. Houghton and I estimated that the biota may be releasing about as much carbon dioxide to the atmosphere annually as is released by the combustion of fossil fuels. In the other paper Bolin arrived at a somewhat lower value for the biotic release by drawing on data on forest harvests collected by the United Nations Food and Agriculture Organization (FAO). Bolin's estimate was about $10^{15}$ grams of carbon. The subject dominated the Dahlem conference, stimulating much discussion.

The possibility of a significant release of carbon dioxide from the biota also dominated two subsequent conferences, one arranged by the Department of Energy (formerly the Energy Research and Development Administration) last March in Miami Beach and the other held in April at Ratzeburg in West Germany under the auspices of the Scientific Committee on Problems of the Environment of the International Council of Scientific Unions. Recent papers in *Science* and other journals support the conclusion that the biota is a net source of atmospheric carbon dioxide and not a sink. The oceanographers' models are clearly in question. What does this mean for the world carbon budget?

The answer is far from clear. Substantially larger quantities of carbon are entering the atmosphere than are stored in it. In addition to the $5 \times 10^{15}$ grams of carbon released annually from the combustion of fossil fuels, another $4 \times 10^{15}$ to $8 \times 10^{15}$ grams, possibly more, may be being released currently through the destruction of forests and the accelerated oxidation of humus. Of this combined amount, $9 \times 10^{15}$ to $13 \times 10^{15}$ grams of carbon per year, only $2.3 \times 10^{15}$ grams accumulate in the atmosphere. The remainder, $7 \times 10^{15}$ to $11 \times 10^{15}$ grams, perhaps more, is stored somewhere on the earth. But where? As we have seen, the present models of oceanic uptake provide for the removal of less than $3 \times 10^{15}$ grams of carbon per year. Oceanographers are now reviewing their assumptions to see if they have overlooked mechanisms




**PROJECTIONS OF FOSSIL-FUEL CONSUMPTION (*black curves*) and carbon dioxide content of the atmosphere (*color*) are shown for minimum and maximum plausible rates of increase. The fuel-consumption rates are taken from a recent study done at the Oak Ridge National Laboratory. The minimum projection assumes an annual increase of 2 percent per year until 2025, followed by a symmetrical decrease. The maximum projection assumes a growth rate of 4.3 percent per year until the rate is limited by the depletion of resources in the middle of the next century. The uncertainty of such predictions makes estimates of the future carbon dioxide content of the atmosphere extremely risky. The uncertainty is greatly aggravated by the newly recognized possibility that the destruction of forests may also be releasing large amounts of carbon dioxide to the atmosphere. The sawtooth part of the curve represents the Mauna Loa measurements that were begun in 1958. At present the fossil fuel burned each year releases to the atmosphere an amount of carbon dioxide equivalent to about two parts per million. About half of that amount of carbon dioxide is actually retained in the atmosphere.**

that would be able to sequester additional amounts of carbon.

Because the issue is so important those who are familiar with terrestrial ecosystems are closely examining their own data, particularly the data covering changes in forest mass. How certain can we be that terrestrial communities are indeed a net source and not a sink? The problem has been addressed in several ways in recent months. The data are not as good as one might hope, but they seem compelling to those who are familiar with them.

The analysis is based first on knowledge of the relative magnitudes of the pools of carbon held within the biota and on the net primary production for each of the major plant communities on the earth. For example, the Whittaker-Likens study shows that tropical rain forests hold about 42 percent of all the carbon locked up in terrestrial vegetation and account for about 32 percent of the total net primary production. Forested areas of all kinds—tropical, temperate and boreal—hold 90 percent of all the carbon held in vegetation and contribute more than 60 percent of the net primary production. The only other large single contributor to the net primary production is the savannas, or grasslands, which account for about 12 percent of it and for only about 3 percent of the standing mass of carbon. All the cultivated land on the earth accounts for about 8 percent of the total net primary production and for less than 1 percent of the standing mass of carbon. The Whittaker-Likens estimates lie about midway between the extremes of other analyses that were carried out under the direction of P. Duvigneaud of the University of Brussels at the Ratzeburg conference. Since the various studies all confirm the importance of forests, particularly the tropical forests, it is essential to establish whether or not these ecosystems are changing in size and, if they are, at what rate.

The data are scarce. We have the experience of Henry C. Darby, a British geographer, who in 1954 published an appraisal of the changes in the forest vegetation of Europe spanning the millennium from A.D. 900 to 1900. In that period the forest cover of western Europe was reduced from about 90 percent to about 20 percent. A similar change took place earlier in the lands of the Mediterranean, particularly the Levant. This great reduction in forest area released a quantity of carbon that was a significant fraction of the total previously held in the atmosphere. It is reasonable to assume that since 1900 continued industrialization and population growth have resulted in similar changes in the forests elsewhere.

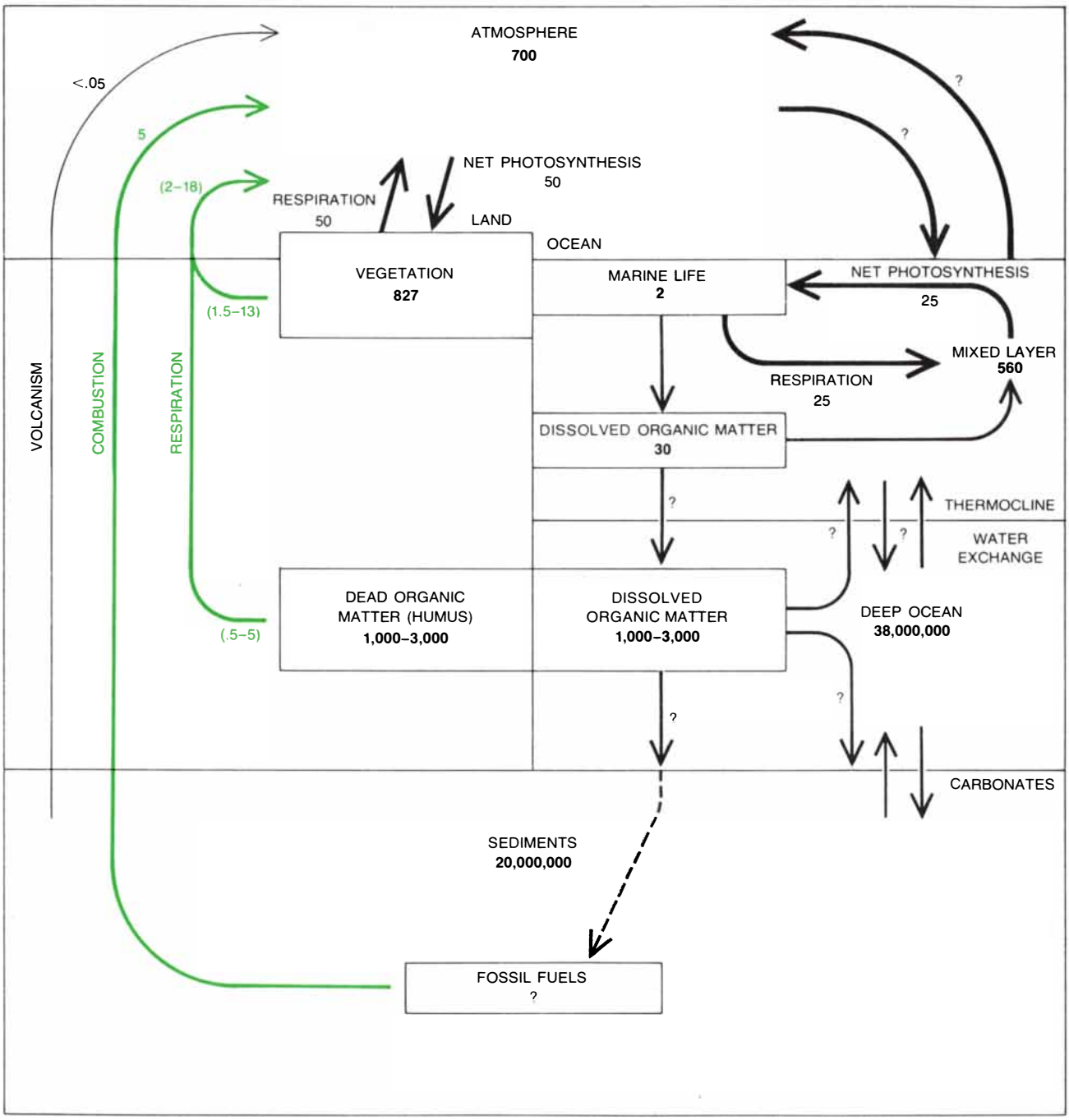
Timothy Wood and Daniel B. Botkin of the Ecosystems Center at Woods Hole recently conducted a study of the changes in the forest area of New England over the period since the arrival of European settlers. They found that up to 1900 there was a continuous reduction in the total standing crop of forests but that since then there has been a period of recovery owing to the abandonment of agriculture and the expansion of forests into the former agricultural land. The recovery, however, has not resulted in a pool of carbon equivalent to that of the original forests. The forests have been harvested regularly and have never reached the stature or extent of the original ones; the standing crop of carbon today is no more than half the original one. The newest data indicate that the increase in the forested area has now ended, probably because of a renewed expansion of agriculture and an intensified harvest of trees.

The Wood-Botkin study shows that in a Temperate Zone forest that was allowed to recover from intensive harvest the storage of carbon amounted to 3 or 4 percent per year of the net primary production over the entire period of recovery, about 70 years. If a similar fraction of the net primary production of other Temperate Zone forests were stored and an equal additional quantity were stored in humus, the amount of carbon accumulated in all the Temperate Zone forests of the earth would come to about $.5 \times 10^{15}$ grams per year. The experience with New England forests suggests the regrowth of forests in the temperate zones probably does not at present provide a large sink for atmospheric carbon dioxide.

Meanwhile there has been a continuous expansion of agriculture into other forested lands, a continuous harvesting of primary forests elsewhere and a general toxification of the earth as a result of human activities. The most conspicuous inroads into the forests have been in the Tropics. The largest remaining forested area on the earth is the Amazon Basin, and we have sought data on the rate of harvest of Amazonian forests. There is no thoroughly satisfactory survey that can be applied to the entire Am-

azon Basin without all the hazards associated with any extrapolation. Nevertheless, Lawrence S. Hamilton of Cornell University and his colleagues have recently published a series of reports on changes in Venezuelan forests.

One of the reports, by J. P. Veillon, shows a 33 percent reduction in forest area in the western llanos of Venezuela between 1950 and 1975. Hamilton, in an introduction to the series, cites FAO data that "suggests an estimate of moist forest cleared per year at . . . 0.6 to 1.5 percent of the still existing area." Studies from other sections of the Amazon Basin attest to the high rate of development of highways, the expansion of agriculture at the expense of forests and the failure of revegetation after the forest has been cleared. Virtually no knowledgeable investigator with experi-




**GLOBAL BALANCE SHEET shows major carbon repositories and annual exchange rates among depots that are in contact. Quantities are expressed in units of $10^{15}$ grams, or billions of metric tons. Annual releases to the atmosphere governed by human activities are shown in color. Land plants fix a net of about $50 \times 10^{15}$ grams of carbon per year. This carbon is either consumed and promptly respired by various terrestrial organisms or stored in the plant mass. The balance between fixation (net photosynthesis) and storage plus the total respiration of all terrestrial organisms determines whether there is a net flux of carbon dioxide to or from the biota. Many biologists now believe that there has been a long-term net flow of carbon dioxide from the biota into the atmosphere and that the flow is continuing. The carbon fixed by marine organisms is either respired or stored. It has been commonly assumed that most of it is respired immediately and recycled. It now seems possible that sinking fecal pellets may carry more carbon into the oceanic depths than had been thought. This transfer would supplement the normally slow diffusion of carbon dioxide into the surface layers of the ocean, where it comes into equilibrium with the carbonate-bicarbonate system. Although the deep ocean provides a virtually unlimited sink for carbon dioxide, gas must enter mixed layer and then penetrate thermocline, a thermally stratified layer that impedes mixing with deeper layers.**

ence in the Tropics believes the Amazonian forest will escape substantial inroads from harvest and clearing over the next 30 years.

The best assumptions available to me suggest that the rate of clearing of tropical forests for all purposes is probably in the range of .5 to 1.5 percent of the existing area annually. If the rate is 1 percent, and if most of the clearing is through the expansion of agriculture and grazing lands into areas that were formerly forested, the release of carbon from forests alone would be about $4.5 \times 10^{15}$ grams per year. By applying such estimates to the entire terrestrial biota, allowing for additional storage through regrowth, my colleagues and I have recently estimated that the net release of carbon dioxide to the atmosphere from the biota is $6 \times 10^{15}$ grams of carbon annually. The magnitude of an additional release through the decay of soil humus is hard to calculate but we have estimated the loss at $2 \times 10^{15}$ grams per year. The total estimated release is therefore about $8 \times 10^{15}$ grams. The uncertainties are so great, however, that we have guessed that the actual loss from the biota might lie between $2 \times 10^{15}$ and $18 \times 10^{15}$ grams per year. Although one must always bear in mind the limitations of the data on which such assumptions are based, it is difficult to avoid the conclusion that the destruction of the forests of the earth is adding carbon dioxide to the atmosphere at a rate comparable to the rate of release from the combustion of fossil fuels, and if the oxidation of humus is included, at an appreciably higher rate.

| | AREA ($10^6$ SQUARE KILOMETERS) | NET PRIMARY PRODUCTION ($10^{15}$ GRAMS OF CARBON PER YEAR) | PLANT MASS ($10^{15}$ GRAMS OF CARBON) |
|---|---|---|---|
| TROPICAL RAIN FOREST | 17.0 | 16.8 | 344.0 |
| TROPICAL SEASONAL FOREST | 7.5 | 5.4 | 117.0 |
| TEMPERATE EVERGREEN FOREST | 5.0 | 2.9 | 79.0 |
| TEMPERATE DECIDUOUS FOREST | 7.0 | 3.8 | 95.0 |
| BOREAL FOREST | 12.0 | 4.3 | 108.0 |
| WOODLAND AND SHRUBLAND | 8.5 | 2.7 | 22.0 |
| SAVANNA | 15.0 | 6.1 | 27.0 |
| TEMPERATE GRASSLAND | 9.0 | 2.4 | 6.3 |
| TUNDRA AND ALPINE MEADOW | 8.0 | 0.5 | 2.3 |
| DESERT SCRUB | 18.0 | 0.7 | 5.9 |
| ROCK, ICE AND SAND | 24.0 | 0.03 | 0.2 |
| CULTIVATED LAND | 14.0 | 4.1 | 6.3 |
| SWAMP AND MARSH | 2.0 | 2.7 | 13.5 |
| LAKE AND STREAM | 2.0 | 0.4 | 0.02 |
| TOTAL CONTINENTAL | 149.0 | 52.8 | 826.5 |
| OPEN OCEAN | 332.0 | 18.7 | 0.45 |
| UPWELLING ZONES | 0.4 | 0.1 | 0.004 |
| CONTINENTAL SHELF | 26.6 | 4.3 | 0.12 |
| ALGAL BED AND REEF | 0.6 | 0.7 | 0.54 |
| ESTUARIES | 1.4 | 1.0 | 0.63 |
| TOTAL MARINE | 361.0 | 24.8 | 1.74 |
| WORLD TOTAL | 510.0 | 77.6 | 828.0 |

**MAJOR PLANT COMMUNITIES of the earth are listed along with their area, their net primary production and the amount of carbon they hold in storage. Net primary production is the amount of carbon a plant community provides annually for harvesting or for the support of various consumer organisms, either wild or domesticated. Although only about 30 percent of the earth's surface is covered by land, the net primary production of terrestrial vegetation is slightly more than twice the primary production of the oceans. The quantity of carbon stored in land plants is some 500 times greater than the quantity stored in marine ecosystems. The carbon stored in trees is roughly equal to the carbon in the atmosphere. The figures in the table were recently compiled by R. H. Whittaker and Gene E. Likens of Cornell University.**

These observations have come as a surprise to many agriculturists and foresters. The assumption in their experience has been that modern agricultural and forestry practices have improved the net primary productivity of terrestrial systems under human management and that the improvement of net primary production must have resulted in a faster storage of carbon in the humus of soils under management than in natural soils. But the average productivity of agriculture, measured in the ecologist's units of total organic matter available to man or other animals, is substantially less than the average primary productivity of the natural communities the agriculture replaced. Agricultural plant communities are managed not for the storage of large quantities of carbon but for the rapid turnover of carbon pools through their utilization by man. This means that lands under agriculture do not store as much carbon as the forests that were replaced by agriculture did. Similarly, grasslands turned to agriculture tend both to lose the organic matter in the soil through decay and not to accumulate additional organic matter.

Forests that are managed intensively, although they may produce increased amounts of wood, tend to have at the time of harvest a standing crop much smaller than the stands of the primary forests they replaced. This again means that the turnover time has been reduced and that the standing pool of carbon never achieves the same size as the pool of the original forests. Thus the replacement of primary forests with secondary forests that are managed for lumber or pulp results in a net release of carbon dioxide. There is a further release of carbon through the decay of humus. The decay is stimulated by the harvest of the forests and may continue for several years after the harvest during the early stages of the plant succession that follows the cutting. All these factors lead to a net release of carbon when unmanaged primary forests are replaced by managed ones.

One further line of evidence seems to support the conclusion that the terrestrial biota has for many decades been a net source of carbon dioxide rather than a sink. By comparing the ratios of carbon isotopes in trees Minze Stuiver of the University of Washington has calculated that in the century between 1850 and 1950 the biota appears to have released

to the atmosphere $1.2 \times 10^{15}$ grams of carbon per year. Over the same period the release from fossil fuels averaged $.6 \times 10^{15}$ grams of carbon per year.

Stuiver exploited the fact that the ratio of carbon 12 to carbon 13 varies among the atmosphere, the biota and fossil fuels. The biota and fossil fuels are enriched slightly in the lighter isotope, carbon 12. In addition a third isotope of carbon, carbon 14, is present only in the atmosphere and in the biota. Carbon 14 is produced in the upper atmosphere through the cosmic-ray bombardment of the common isotope of nitrogen, nitrogen 14. It has also been produced in large quantities by atomic bomb tests in the atmosphere. Because carbon 14 has a half-life of some 6,000 years it has long since disappeared from fossil fuels, which were formed millions of years ago. The combustion of fossil fuels therefore releases a pool of carbon that is deficient in carbon 14 and tends to dilute the carbon 14 of the atmosphere. By measuring the concentration of various carbon isotopes in the rings of trees of known age and comparing the ratios with those of the atmosphere and of fossil fuels Stuiver was able to estimate the amount of carbon released from the biota. The measurements are technically difficult; moreover, the ratios of carbon isotopes in trees are affected by various environmental factors that make the results less clear-cut than one might like. Nevertheless, the technique offers an important additional means for testing the magnitude of the release of carbon dioxide from the biota.

| FORESTS: | PLANT MASS ($10^3$ GRAMS OF CARBON PER SQUARE METER) | NET PRIMARY PRODUCTION (GRAMS OF CARBON PER SQUARE METER PER YEAR) | |
|---|---|---|---|
| | | RANGE | MEAN |
| TROPICAL (WET) | 3.0–36.0 | 450–1,600 | 990 |
| TEMPERATE | 3.0–90.0 | 270–1,125 | 560 |
| BOREAL | 3.0–18.0 | 180–900 | 360 |
| SAVANNAS | 0.1–7.0 | 90–900 | 400 |
| GRASSLANDS | 0.1–2.3 | 90–675 | 270 |
| CULTIVATED LAND | 0.2–5.4 | 45–2,800 | 290 |

**TRANSFER OF LAND INTO AGRICULTURE usually results in a sharp reduction in the carbon stored in the biomass and a somewhat smaller reduction in the carbon dioxide removed annually from the atmosphere and fixed by photosynthesis. The table, also based on the work of Whittaker and Likens, indicates the range of published estimates for the standing biomass and the net primary production of major natural ecosystems compared with cultivated land.**

In view of the available evidence there seems to be little question that the increase in carbon dioxide content of the atmosphere is due not only to the combustion of fossil fuels but also to the destruction of forests. G. Evelyn Hutchinson of Yale University made this point in a chapter in *The Earth as a Planet,* a book edited by Gerard P. Kuiper in 1954. Hutchinson guessed that the release of carbon from the biota was about equal to the release from fossil fuel. The best evidence we have at present indicates that this relation persists.

There is enough carbon held within the biota for the relation to continue for another decade or two through the expected peak in the worldwide consumption of oil. By that time, if not sooner, the earth will have been committed to climatic change as a result of the accumulation of carbon dioxide in the atmosphere—if indeed such a change is to take place. The question of whether or not it will take place depends on the scale of the carbon dioxide effect, an appraisal that is extremely difficult to make with any certainty. It is now recognized that changes in the output of the sun, changes in the reflectivity of the earth (as it is influenced by the extent of cloud, snow and ice cover) and other factors all influence the climate. Whether or not the carbon dioxide content of the atmosphere is great enough to be the dominant factor remains to be seen. If the carbon dioxide effect is indeed dominant, the probability is that the earth will be warmed differentially, with temperatures increasing toward the poles. Such a change can be expected to move the desert zones poleward, enlarging the areas of aridity and reducing the areas suitable for agriculture. The prospect is not encouraging for a world whose human population may double within the next 30 to 35 years.

If the evidence were overwhelming that the risk of a deleterious change in climate over the next few decades was unacceptably great, the course of action would be clear enough. The burning of fossil fuels would be restricted to reduce that source of carbon dioxide. Strong moves would also be made to prevent the harvest of primary forests around the world, to expand the areas devoted to forests and to allow such areas to develop massive standing crops of trees. Whether such drastic measures could be effected is much in doubt; the social problems that would result would clearly be profound.

Other suggestions have been made, including the suggestion that because the availability of phosphorus is considered by some to limit net primary production in the oceans the advanced nations should apply some of their industrial energy to the mining of phosphorus with the objective of transferring it as rapidly as possible to the relatively infertile oceans, thereby stimulating photosynthesis and the storage of carbon. The scheme is superficially attractive because it seems to offer a way of speeding the storage of carbon in the oceans. The possibility exists, however, that any additional photosynthesis stimulated in the ocean would be offset by an equivalent stimulation of respiration, so that there would be no net increase in carbon storage. There are many questions about the feasibility of such a step, including the basic one of whether phosphorus is indeed critically limiting in oceanic waters. There is a reasonable possibility that nitrogen is also limiting, in which case the stimulation of marine photosynthesis could prove to be substantially more difficult than has been assumed.

Nevertheless, recent experience has emphasized to all of us who work in these fields that important details of the world carbon budget are substantially unknown at present. Some improvement in our knowledge of such details seems feasible through measurement of the changes in the areas and structures of forests worldwide from satellite photographs. Investigations are also under way to see if biotic mechanisms might facilitate the transfer of considerably more carbon into the oceanic depths than we assume is being transferred today. The immediate prospect of establishing with high precision all the details of the world's carbon budget, however, is not bright.

The potential hazards associated with a steady increase in the carbon dioxide content of the atmosphere will loom large in the coming decades and will doubtless bear heavily on such decisions as whether to accelerate the development of power plants based on nuclear fuel instead of those based on coal and whether to preserve forest areas instead of encroaching on them (and, if the forests are to be preserved, how to provide the new lands that are almost certain to be needed for agriculture). There is almost no aspect of national and international policy that can remain unaffected by the prospect of global climatic change. Carbon dioxide, until now an apparently innocuous trace gas in the atmosphere, may be moving rapidly toward a central role as a major threat to the present world order.

# The Surgical Replacement of the Human Knee Joint

*Improvements in design, materials and surgical methods have advanced the artificial knee to the point where it can provide substantial benefits to a person suffering from severe arthritis*

by David A. Sonstegard, Larry S. Matthews and Herbert Kaufer

In normal activities such as walking, running, kneeling, climbing stairs and getting in and out of chairs the load put on the human knee joint can exceed five times the weight of the body. Nowadays large numbers of people make further (often outrageous) demands on the knee in participating in such sports as football, soccer, tennis and long-distance running. It is no wonder that many people go into their later years with one knee or perhaps both knees so badly deteriorated as to be crippling.

It is now possible to replace such a knee with a mechanical device that imitates the subtle and complex motions of which a sound knee is capable. At present the prosthesis is given only to people who are severely crippled and unlikely to put large demands on the knee. The artificial knee is not yet an answer to the problems of people who are youthfully active, primarily because the loads they put on the knee cannot be withstood by the device under currently available techniques for attaching it to bones. In this article we shall describe several artificial knees, emphasizing particularly the Spherocentric knee that has been developed by our group at the University of Michigan.

The work on artificial knees has grown out of the successful development over the past 15 years of artificial hip joints. Much of the credit for the introduction of materials and configurations that would serve adequately as a hip joint and for the surgical techniques whereby the device is put in place belongs to the British surgeon John Charnley of the Centre for Hip Surgery at Wrightington Hospital in Lancashire.

The natural hip joint is a ball-and-socket mechanism that allows motion in all planes through its center of rotation. (The artificial hip is similarly designed and allows a similar range of motion.) Moreover, the hip joint is surrounded and protected by a large mass of muscle and connective tissue. The knee, in contrast, has a more exacting range of motion and is relatively unprotected by surrounding tissue. It is highly vulnerable to injury from blows as well as from sudden stops and turns. For these reasons it has not been easy to translate the success with artificial hips into the same with artificial knees. Even artificial hips, however, are not suitable for use in a severely injured joint except in highly selective circumstances. A fracture or a severe anatomical disruption prevents the appropriate alignment and skeletal support of a prosthesis. Therefore the people who have received artificial hips and knees have been chronic, disabled arthritis patients.

Arthritis literally means "joint inflammation." The term is applied to nearly a hundred diseases, all of which have as symptoms persistent or recurring pain, stiffness, tenderness and swelling in one joint or more. The number of people thus afflicted has been estimated by the Arthritis Foundation as 363 million, or 10 percent of the world's population. Among them are more than 50 million people in the U.S., of whom some 20 million require medical care and 3.5 million are significantly disabled.

The cause of arthritis is not known, and at present there is no cure for most types of the disease. The major forms are osteoarthritis and rheumatoid arthritis. Osteoarthritis is a degenerative, wear-and-tear type of arthritis that progresses with use of the joint and with age. Rheumatoid arthritis, an inflammatory disease of the soft tissues, is the most disruptive and disabling type of the disease. For both types of arthritis the treatment includes drugs, physical therapy, rest, bracing and the application of heat. Notwithstanding such efforts, both osteoarthritis and rheumatoid arthritis can progress to cause severe pain, instability of the joint, deformity, limitation of activity and disability. Then it is likely that surgery will be necessary. Most replacements of hip and knee joints are made under these circumstances.

One can best grasp what is involved in the replacement of a knee joint by considering the anatomy of a normal knee [*see illustration on page 47*]. The joint includes two articulations, or movable parts, that transmit force and relative motion. One is the patellofemoral bearing surface between the patella (the kneecap) and the femur (the thighbone). The other is the tibiofemoral articulation between the tibia (the larger of the two long bones between the knee and the ankle) and the femur.

Three basic groups of muscles, under the control of the nervous system, provide the forces required for the functioning of the knee. The hamstring group and the gastrocnemius group contract to make the leg bend at the knee. The quadriceps group, utilizing the mechanical advantage provided by the patella, straightens the leg. These two movements are known respectively as flexion and extension.

The ligaments of the knee, composed mainly of parallel fibers of collagen, are pliant and flexible, so that they allow a considerable but nonetheless remarkably controlled freedom of movement. Along their length the ligaments form strong and relatively inextensible connections of bone to bone, thereby providing both stability to the joint and constraint of motion. The patellar ligament links the patella and the tibia and provides a sliding motion at the patellofemoral articulation during flexion and extension. Two pairs of ligaments, the collaterals and the cruciates, stabilize the tibiofemoral articulation.

The tibiofemoral articulation includes two crescent-shaped structures of fibrous cartilage: the medial meniscus (on the inside of the knee) and the lateral meniscus (on the outside). The upper surfaces of the menisci are concave,

thus deepening the tibial plateaus for articulation with the femoral condyles, the knoblike structures at the bottom of the femur. Attachments to the relatively flat tibial surface allow motion of the menisci with respect to the tibia. (The common sports injury described as a "torn cartilage" involves one or both menisci.)

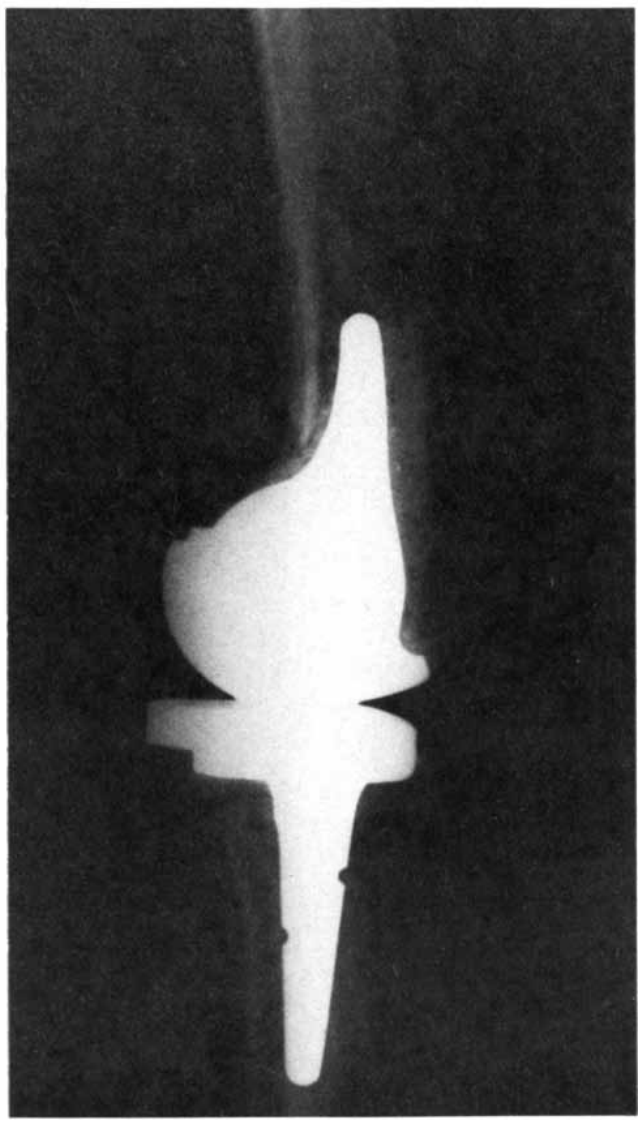
The medial and lateral femoral condyles are complex surfaces with continuously changing radii of curvature. These articular surfaces, in concert with the ligaments and the geometric form of the menisci and the tibia, provide the stabilized and supple range of motion of the knee.

The character of the motion is distinctive. For example, in flexion and extension the knee acts not as a simple hinge but rather as an infinite succession of uniquely placed centers of rotation, each center acting at a particular relative orientation of the femur and the tibia. The tibiofemoral articulation hence includes a subtly coordinated combination of rolling, gliding and rotational movements. At full extension an outward rotation of the tibia with respect to the femur provides a locking mechanism that enables one to stand comfortably.

The load-bearing surfaces of the patella, the tibia and the femur are composed of a relatively thin layer of cartilage, which is a compliant tissue. The cartilage is lubricated by the synovial fluid, which is confined to the space in the joint by a capsule. Although much remains to be learned about the natural lubrication of joints, it is known that the combination of synovial fluid and cartilage provides a smooth, durable, low-friction mechanism the best man-made joints cannot equal.

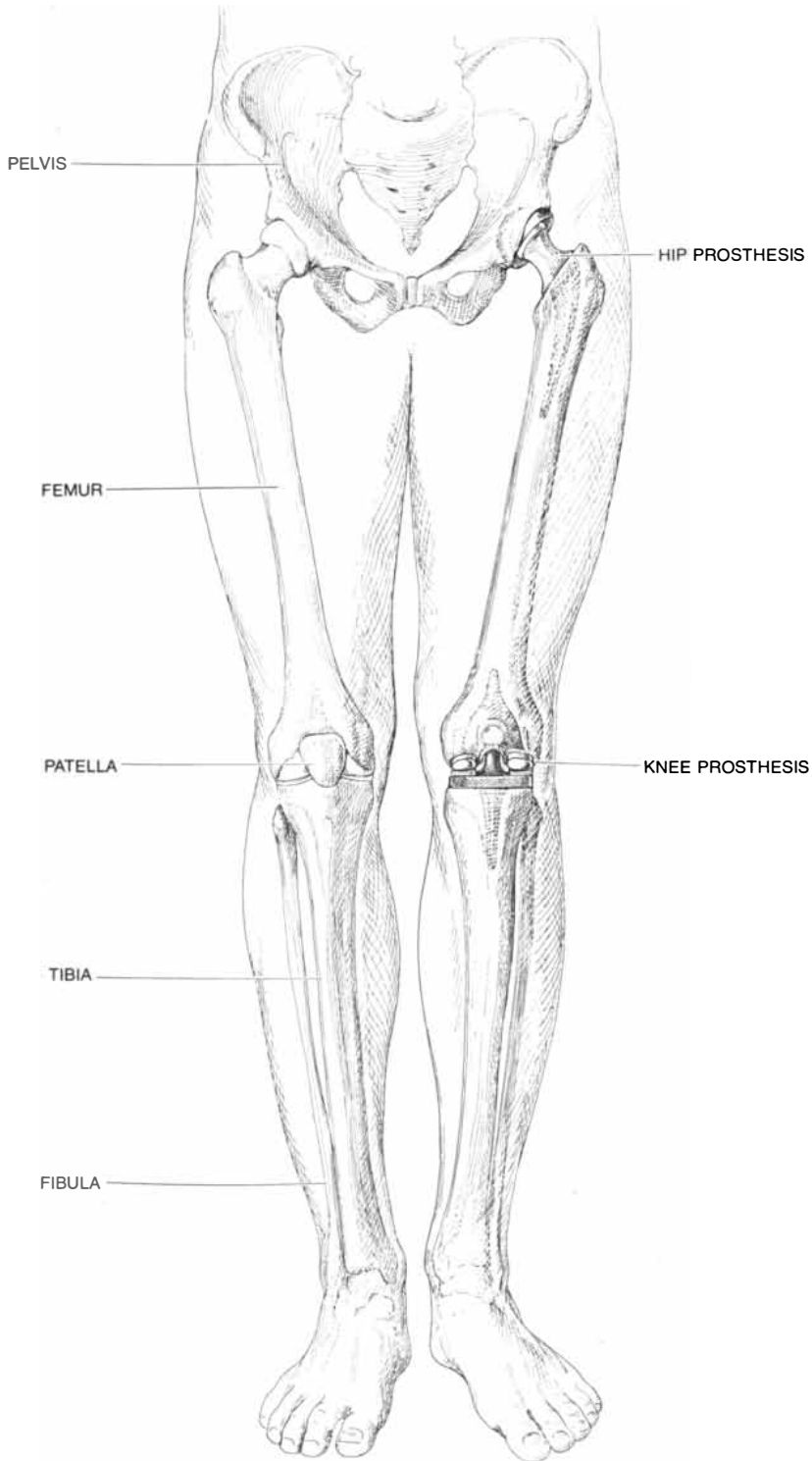
The use of metallic implants for the repair or support of the human skeleton is perhaps as old as surgery. The body, however, puts severe demands on metal. Not until this century were the requirements of biocompatibility, sterility, strength and resistance to corrosion adequately defined. Only a small number of metals have proved to be satisfactory. In the U.S. most artificial joints are made of either a stainless steel or an alloy of cobalt, chromium and molybdenum. Titanium is being used increasingly.

By the early 1950's the work on the suitability of materials had progressed to the point where the first notably successful knee replacement was introduced by Börje Walldius of the Karolinska Institute in Stockholm. His prosthesis had a fixed-hinge design and was made first of an acrylic material and lat-



**ARTIFICIAL KNEE that was implanted in the left leg of a patient with severe arthritis appears in these X rays. The prosthesis is the Spherocentric device developed by the authors at the University of Michigan and portrayed on the cover of this issue. Among the distinctive features of the Spherocentric device is the fact that it approximates the triaxial motions of which a normal knee is capable and that it provides a gradual deceleration of motion rather than abrupt stops. More than 120 of the knees have been installed at the university.**

er of an alloy of cobalt, chromium and molybdenum. The device is fixed to the bone by means of careful sculpturing and fitting of the components to the ends of the tibia and the femur and by stems that extend into the hollow intramedullary spaces of these long bones. The Walldius knee has provided relief from pain and return of function for many arthritic patients. When it has failed, the usual reasons have been that it has worked loose from the bone or that tissues have become inflamed by debris of wear from the load-bearing surfaces (metal on metal). The basic fixed-hinge design, with modifications that reflect technological advances in bearing-surface materials and in attachment of the device, is still chosen by some surgeons to replace a severely disabled knee.




**MAIN SKELETAL STRUCTURES involved in the functioning of the human leg are identified. Prostheses are shown in the left hip and left knee of the figure. The development of artificial knee joints has come about in part from the successful development of artificial hip joints.**

Two important technological improvements have resulted from Charnley's work on artificial hip joints. One improvement is his development of bearing surfaces combining polished metal and polyethylene of high molecular weight. This combination provides a bearing surface that has a low coefficient of friction, a low rate of wear and a small amount of debris from wear (moreover, a debris that is biologically inert).

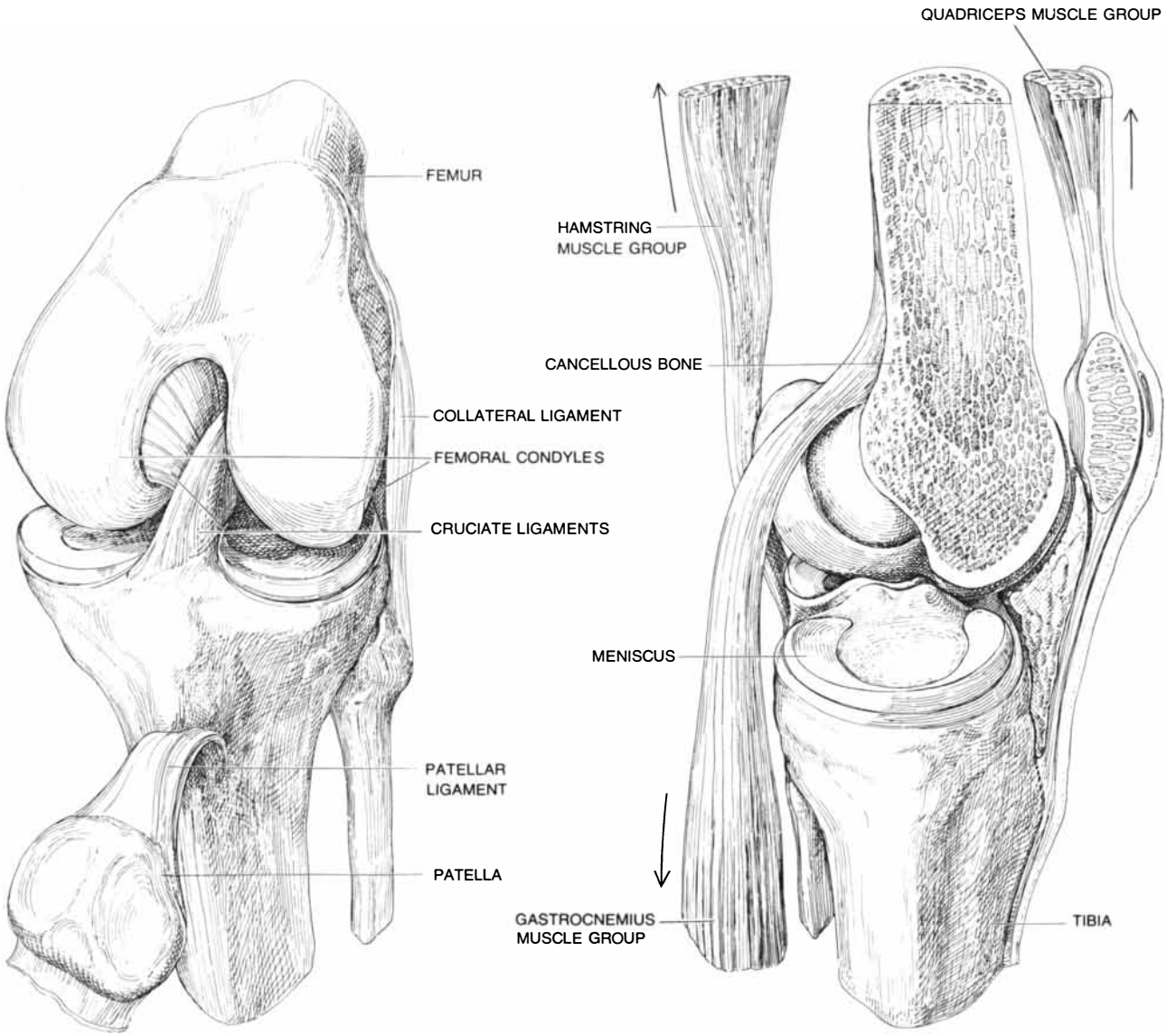
Charnley's second improvement is the use of a cement, polymethyl methacrylate, for fixing the components of a prosthesis to bone. Polymethyl methacrylate is not a glue or an adhesive that works chemically but a true cement that works mechanically. It is applied in a doughlike state so that it can flow around bumps and into holes in the bone; when it hardens, it forms a mechanically interlocked attachment.

These improvements are responsible for the success of the Charnley total hip. Its two components are a spherical hip socket made of polyethylene and a mating femoral component made of metal. Each component is fixed to the skeleton with polymethyl methacrylate.

The first knee-replacement design that utilized these advances was introduced in 1971 by Frank H. Gunston, an orthopedic surgeon associated with the Winnipeg General Hospital and the University of Manitoba. The Gunston knee is a track-and-runner device: the runners, made of metal, are cemented to the femur, and the tracks, made of polyethylene, are cemented to the tibia. The tracks are straight from front to back and slightly concave from side to side; the motion of the runners in the tracks is relatively unconstrained.

In recent years much work has been done to define the ideal artificial knee. The questions at issue include the amount of relative motion to be provided between femoral and tibial components, the materials to use on load-bearing surfaces, the structural integrity of plastic components, the resurfacing of the patellofemoral joint, the amount of bone to be removed and the criteria for choosing patients. More than 80 designs for artificial knees are estimated to be available in the U.S. and abroad. Many of them differ only slightly in configuration and function, so that one must entertain misgivings about this proliferation. The situation does, however, reflect the intense effort devoted to the design of knee replacements.

We shall focus on two general types of knee replacement. The first, which rep-




**NORMAL LEFT KNEE is depicted in views from the front and side. In the front view the knee is bent somewhat, exposing the interior of the knee and the arrangement of the ligaments. The side view portrays the main muscle groups that control the knee. Bending is done by the hamstring and the gastrocnemius groups, straightening by the quadriceps group. Part of the femur has been cut away in side view.**

resents the largest group of artificial knees, is the nonarticulated or unlinked device. Here the emphasis is on replacing the worn and diseased load-bearing surfaces of the knee. Such devices do not include a mechanical link, hinge or similar articulation between the femoral and tibial components. Instead the stability of the joint is provided by the patient's ligaments and muscles, sometimes with help from geometric features of the prosthesis. The Gunston knee is an example.

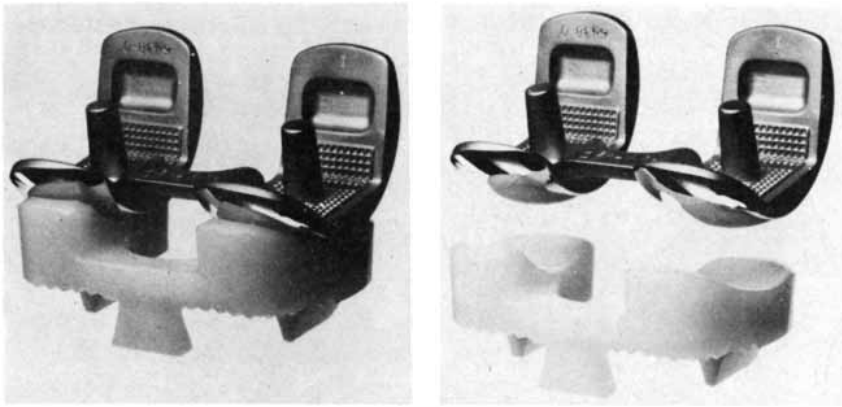
The second type is the articulated device. It provides a constraining mechanical linkage between the femoral and tibial components. A classic example is the Walldius knee.

One of the most commonly implanted nonarticulated devices in the U.S. is the Geometric knee, which grew out of work by orthopedic investigators in several institutions. It is conceptually similar to the Gunston knee but has two components instead of four. The surgeon thus confronts fewer problems of installation and alignment. The bearing surfaces are more congruous, thereby providing larger load-bearing areas and some constraint of motion.

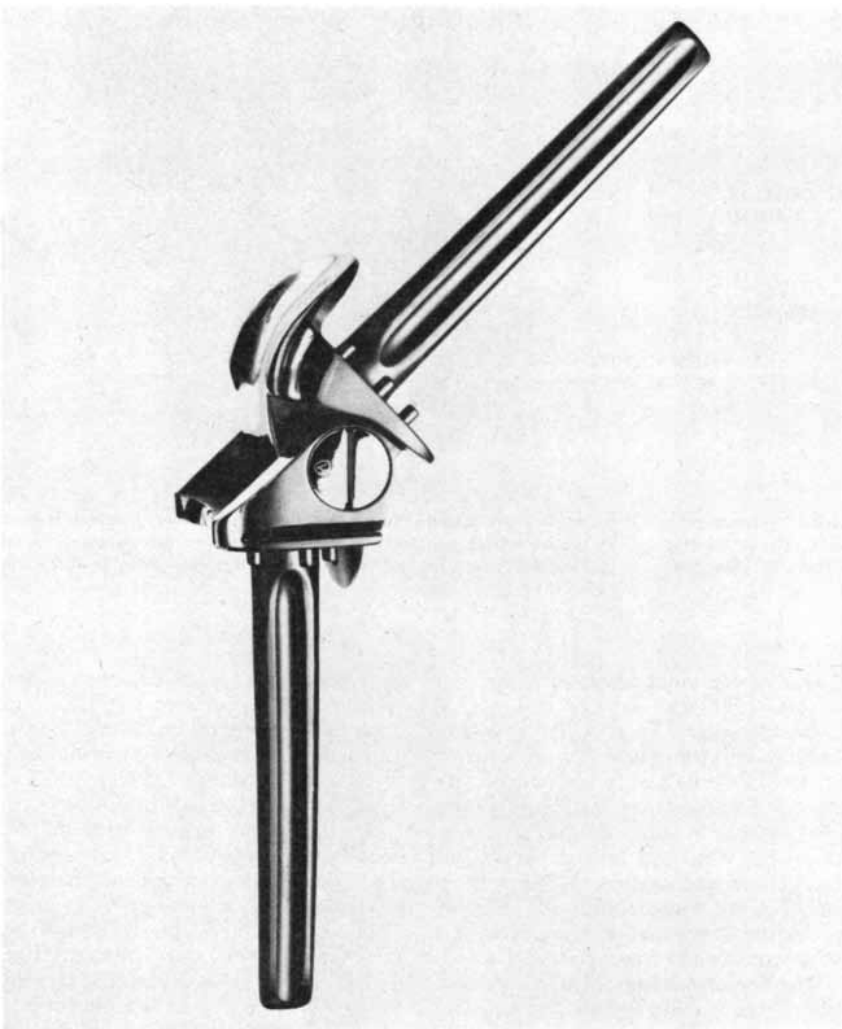
The femoral component consists of two metal bearing surfaces attached to the condyles. The surfaces are made of metal (an alloy of cobalt, chromium and molybdenum) and are in the form of truncated spheres. The tibial component, made of polyethylene, has mating spherical concavities of slightly larger diameter. The device hence allows up to 130 degrees of flexion (the normal knee is capable of about 135 degrees) and small lateral and axial rotations.

Both components have a bridge connecting their symmetrical halves across the front. Each component therefore has the shape of a *U* as seen from above. This feature enables the surgeon to retain the cruciate and collateral ligaments. (With all replacement devices the anatomy of the knee makes it possible to install a prosthesis without removing muscles.)

To install a Geometric knee the surgeon first removes a thin layer of the



**GEOMETRIC KNEE is representative of a class of artificial knees that are nonarticulated, meaning that they do not have a mechanical link between the femoral and tibial components. The prosthesis is shown assembled (*left*) and disassembled (*right*). Such a device depends for its stability on the patient's ligaments and muscles. In effect a nonarticulated knee provides the patient with resurfaced load-bearing areas in the knee. The femoral component of the Geometric knee has two bearing surfaces made of metal; the mating concave parts of the tibial component are made of polyethylene. The device can provide as much as 130 degrees of flexion.**



**WALLDIUS DEVICE was one of the first articulated artificial knee joints. It was introduced in the early 1950's by Börje Walldius of the Karolinska Institute in Stockholm. Having a fixed-hinge design, the Walldius device provides a somewhat smaller range of motion than a normal knee. In its present form the knee is made of an alloy of cobalt, chromium and molybdenum.**

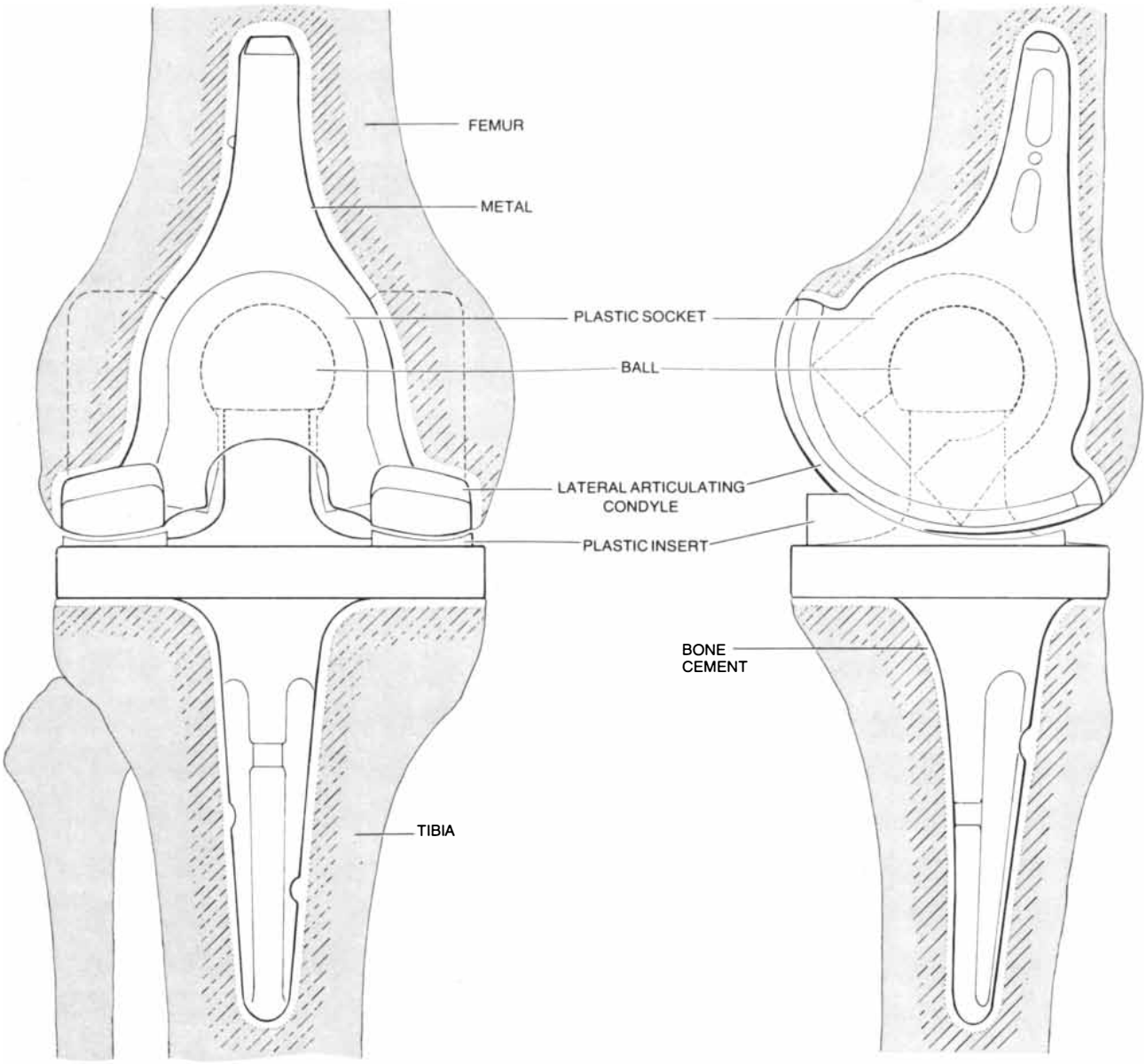
bearing surfaces of the diseased knee. The ends of the bones are then sculptured to achieve the right fit and alignment of the prosthetic components, which are attached with polymethyl methacrylate. The components have stems, grooves and dovetails designed to enhance the action of the cement.

The patient with a Geometric knee thus has what amounts to resurfaced load-bearing areas in his arthritic joint. In some 10 years of use the Geometric knee has proved that it can provide relief from pain and can enable the recipient to be reasonably active, certainly more so than before the operation. The devices do fail at a low rate, most often because the tibial component works loose or becomes deformed. In such cases it is often possible for another device, such as the Spherocentric knee, to be implanted.

The Spherocentric knee is the result of an effort our group began in 1971 to design an articulated artificial knee. Nonarticulated devices such as the Geometric knee were serving well for people whose problem was mainly with the bearing surface and whose diseased knee had good ligaments. For the severely disabled arthritic knee, with gross deformity and instability, the fixed-hinge device of the Walldius type was the only design being used widely. In view of the advancing state of the art we thought an alternative was appropriate and possible.

We set ourselves several goals and criteria of design. Among them were to achieve an articulated, mechanical interlock between the femoral and tibial components to provide intrinsic stability independent of the ligaments; to provide the triaxial motion characteristic of the normal knee; to limit the range of motion by gradual deceleration rather than by abrupt impactive stops; to put metal against polyethylene at all bearing surfaces; to support all the polyethylene components with metal in order to minimize the tendency of the plastic to become deformed; to employ polymethyl methacrylate for fixation; to ensure that any polyethylene bearing surface could be replaced if it became worn, and to achieve a device the surgeon could install without undue difficulty.

The criterion of triaxial motion meant that a ball-and-socket arrangement would be the best means of achieving a mechanical interlock between components. A track-and-runner scheme, bearing the same relation to the ball and socket that a set of outriggers bears to a boat, would provide the range and limitations of motion. We achieved both arrangements with a femoral socket and two tibial inserts contrived so that wherever polyethylene formed part of a bearing surface it was supported by metal [*see illustration on page 51*]. The aim of




**METHOD OF OPERATION of a Spherocentric artificial knee joint is indicated in views from the front and side. The Spherocentric knee is an articulated device, the articulation being provided by the ball-and-socket arrangement. The bearing surfaces are metal on plastic.**

controlled deceleration at full extension of the knee was ultimately achieved by appropriate placement of the ball and socket with respect to the outriggers, which had a carefully chosen set of changing radii of curvature and hence provided a camlike action.

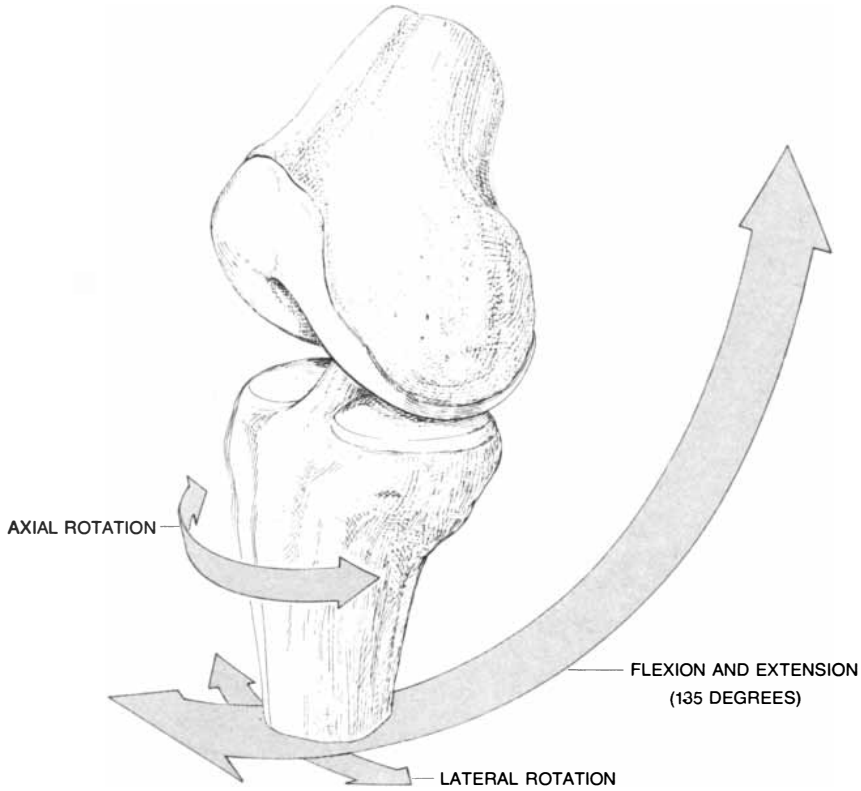
In loading tests with bones from cadavers we found that both the Walldius device and the Spherocentric one sustained higher loads before failing than a natural knee does. Of particular interest were the dynamic loading characteristics of each device in arresting motion at the end of extension. The loading can be measured as energy stored.

We tested several versions of the Spherocentric design in our laboratories with the objective of closely matching the character of the energy storage of the device to that of a normal knee. The final version of the knee provides energy-storage loadings that are about a third of those experienced with the Walldius device, which has metal-to-metal stops. Since the loadings involved in arresting extension are transmitted to the areas where the implant is fixed to the bone by cement, it seems probable that the Spherocentric design will incur fewer loosening problems than the Walldius knee and similarly constrained fixed-hinge devices incur.

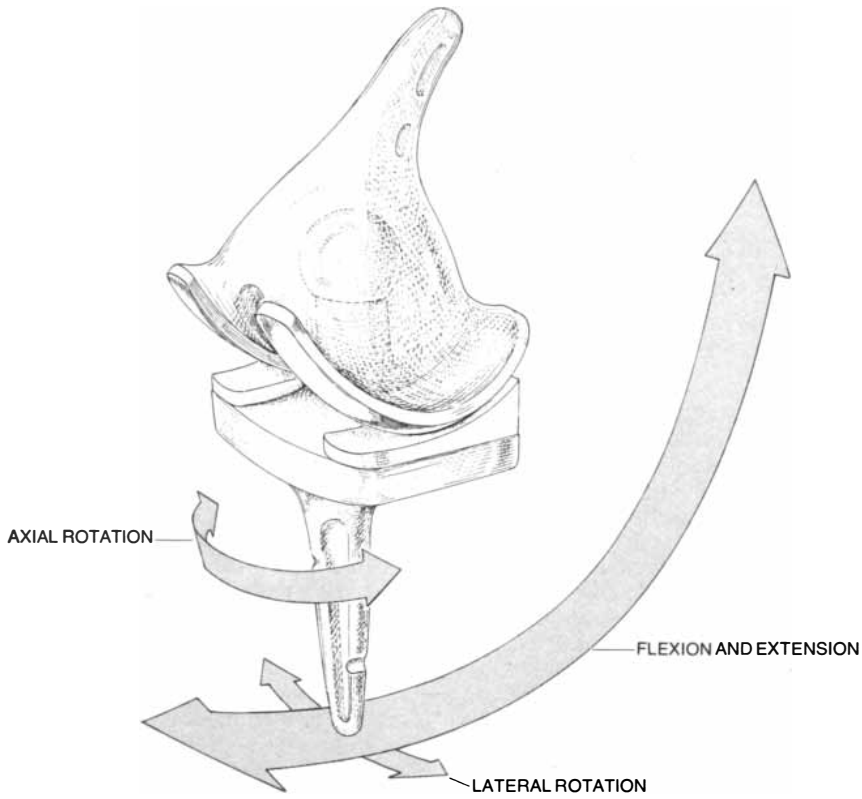
In 1973 we were ready to begin clinical trials of the Spherocentric knee. Our criteria for choosing patients were (and still are) severe deformity, gross instability or failure of a previous implant. The initial trial involved 22 patients and 25 implants. We have followed these patients since they received their artificial knee and have found that they experienced a marked relief of pain and have been able to be much more active.

Up to now more than 120 Spherocentric knees have been installed at the University of Michigan Medical Center. Nevertheless, the device must still be classified as experimental. The features of triaxial motion and controlled deceleration do appear to have reduced the problem of loosening of components, but only further experience with the devices will show whether loosening can be avoided over long periods.

A recent survey under the Rehabilitation Engineering Program of North-




**TRIAXIAL MOTIONS of which a normal knee is capable are portrayed somewhat schematically. In addition to bending through 135 degrees the knee can rotate axially, as it does when one swivels one's foot from side to side at the ankle, and laterally, which means that the tibia rotates with respect to the femur. Such a rotation occurs as the leg is bent and straightened.**




**RANGE OF MOVEMENT allowed by the Spherocentric knee is depicted. Ideally a patient with such a device can achieve 120 degrees of flexion and extension (around the *Y* axis), 30 degrees of axial rotation (the *Z* axis) and five degrees of lateral rotation (the *X* axis). Motion is controlled mainly by the tracks and runners, which have a set of changing radii of curvature.**

western University indicated that 80,000 total hip replacements and 30,000 knee replacements were performed in the U.S. in 1976. Surgeons responding to a questionnaire that was part of the survey indicated that if more reliable prostheses were available, they would do 1.12 times as many hip replacements and 1.76 times as many knee replacements. The figures suggest that considerable effort will be required before the replacement of the human knee can be considered as being more than a qualified success.

In our view the effort must be made in two areas. The first has to do with the design of artificial knees. Promising new implant materials, including metals, polymers and ceramics, are appearing. They should enhance mechanical performance and provide more latitude in design. Continuing research on the loading of both normal and diseased knees will lead to better tests of new designs in the laboratory.

The second area where improvement is needed is the attachment of implants to bones. Loosening is the primary cause of failure of artificial knees (roughly 10 percent in two years). At present not enough is known about the phenomenon to be able to predict which patients will be most likely to experience it. Certainly the frequency and severity of loosening increase with time, and one can say that the trouble is quite likely to occur in patients who are either active or heavy and thus put great demands on a prosthesis.

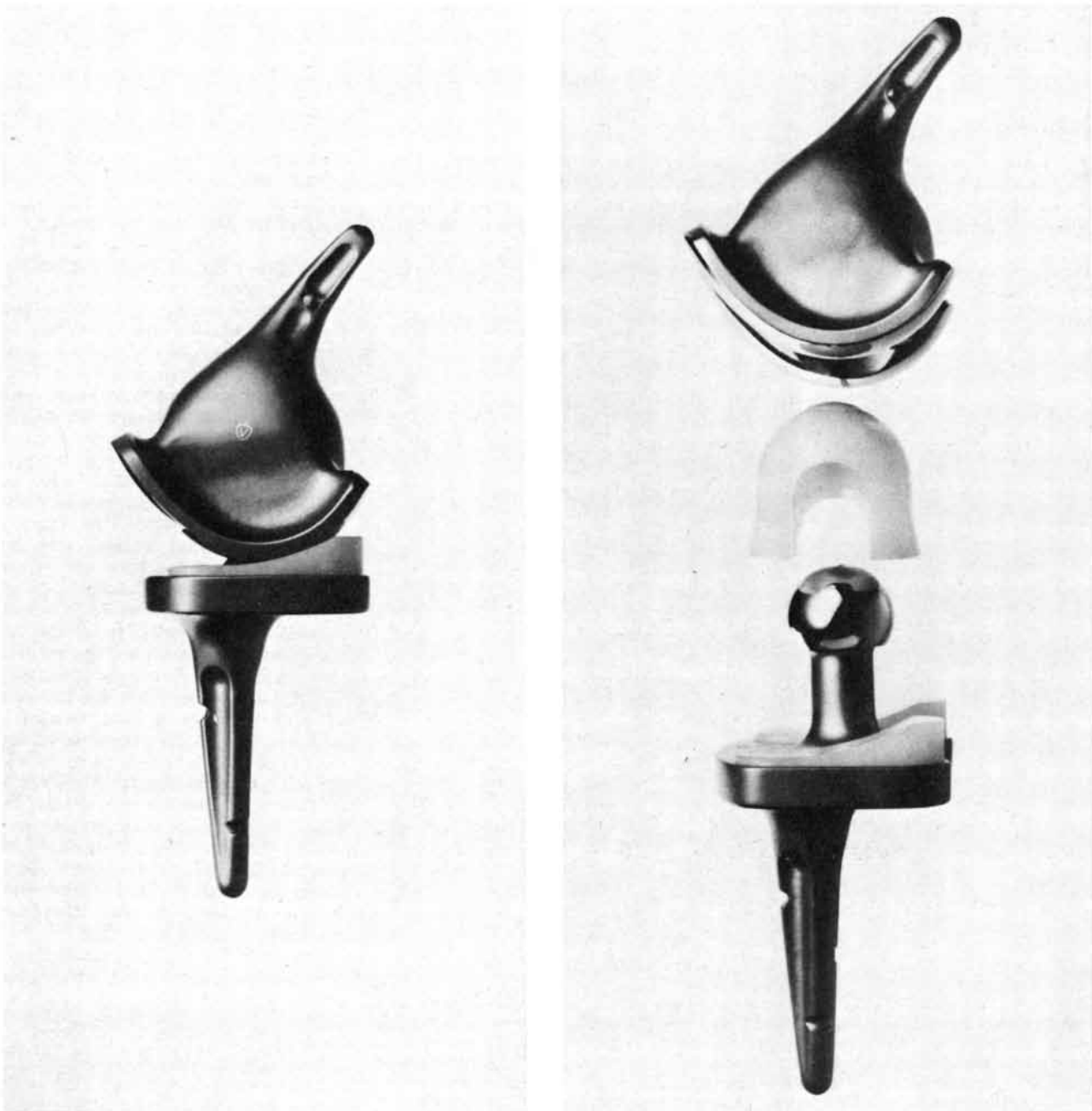
We have been studying this problem in our laboratories by imposing loads on junctions of bone and cement. The data indicate that polymethyl methacrylate fixes better to cancellous bone (the spongy material near the ends of long bones) than to cortical bone (the harder material at the ends). The data also reveal large variations in the properties of cancellous bone in different people, of different bones in one individual and even of different areas in a single bone.

These studies have been made with bones from cadavers, whereas living bone is an active tissue that develops strength and shape in response to loadings. We and many other investigators believe the loosening of prostheses is related to that mechanism. After a device has been implanted the bone adjacent to the cement actively resorbs and remodels itself to accommodate the new loadings. A tenuous balance then exists, which on occasion is dominated by resorption and results in loosening. Only carefully designed experiments with living animals can explore the problem of bone and cement more fruitfully.

Two alternative means of fixation are under study by other investigators. One takes advantage of the piezoelectric property of bone. Electric potentials

generated by natural loads on a bone or imposed from outside may stimulate the growth of the bone and so enhance the fixation of a prosthesis. The second approach is to implant devices with porous surfaces that provide places for bone to grow, so that the prosthesis becomes directly fixed. Appropriately sized matrixes of ceramics, polymers and metals have demonstrated considerable promise in this direction.

The replacement of the human knee joint is an active, rapidly changing field of endeavor. The 30,000 such replacements made in 1976, primarily in patients who were elderly and chronically disabled by arthritis, have provided reduction of pain and increased function for most of the recipients. Several things can be expected to keep the state of the art advancing. They include the improvement of materials for making prostheses, the identification of appropriate laboratory techniques for testing designs and the refinement of gait tests in the laboratory for evaluating a patient before and after an operation. Nevertheless, the primary cause of clinical failure, which is that the components of prostheses tend to work loose, must be overcome before the surgical replacement of the human knee joint can become more generally applicable. Even then patients, surgeons and engineers should not forget that the natural human knee joint is a remarkable product of evolutionary engineering. Although it can be imitated with increasing success, it will never be duplicated.



**THREE COMPONENTS of the Spherocentric prosthesis are shown assembled at the left and disassembled at the right. The metal bearing surfaces (the ball and the runners) are made of a highly polished alloy of cobalt, chromium and molybdenum, and the plastic bearing surfaces (the socket and the tracks) are made of polyethylene of high molecular weight. To install the device the surgeon implants the femoral component and then the tibial component. The polyethylene socket is placed over the ball and pressed into the femoral component, where locking devices hold it in place. A feature of the Spherocentric knee is that all plastic bearing surfaces are supported by metal.**

# The Three-dimensional Structure of Transfer RNA

*This nucleic acid plays a key role in translating the genetic code into the sequence of amino acids in a protein. The determination of its structure has clarified the mechanism of protein synthesis*

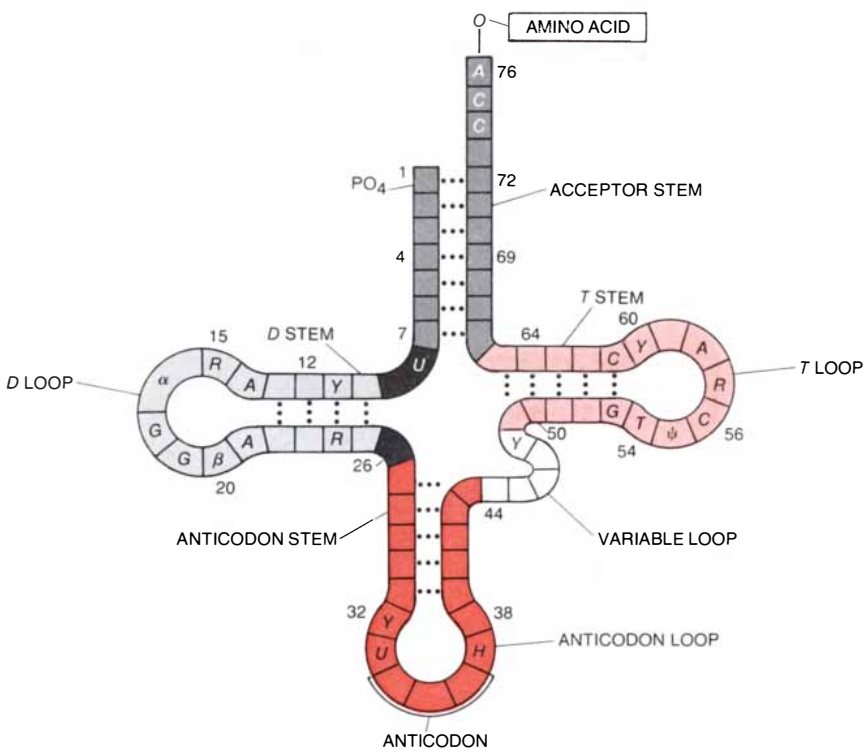
by Alexander Rich and Sung Hou Kim

It is now widely known that the instructions for the assembly and organization of a living system are embodied in the DNA molecules contained within the living cell. The sequence of nucleotide bases along the linear chain of the DNA molecule specifies the structure of the thousands of proteins that are the construction materials of the cell and the catalysts of its intricate biochemical reactions. By itself, however, a DNA molecule is rather like a strip of magnetic recording tape: the information embodied in its structure cannot be expressed without a decoding mechanism.

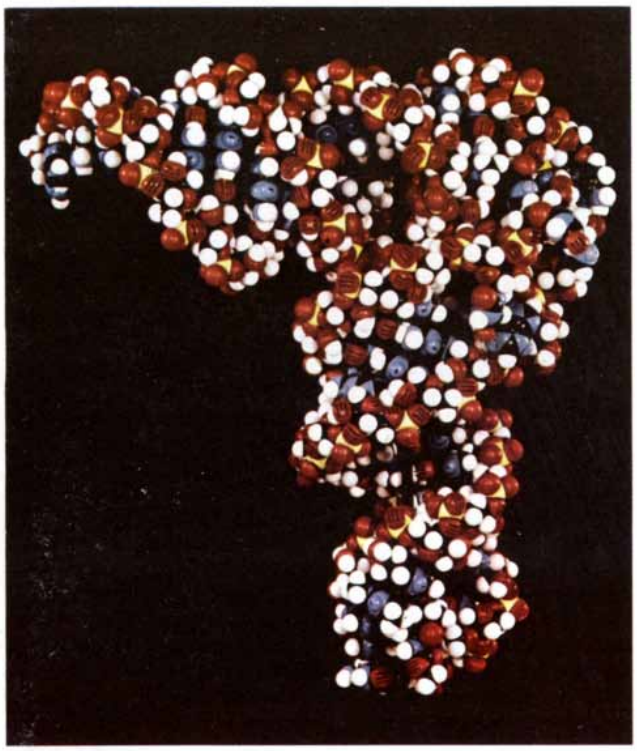
The development of such a decoding mechanism was one of the crucial events in the origin of life some four billion years ago. A basic biochemical system gradually evolved in which the nucleotide sequence of DNA is first transcribed into the complementary sequence of messenger RNA (abbreviated mRNA). The messenger RNA then directs the assembly of amino acids into the specific linear sequence characteristic of a given protein, a process called translation.

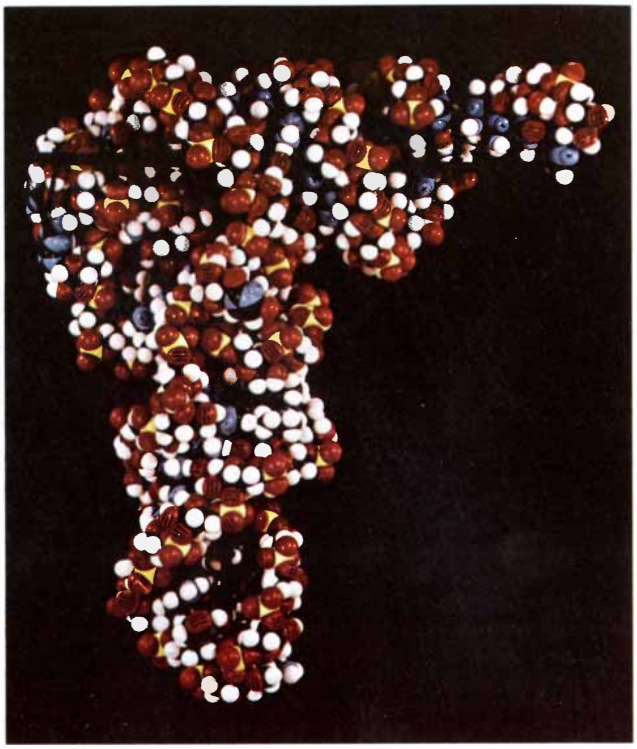
A central role in translation is played by another kind of RNA: transfer RNA (tRNA). The molecules of transfer RNA form a class of small globular polynucleotide chains (as distinct from fibrous polynucleotide chains such as DNA and mRNA) about 75 to 90 nucleotides long. They act as vehicles for transferring amino acids from the free state inside the cell into the assembled chain of the protein. This vital function as an intermediary between the nucleic acid language of the genetic code and the amino acid language of the working cell has made transfer RNA a major subject of research in molecular biology. Recently, in an important step toward the goal of understanding the process of translation in precise molecular terms, the three-dimensional structure of a tRNA molecule has been worked out at high resolution.

The translation of the nucleotide sequence of messenger RNA into protein proceeds in two major steps. First an amino acid molecule is attached to a particular transfer-RNA molecule, a reaction catalyzed by a large enzyme called an aminoacyl-tRNA synthetase. There are many different types of synthetase in living cells, each specific for one of the 20 different amino acids found in proteins. For example, leucyl-tRNA synthetase selectively binds to itself both the amino acid leucine and

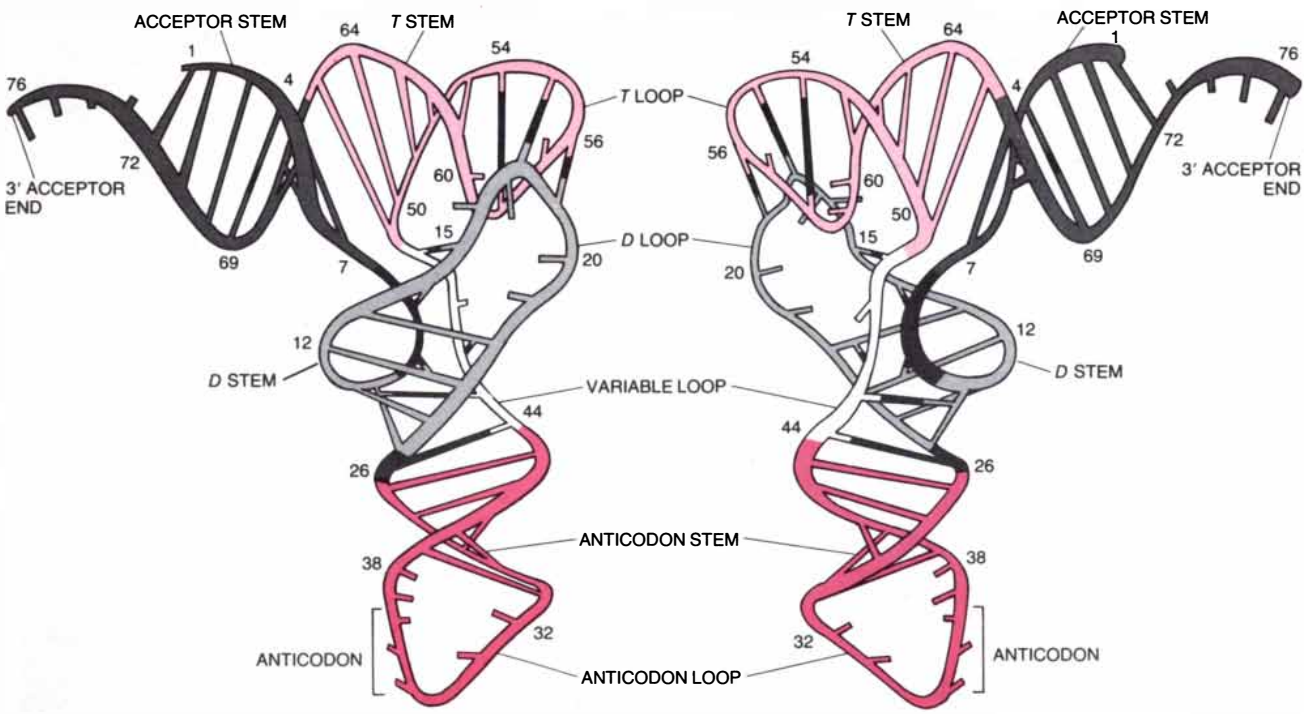



**CLOVERLEAF DIAGRAM is the two-dimensional folding pattern of the transfer-RNA (tRNA) molecule, which was first deduced in 1965 from the sequence of nucleotide building blocks in yeast alanine tRNA. Since then the diagram has been found to fit the nucleotide sequences of about 100 tRNA's isolated from plant, animal and bacterial cells. Nucleotide bases found in the same positions in all tRNA sequences are indicated. The ladderlike stems are made up of complementary bases in different parts of the polynucleotide chain that pair up and form hydrogen bonds, causing the chain to fold back on itself. The number of nucleotides in the various stems and loops is generally constant except for two parts of the *D* loop designated $\alpha$ and $\beta$ (which consist of from one to three nucleotides in different tRNA's) and the variable loop (which usually has four or five nucleotides but may have as many as 21). Abbreviations are *A* (adenosine), *G* (guanosine), *C* (cytidine), *U* (uridine), *R* (adenosine or guanosine), *Y* (cytidine or uridine), *T* (ribothymidine), $\psi$ (pseudouridine), *H* (modified adenosine or guanosine).**

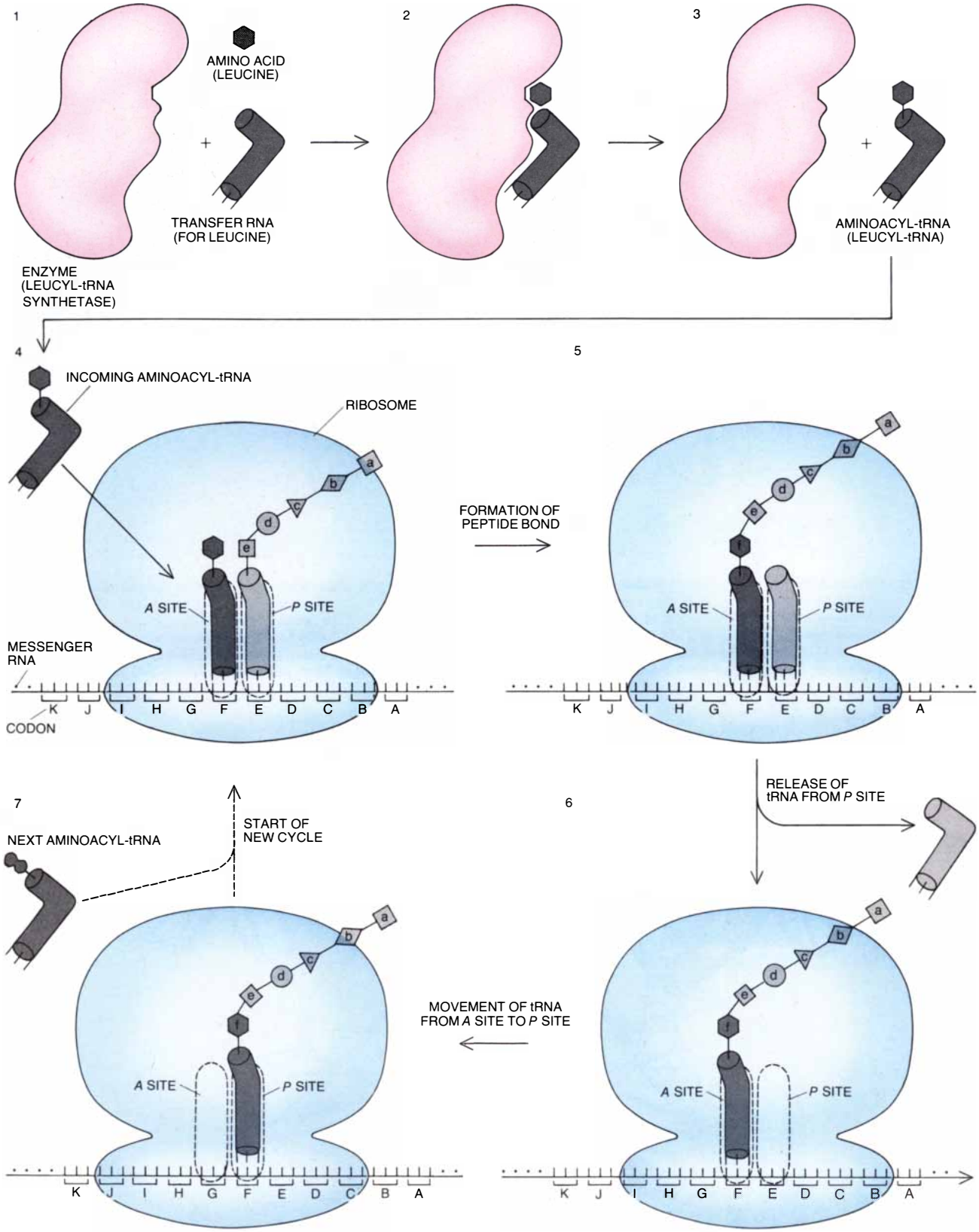




**SPACE-FILLING MODEL of yeast phenylalanine tRNA approximates the actual shape of the molecule. It was constructed on the basis of X-ray-diffraction analyses conducted in the authors' laboratories at the Massachusetts Institute of Technology and the Duke University School of Medicine. The polynucleotide chain of tRNA is folded into a compact *L*-shaped structure. During protein synthesis the amino acid phenylalanine is joined to the end of the horizontal arm of the *L*. Three nucleotide bases at the end of the vertical arm then recognize the genetic code for phenylalanine on the strand of messenger RNA (mRNA). Finally the amino acid is transferred to the growing protein chain. In this molecular model carbon is black, oxygen red, nitrogen blue, phosphorus yellow and hydrogen white.**




**FOLDING PATTERN of the polynucleotide chain in yeast phenylalanine transfer RNA is diagrammed. The sugar-phosphate backbone of the molecule is represented as a coiled tube, with the cross rungs standing for the nucleotide base pairs in the stem regions. The short rungs indicate bases that are not involved in base-base hydrogen bonding. The shading refers to cloverleaf diagram on opposite page.**




**FUNCTION OF TRANSFER RNA in the synthesis of a protein molecule is to make a chain of amino acids that reflects the nucleotide sequence of the template represented by messenger RNA. First a large enzyme called an aminoacyl-tRNA synthetase joins a specific transfer-RNA molecule to its corresponding amino acid with a covalent bond (*1–3*). The transfer RNA with the amino acid attached to it binds at the *A* site to the ribosome: the organelle where the amino acids are linked into the polypeptide chain of a protein. This interaction requires specific hydrogen bonding between the three codon bases on the messenger-RNA strand that specify an amino acid and the three anticodon bases of the transfer RNA (*4*). A transfer-RNA molecule in the adjacent *P* site then transfers the growing polypeptide chain to the tRNA in the *A* site (*5*). The "empty" tRNA leaves the *P* site and the ribosome moves along the messenger RNA a distance of one codon, so that the transfer RNA carrying the polypeptide chain is shifted from the *A* site to the *P* site (*6, 7*). Then the cycle begins anew.**

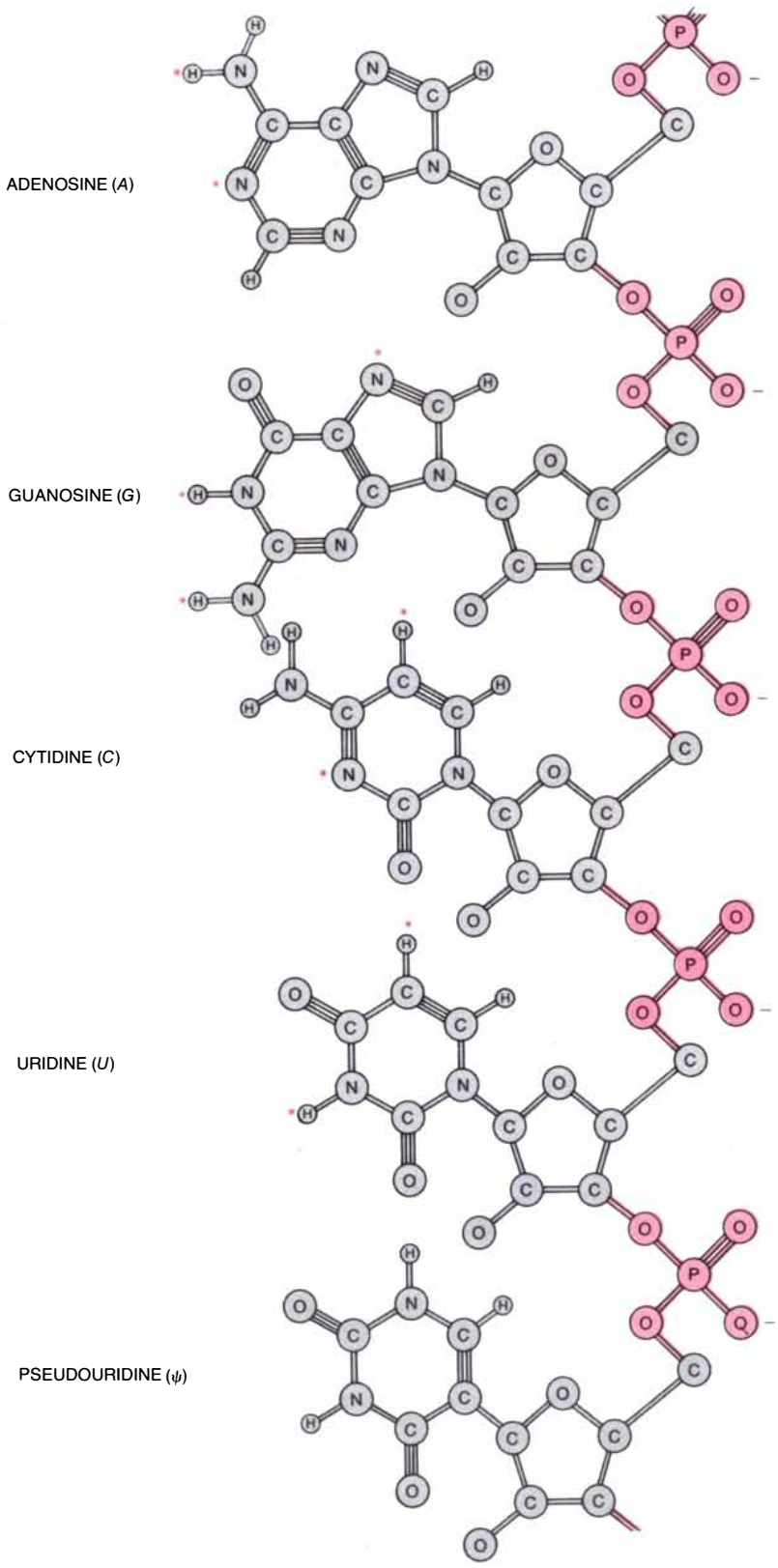
the tRNA for leucine; a complex of leucine and leucine tRNA is then formed and released. Once a tRNA has an amino acid attached to it, it is ready to participate in the second major step of protein synthesis.

This second step, the joining of the amino acids into a chain, is carried out inside the cellular organelle known as the ribosome, an aggregate of more than 50 different protein molecules and three RNA molecules. The ribosome is an intricate piece of molecular machinery designed to help translate the polynucleotide sequence of messenger RNA into the polypeptide sequence of protein. Although the exact details of the process have not been worked out, its general features are known.

Each amino acid in a protein is specified by a group of three adjacent nucleotide bases, designated a codon, on the messenger-RNA strand. There are four kinds of nucleotide base in messenger RNA, and so there is a total of $4^3$, or 64, possible codons. The relation between the codons and the amino acids they specify is the genetic code. The fact that the code appears to be the same in all living organisms is a remarkable proof of the unity of life at the molecular level.

Inside the ribosome are two sites that are involved in translation. One of them is the *A* site, which stands for aminoacyl-tRNA binding site. It is at this position that the transfer-RNA molecule and its attached amino acid are bound to the ribosome. The tRNA is positioned there partly by a set of specific interactions with the messenger RNA, which has already become associated with the ribosome. Three special nucleotide bases in the transfer-RNA molecule, designated the anticodon, interact with three complementary codon bases in the messenger RNA. The interaction involves the weak directional bonds known as hydrogen bonds, in which a hydrogen atom with a slight positive charge is shared by two other atoms with a slight negative charge. Hydrogen bonding is also the force that holds together the complementary nucleotide bases in the double helix of DNA: the base guanine on one strand of the helix is always paired with the base cytosine on the other strand, and the base adenine is always paired with the base thymine.

Immediately adjacent to the *A* site in the ribosome is the peptidyl-tRNA binding site, or *P* site. The transfer-RNA molecule with the growing chain of amino acids attached to it is bound to this site and specifically interacts with the next codon triplet of bases on the messenger-RNA chain. In the course of protein synthesis the growing polypeptide chain is cleaved from the tRNA molecule in the *P* site and is transferred to the end of the single amino acid attached to




**NUCLEOSIDES, consisting of a nucleotide base attached to the sugar ribose, are joined by negatively charged phosphate ($PO_4$) groups to form the polynucleotide chain of transfer RNA. The four major nucleosides in the molecule are adenosine, guanosine, cytidine and uridine. Transfer RNA also incorporates many modified nucleosides, more than 50 of which have been identified. The commonest modification is the replacement of a hydrogen atom by a methyl group ($CH_3$). This reaction is catalyzed by special enzymes and occurs at the sites indicated by an asterisk. Other structural modifications also occur. For example, the nucleoside pseudouridine (ψ) has its base attached to the ribose through a carbon atom instead of a nitrogen atom.**

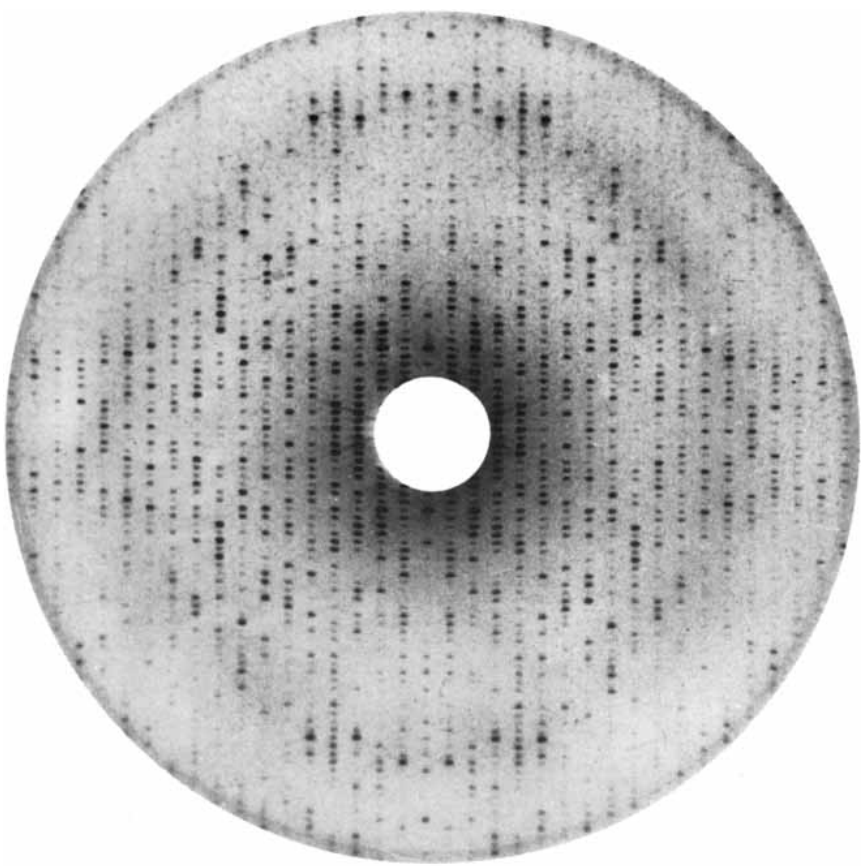
the tRNA molecule in the *A* site. Once the transfer has been accomplished (the cleavage and rejoining reactions are carried out by an enzyme in the ribosome) the growing polypeptide chain has been elongated by one amino acid. The "empty" tRNA molecule is then released from the *P* site, and the messenger RNA and the newly elongated peptidyl-tRNA are shifted from the *A* site to the *P* site. A new transfer RNA with an amino acid attached to it now finds its way into the ribosome and becomes lodged in the vacated *A* site through the specific interaction between its anticodon bases and those of the next codon on the messenger-RNA strand. The system is now back to its starting point, ready to begin another cycle of events in which one more amino acid will be added to the chain. This stepwise addition is repeated until the complete protein has been synthesized.

The process of polypeptide-chain elongation is fairly rapid: it occurs as many as 20 times a second in a bacterial cell and about once every second in a mammalian cell. For example, the hemoglobin molecule is a large protein consisting of four polypeptide chains with about 140 amino acids each. The synthesis of one such chain would take seven seconds in a bacterial cell and two or three minutes in a mammalian cell. Even though this rate of synthesis is fairly high there are surprisingly few errors in translation, because the machinery of the ribosome ensures a careful fit between each transfer-RNA molecule and the messenger RNA. The process is also very efficient, because there are usually several ribosomes at work translating a single strand of messenger RNA.

In order to understand how transfer RNA carries an amino acid into the ribosome and transfers it to the growing polypeptide chain it is essential to have a knowledge of the three-dimensional structure of the tRNA molecule. One of the first clues to that structure emerged from the nucleotide sequence of a yeast tRNA specific for the amino acid alanine, which was determined in 1965 by Robert W. Holley and his colleagues at Cornell University. These workers noted that there were certain regions of the sequence that would be complementary if the chain were folded back on itself. Specifically, these regions could form hydrogen bonds with each other, much like the base pairing in the double helix of DNA (except that in RNA's the base adenine is paired with uracil instead of thymine). The polynucleotide chain of transfer RNA could thus be arranged in such a way that it would contain hydrogen-bonded double-strand regions called stems and nonbonded regions called loops. The postulated combination of stems and loops resembled a four-leaf clover, and so it became known as the cloverleaf diagram.

One feature of the nucleotide sequence of transfer RNA is that it includes many unusual bases, most of them common RNA bases that have been modified by the addition of one or more methyl groups ($CH_3$). Because of this feature some parts of the cloverleaf diagram have been named for the modified bases that occur in them. For example, the *T* loop is so named because it includes thymine (*T*), which is found in DNA but is not found in any RNA species other than transfer RNA. Similarly, the *D* loop usually includes the modified base dihydrouracil (*D*). Other regions of the cloverleaf are the variable loop, which in different tRNA's has different numbers of nucleotides (ranging from four to 21), the anticodon loop, which includes the three bases of the anticodon, and the acceptor stem, which accepts the amino acid specific to that particular tRNA.

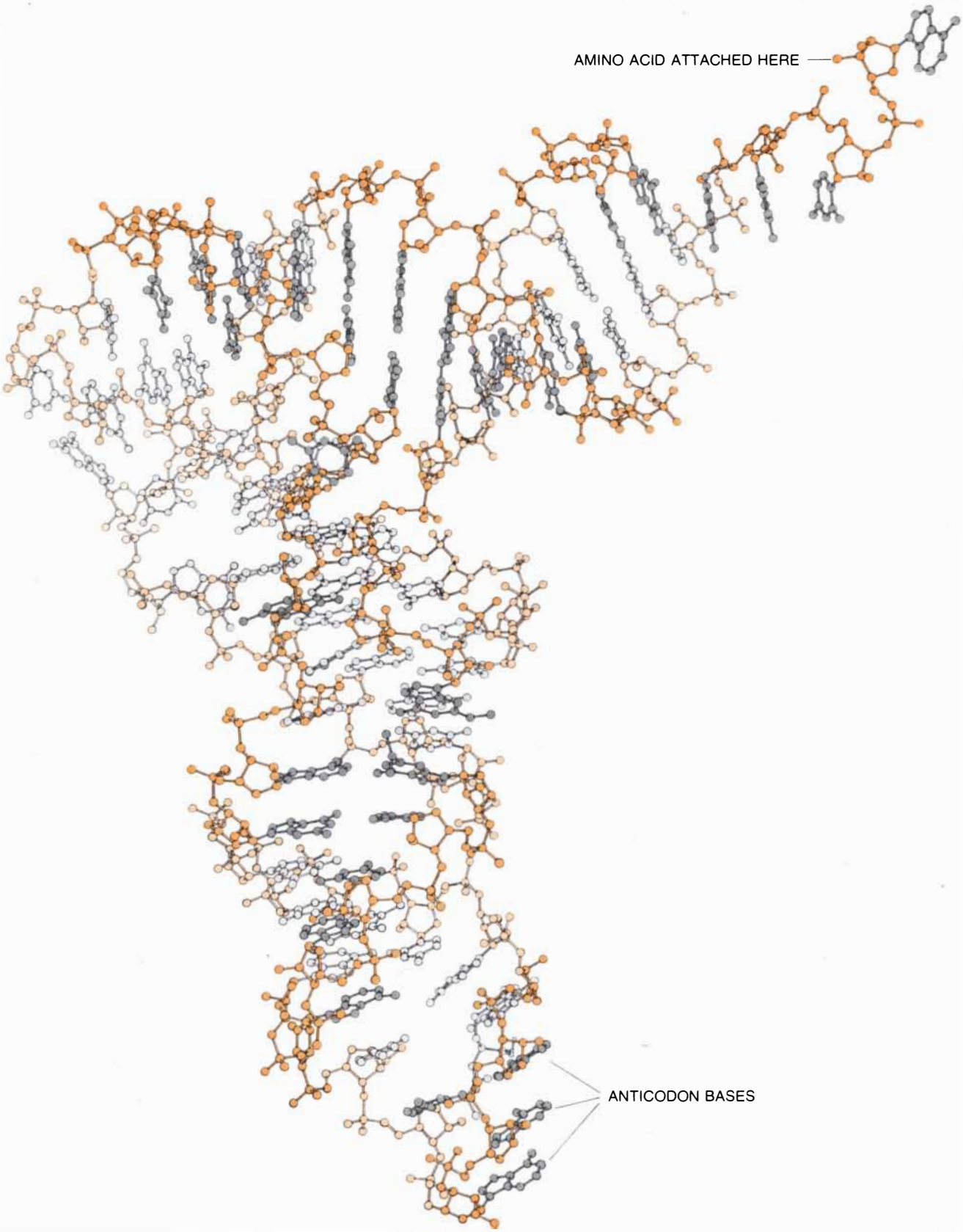
An interesting feature of the cloverleaf diagram is the presence of nucleotide sequences that are constant in all 100 of the tRNA sequences that have been determined so far. The number of base pairs in the stem regions is also constant: seven in the acceptor stem, five in the *T* stem, five in the anticodon stem and three or four in the *D* stem. These features are maintained in tRNA molecules from plants, animals, bacteria and viruses. Indeed, the pattern of stems, loops and constant nucleotides found in the tRNA cloverleaf appears to have the same universality as the genetic code. Much of the explanation for this constancy was later provided by the three-dimensional structure of tRNA.



**X-RAY-DIFFRACTION PATTERN was one of patterns utilized by the authors to deduce the three-dimensional structure of transfer RNA. The pattern was created by directing X rays into a crystalline array of transfer-RNA molecules and capturing the scattered beams on a piece of film. The spots contain information about distribution of electrons within the crystal.**

Today the three-dimensional structure of large biological molecules is commonly determined by means of the X-ray-diffraction analysis of molecular crystals. A molecular crystal is an assembly of molecules packed together in a regular three-dimensional array. When X rays with a wavelength comparable to the distance between atoms are directed into the crystal, they are diffracted, or scattered, in a variety of directions by the electron clouds of the atoms in the crystal lattice. The diffraction pattern of the crystal can be detected as a series of spots on a piece of X-ray film, with the blackening of the emulsion being proportional to the intensity of each scattered beam.

This pattern contains a great deal of information about the structure of the crystal. For one thing, the amplitude of the wave scattered by an atom is proportional to the number of electrons in the atom, so that a carbon atom will scatter




**DETAILED SKELETAL MODEL of yeast phenylalanine tRNA shows the hydrogen-bonding interactions between the nucleotide bases. It was derived in 1974 from an X-ray-crystallographic study at a resolution of three angstroms. Projection shown here was generated on a computer by one of the authors (Kim). Ribose-phosphate backbone of the molecule is shaded in color; the bases are shaded in gray.**

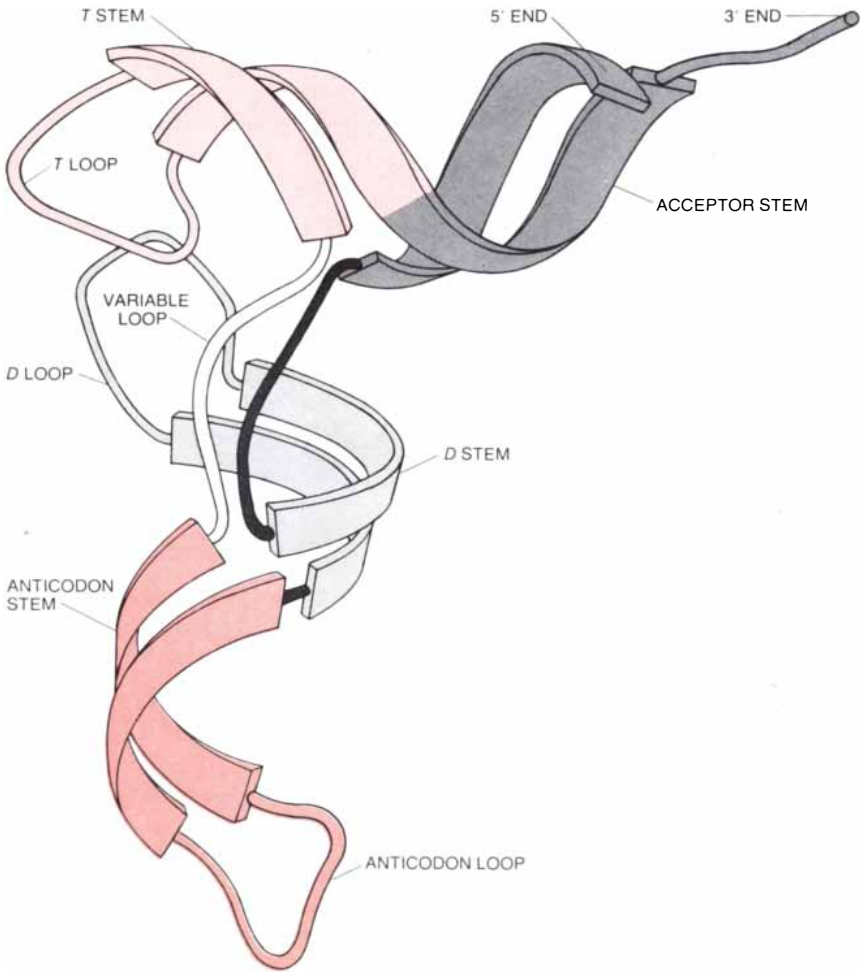
X rays six times more strongly than a hydrogen atom. Secondly, the scattered waves recombine inside the crystal lattice; depending on whether they are in phase or out of phase, they will either reinforce or cancel one another. The way the scattered waves recombine depends only on the arrangement of the atoms in the crystal, and so it is possible to reconstruct the image of a molecule from its diffraction pattern.

To analyze the three-dimensional structure of a large protein or nucleic acid molecule a crystal of the substance is first prepared. Then the crystal is mounted in a capillary tube and positioned in a precise orientation with respect to the X-ray beam and the film. The crystal is rotated along each of its axes to yield a series of X-ray photographs in which there is a regular array of spots of various intensities. Each of these photographs is actually a two-dimensional section through a three-dimensional array of spots.

Next the intensities of all the spots in the diffraction patterns are measured, either from the film or through the use of a Geiger counter. Additional information is needed, however, before one can establish the three-dimensional structure, namely the phases of the scattered X-ray beams with respect to an arbitrary fixed point in the crystal. This information is obtained by inserting heavy-metal atoms such as those of platinum or gold into the crystal lattice as markers. The addition of these atoms changes the diffraction pattern slightly and enables one to calculate the phases of the diffracted beams.

With this information in hand it is possible to calculate the density of the electrons at a large number of regularly spaced points in the crystal, making use of a Fourier series: a sum of sine and cosine terms. A high-speed computer is needed to handle the enormous number of terms (more than a billion) involved in determining the structure of a large protein or nucleic acid molecule. The first such molecule whose structure was determined in this way was the protein myoglobin; the feat was accomplished in 1958. Today the technique is almost routinely exploited for the structural analysis of large molecules.

The end product of the technique is a three-dimensional map showing the distribution of the electrons in the crystal. The map is usually drawn as a series of parallel sections stacked on top of one another, each section being a transparent plastic sheet on which the electron-density distribution is represented by black contour lines resembling those of a topographical map. The critical factor in the interpretation of the electron-density map is its resolution, which is determined by the number of scattered-beam intensities incorporated in the Fourier series. For example, a map at a resolution of six angstroms, derived from the innermost spots of the diffraction pattern, may reveal the general shape of the molecule but few additional structural details. (An angstrom is $10^{-10}$ meter, about the diameter of a hydrogen atom.) Maps of higher resolution are needed to delineate groups of atoms, which may be three to four angstroms apart, or individual atoms, which are from one to two angstroms apart. A large molecule is usually analyzed at different levels of resolution, making it possible to visualize different features of the structure. The ultimate resolution of an X-ray analysis, however, is determined by the degree of perfection of the crystal. For large biological molecules the best resolution one can usually obtain is about two angstroms.



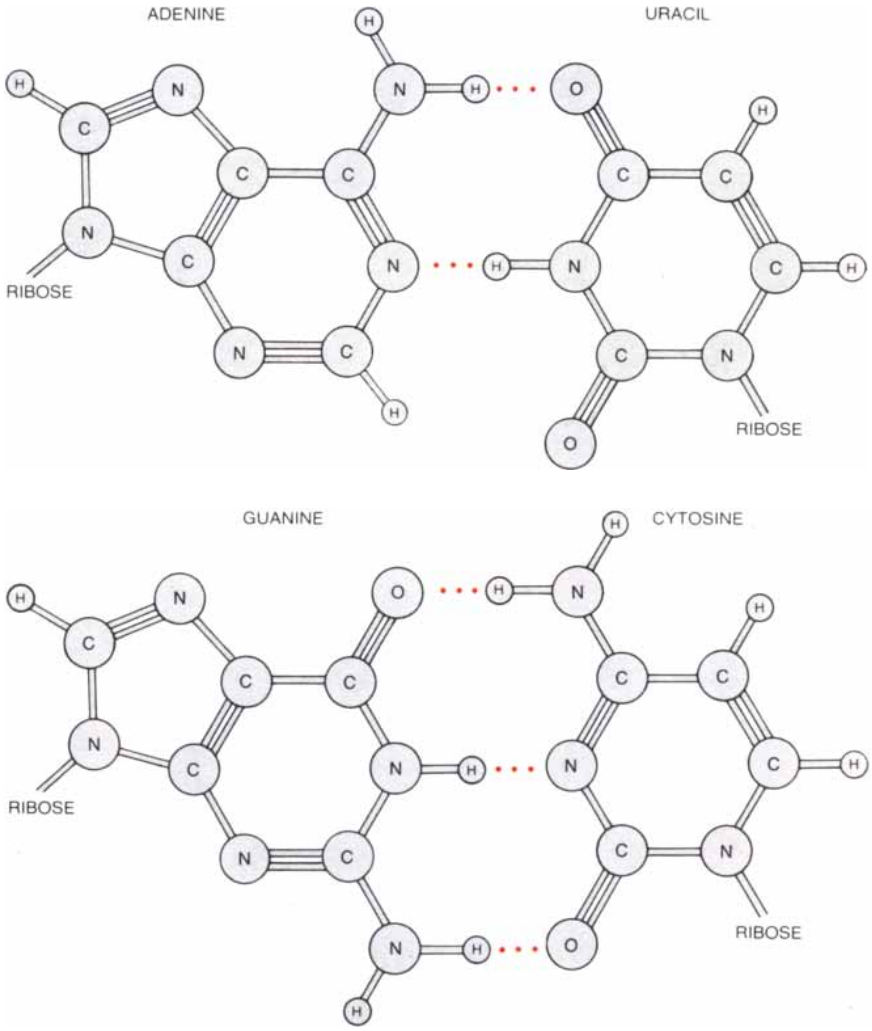

**HELICAL SEGMENTS of the tRNA molecule, corresponding to the four stems of the cloverleaf diagram, are represented by ribbons in this schematic view. The two helical regions are arranged at right angles to provide the structural framework for the *L*-shaped folding pattern. Each region consists of about 10 base pairs, corresponding to roughly one turn of the double helix. The helix in these regions is similar to the double helix of DNA, except that in transfer RNA the two strands of the helix are formed by different parts of same polynucleotide chain.**

With transfer RNA the first step of the process—crystallizing the molecule—turned out to be a major hurdle. In 1968 our group at the Massachusetts Institute of Technology and workers in five other laboratories discovered that it was possible to crystallize different species of tRNA by dissolving them in various mixtures of solvents and allowing the solvents to evaporate slowly. This advance caused great excitement among molecular biologists, since it seemed that the major hurdle had been overcome and that the three-dimensional structure of transfer RNA was within reach. Our elation was soon followed by some degree of despair when it was realized that although many different species of tRNA had been crystallized, most of the crystals were quite disordered. As a result the crystals provided diffraction patterns with very low resolution (usually between 10 and 20 angstroms) and hence could reveal little of the detailed structure of the molecule. Although it was exciting to discover that tRNA was crystallizable, it was frustrating to realize that further work had to be done before suitable material was available for X-ray-diffraction analysis.

Together with Gary J. Quigley and Fred L. Suddath we made a concerted

effort to find conditions where tRNA would form a well-ordered crystal that would produce an X-ray-diffraction pattern with sufficient resolution to reveal the three-dimensional structure of the molecule. For two years we surveyed a large number of different tRNA species and crystallizing conditions. Finally we made an important discovery: the addition of spermine, a small positively charged molecule, resulted in the formation of a highly ordered crystal of a tRNA extracted from yeast cells that was specific for the amino acid phenylalanine. The spermine-stabilized crystal had a diffraction pattern that extended out to a resolution of nearly two angstroms.

Late in 1972, working with Alexander McPherson, Daryll Sneden, Jung-Ja Park Kim and Jon Weinzierl, we obtained an electron-density map of the crystal in which we were able to trace the backbone of the polynucleotide chain of the tRNA at a resolution of four angstroms. At that resolution it was not possible to perceive the individual bases of the polynucleotide chain, but the electron-dense phosphate ($PO_4$) groups along the backbone of the molecule could be seen as a string of beads coiled in three-dimensional space. To our great surprise the polynucleotide chain was organized in such a way that the molecule was shaped like an *L*, with one arm of the *L* made up of the acceptor stem and the *T* stem and the other arm made up of the *D* stem and the anticodon stem. The complementary hydrogen-bonded sequences that had been identified in the cloverleaf diagram were clearly seen as RNA double helixes. The various loops occupied strategic positions either at one end of the molecule or at the corner of it, where the *T* and *D* loops were coiled together in a complex manner.



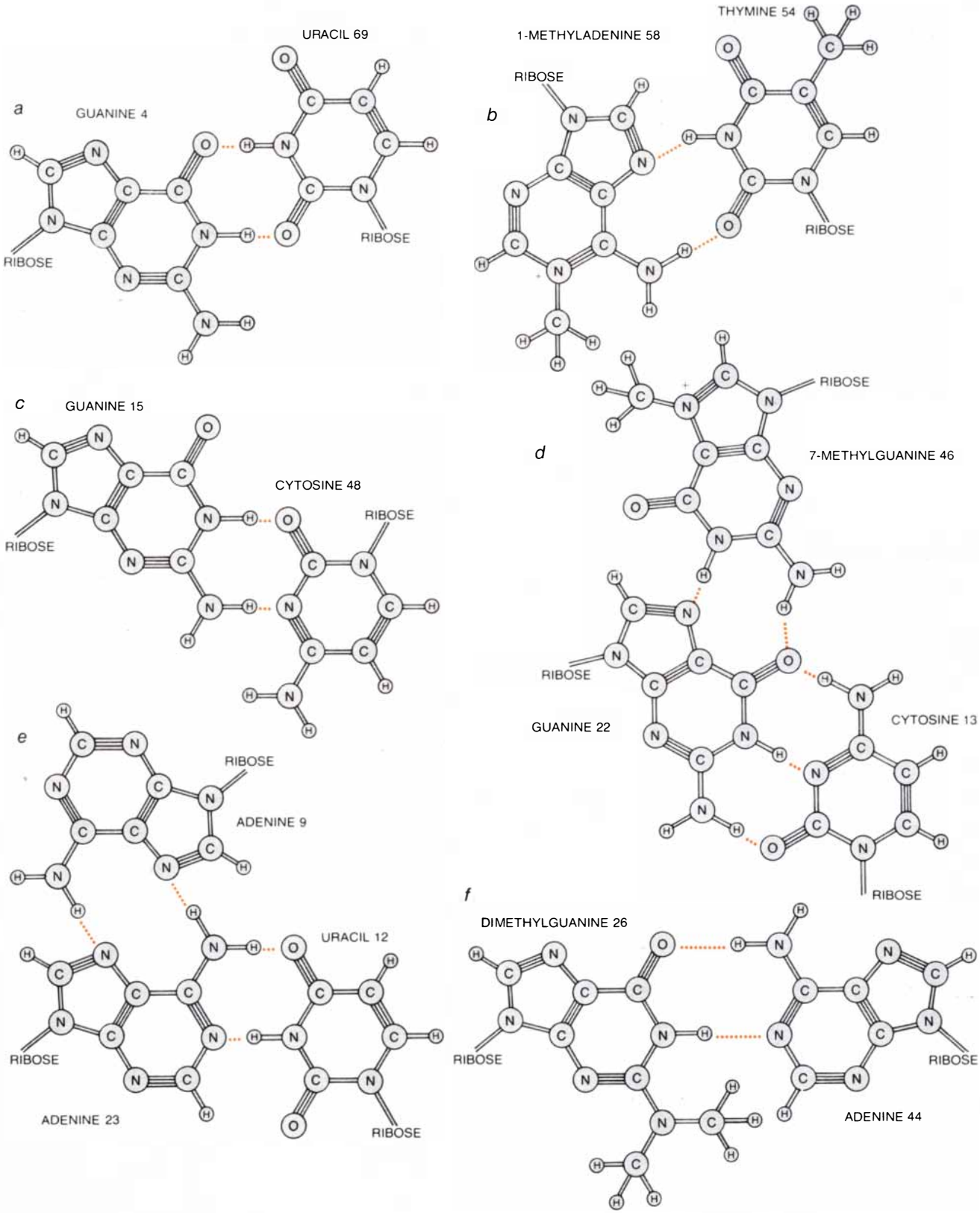

**COMPLEMENTARY HYDROGEN BONDING between bases in the helical regions of the transfer-RNA molecule follows the pattern first proposed by James D. Watson and Francis H. C. Crick for double helix of DNA (except that in tRNA uracil replaces thymine). Adenine and uracil pair with two hydrogen bonds, whereas guanine and cytosine pair with three.**

This folding of the molecule was entirely unexpected. Over the preceding few years a number of investigators had recognized the features common to the cloverleafs of all transfer RNA's and had tried to predict how the tRNA molecule might be folded. As is so often the case, however, nature proved to be subtler than had been imagined. The *L*-shaped folding served to explain a number of chemical observations that had accumulated, and it also made people wonder what functional purpose was served by this unusual shape.

By mid-1974, together with Joel L. Sussman, Andrew H.-J. Wang and Nadrian C. Seeman, we had interpreted the electron-density map at a resolution of three angstroms. The overall form of the molecule was the same as the one apparent at four angstroms, but now many more details were visible, including the positions of most of the nucleotide bases. At about this time Jon Robertus, Brian F. C. Clark, Aaron Klug and their colleagues at the British Medical Research Council Laboratory of Molecular Biology in Cambridge described their X-ray-crystallographic analysis of a transfer RNA at a resolution of three angstroms. Their tRNA was the same spermine-stabilized yeast phenylalanine tRNA, but it was in a different crystal form. Even though the molecule was packed differently in the crystal lattice, comparison of the two three-dimensional structures resulting from the analyses showed that the structures were virtually identical. This agreement between the findings of the two groups provided important evidence that the structure of the tRNA molecule is independent of how it is packed in a crystal.

The map of the tRNA molecule at a resolution of three angstroms confirmed our earlier finding that it is organized into two columns of nucleotide bases stacked at right angles to each other. These columns have both helical and nonhelical regions corresponding to the stems and loops of the cloverleaf diagram. The high-resolution map further revealed that the two helical regions each consist of about 10 base pairs, corresponding to one turn of the double helix, and possess the same type of hydrogen bonding between complementary nucleotide bases as that found in the double helix of DNA.

In the nonhelical parts of the tRNA molecule many of the nucleotide bases are oriented with their hydrogen-bonding groups pointed toward the interior of the molecule, where they participate in a variety of unusual hydrogen-bonding interactions known as tertiary interactions. Such bonds may occur between two or three bases that are not usually considered complementary, between a base and the ribose-phosphate backbone of the transfer-RNA chain or even between different parts of the backbone itself. The fact that several tertiary interactions in tRNA involve the hydroxyl (OH) groups of the sugar ribose is of particular interest, because hydroxyl groups are absent from the sugar molecules of DNA. Such tertiary interactions




**UNUSUAL INTERACTIONS between bases stabilize the folding pattern of the transfer-RNA molecule. In the acceptor stem the normally noncomplementary bases guanine and uracil are held together by two hydrogen bonds as the result of a slight lateral "wobble," or displacement, in one of the bases (*a*). In the *T* loop 1-methyladenine is paired with thymine, a modified form of uracil that has an added methyl group (*b*). In the core region of the molecule, immediately below the corner, guanine and cytosine are paired, but with two hydrogen bonds instead of the usual three. This pairing is of the *trans* type because the ribose groups fall on opposite sides of the pair (*c*). Also in the core region are two complex systems of hydrogen bonding involving three bases in the same plane (*d, e*). In the region joining the *D* stem and the anticodon stem a dimethylated guanine is paired with an adenine by two hydrogen bonds (*f*). Because of the bulky methyl groups on the guanine this base pair is not planar; the two bases are tilted about 25 degrees away from each other like the blades of a propeller. The dimethyl guanine is stacked at the bottom of the *D* stem and the adenine is stacked at the top of the anticodon stem, an arrangement that stabilizes the junction between the two stems. For a more schematic view of these interactions see the diagram on page 62.**

are simply not needed in a regular linear nucleotide chain such as that of DNA, but they are essential for stabilizing the complex coiling of the polynucleotide chain in tRNA.

One unusual hydrogen-bonding arrangement was found in the acceptor stem, where the pair of nucleotide bases guanine-uracil occurs in place of the normal pair guanine-cytosine or adenine-uracil. The possibility of such a pairing had been suggested several years earlier when Francis H. C. Crick made the observation that it was likely certain additional types of base pairing would be found at the position of the third base in the interaction between the messenger-RNA codon and the transfer-RNA anticodon. One of the "unconventional" arrangements Crick had postulated was a guanine-uracil pair that would be connected by two hydrogen bonds as a result of a "wobble," or slight lateral displacement, in one of the bases. Continued analysis and refinement of the electron-density map at a resolution of 2.5 angstroms confirmed the wobble type of pairing between guanine and uracil in the acceptor stem.

Several other novel arrangements of hydrogen bonds have been discovered among the tertiary base-base interactions in the transfer-RNA molecule [*see illustration on opposite page*]. The variety of these interactions was one of the most surprising findings to emerge from our structure-determination work.

Most of the flat nucleotide bases in transfer RNA are organized in two stacked columns that form the arms of the *L*-shaped molecule. This arrangement explains the unusual stability of tRNA. If one heats a solution containing tRNA molecules, they will denature, that is, the polynucleotide chain will unravel and assume random conformations in the solution. As soon as the solution cools, however, the molecule will immediately snap back to its native conformation. This behavior is quite different from that exhibited by most proteins, which denature irreversibly; egg albumin, for example, turns white and opaque when the egg is boiled and stays that way when the egg is cooled.

Why does the transfer-RNA molecule revert so readily to its native structure? It is known that the stacking interaction between the adjacent nucleotide bases in the interior of the DNA double helix is one of the major stabilizing features of that molecule. Similarly, the bases of tRNA are predominantly hydrophobic (water-repelling), so that they retreat from the surrounding solvent into the interior of the folded polynucleotide chain; this behavior helps to return the tRNA molecule to its native—and stablest—conformation. In proteins there is usually no comparable interaction that will make the polypeptide chain refold spontaneously. Thus it appears that the structure of tRNA is organized to preserve the stabilizing feature of the stacking interactions between bases. At the same time some very complex molecular architecture holds the two stacked columns at right angles to each other.

An important aspect of the tertiary interactions found in yeast phenylalanine tRNA is the fact that many of them involve bases that are the same in the polynucleotide sequences of all tRNA's. Moreover, bases occurring in regions of the polynucleotide chain that have variable numbers of nucleotides are usually unstacked and located in loops that protrude from the surface of the tRNA molecule. These findings suggest that the structural framework of yeast phenylalanine tRNA may accommodate the nucleotide sequences found in other tRNA's. For example, in yeast phenylalanine tRNA one variable region of the *D* loop contains two nucleotides, and this segment of the polynucleotide chain arches away from the molecule and returns. If there were more nucleotides in this region, it is likely that the bulge would be somewhat larger; conversely, if there were fewer nucleotides, it would be smaller. The size of such variable loops, however, would not affect the overall folding pattern of the molecule.

A number of important problems concerning the three-dimensional structure of transfer RNA's in general remain unsolved. It is not clear, for example, what the detailed structure will be for tRNA's with very large variable loops. The structure of "initiator" tRNA's, which start the synthesis of proteins by laying down the first amino acid, is also of interest. Some initiator tRNA's have polynucleotide sequences that depart somewhat from the sequences common to other tRNA's, particularly in the *T* loop. It is quite likely that these differences are associated with a structure slightly different from that of yeast phenylalanine tRNA.

Our crystals of yeast phenylalanine tRNA contain almost 75 percent water. It is important to ask whether the molecule has the same form in solution (where it is biologically active) that it has in the crystal. Fortunately there have been numerous investigations of yeast phenylalanine tRNA in solution. These studies make it possible to correlate the structure observed in the crystal with various chemical characteristics of the molecule. For example, one of the features of yeast phenylalanine tRNA in solution is that some nucleotides seem to be readily available for chemical modification when chemical reagents are added to the solution, whereas other nucleotides are not. This disparity was puzzling until the structure of the molecule in the crystal emerged. Then it became apparent that only certain nucleotides, such as those that protrude from the molecule in the crystalline state, are readily available for chemical modification. In general there is an excellent correlation between the susceptibility of a region of the tRNA molecule to chemical modification and the accessibility of that region of the molecule in the crystalline state.

Several other types of experiments carried out in solution can be interpreted in the light of the three-dimensional structure, including experiments based on nuclear magnetic resonance, which is sensitive to the three-dimensional structure of a molecule. Several investigators have found a good correlation between the nuclear-magnetic-resonance signals obtained from transfer-RNA molecules in solution and the three-dimensional structure deduced from X-ray-diffraction analysis of yeast phenylalanine tRNA in the crystal. These and other findings provide convincing evidence that the structure of the tRNA molecule in the crystal is the structure of the biologically active form of the molecule.

Such correlations are important because the principal reason for determining the structure of a biological molecule is to perceive how it functions in a biological system. What has the structure of transfer RNA taught us about how the molecule works? Here one can speak with considerably less confidence because the necessary experiments have not yet been carried out. First one would like to know how the enzyme aminoacyl-tRNA synthetase recognizes and selects only the correct tRNA for attachment to a specific amino acid. If this process is to be understood fully, it will be necessary to determine the three-dimensional structure of the synthetase when it is complexed with the tRNA, so that the nature of the specific interactions between the enzyme and the nucleic acid can be perceived. Studies of this kind are now under way in many laboratories, and the answers should be forthcoming in the near future. Already some experiments suggest that in the recognition process certain regions of the tRNA molecule are more important than others.

Another major question is: Why does the tRNA molecule have an *L*-shaped form in which the anticodon is more than 76 angstroms away from the attached amino acid? The definitive answer has not yet been obtained, but it is quite likely that the shape of tRNA is related to its essential transfer function inside the ribosome. For two adjacent tRNA's in the *A* and *P* sites to be brought close together on the messenger-RNA strand so that the growing polypeptide chain can be transferred from one to the other, it may have been necessary for the cell to have evolved a tRNA molecule bent in this peculiar fashion. Perhaps the acceptor arm of the *L* rotates inside the ribosome so that the protein chain can be transferred to the tRNA bound to the next codon on

the messenger-RNA chain. This second tRNA would then take the chain onto its own amino acid and in turn pass the chain along. The considerable distance between the end of the acceptor stem and the anticodon loop of the tRNA molecule may also be functionally important in that different ribosomal proteins can simultaneously interact with several regions of the tRNA in order to help maintain the precision of protein synthesis.

The view of the tRNA molecule that has been obtained from X-ray-diffraction analyses of molecules in a crystal is essentially a static one. In its natural environment within the cell the molecule may undergo conformational changes, particularly when it interacts with large molecular structures such as the ribosome. Recent experiments suggest that inside the ribosome the *D* loop and the *T* loop of the tRNA molecule may move away from each other when the molecule shifts from the *A* site to the *P* site. It is also possible that the shape of the anticodon loop is modified when it comes in contact with the messenger RNA inside the ribosome. A fuller evaluation of these proposals will have to await the results of further research.

At the beginning of this article we described the role of transfer RNA in protein synthesis in some detail because that is the molecule's most essential role in biological systems. Without the tRNA molecule genetic information could not be expressed in the synthesis of proteins. In addition tRNA participates in a variety of other processes, some of which are of great importance. For example, tRNA molecules can donate amino acids to preformed protein molecules or to the molecular structure of the cell wall in bacteria independently of the ribosome.

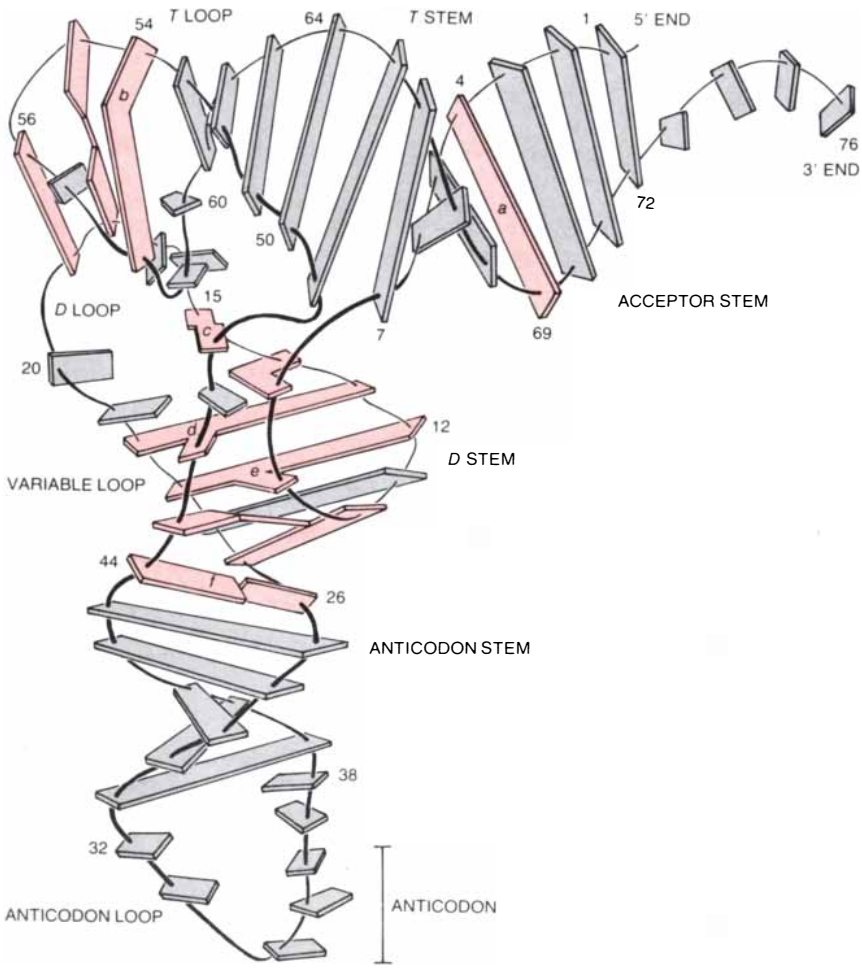
Another process in which tRNA participates is the control of gene expression. Certain tRNA's with an amino acid attached to them are known to determine whether or not genes, that is, segments of DNA, will be expressed by regulating their transcription into messenger RNA. The detailed mechanism is not known, but in some systems the control function is associated with a particular modified nucleotide in the tRNA molecule, for example a uracil that has been converted into a pseudouracil. It is thought that tRNA helps to control the expression of many different genes, although the exact number is not known. The instances that have been most intensively studied are those of the genes that regulate the synthesis of amino acids, where tRNA-mediated regulation plays a major role.

Other observations suggest that transfer RNA may be involved in still more types of biochemical regulation. For example, in the course of embryonic development one kind of modification of certain nucleotides in a tRNA gives way to another kind. Similarly, when a normal cell becomes cancerous, the kinds of modifications of nucleotides in its tRNA molecules change substantially. It is not yet known whether these transformations are associated with the regulatory functions of tRNA.

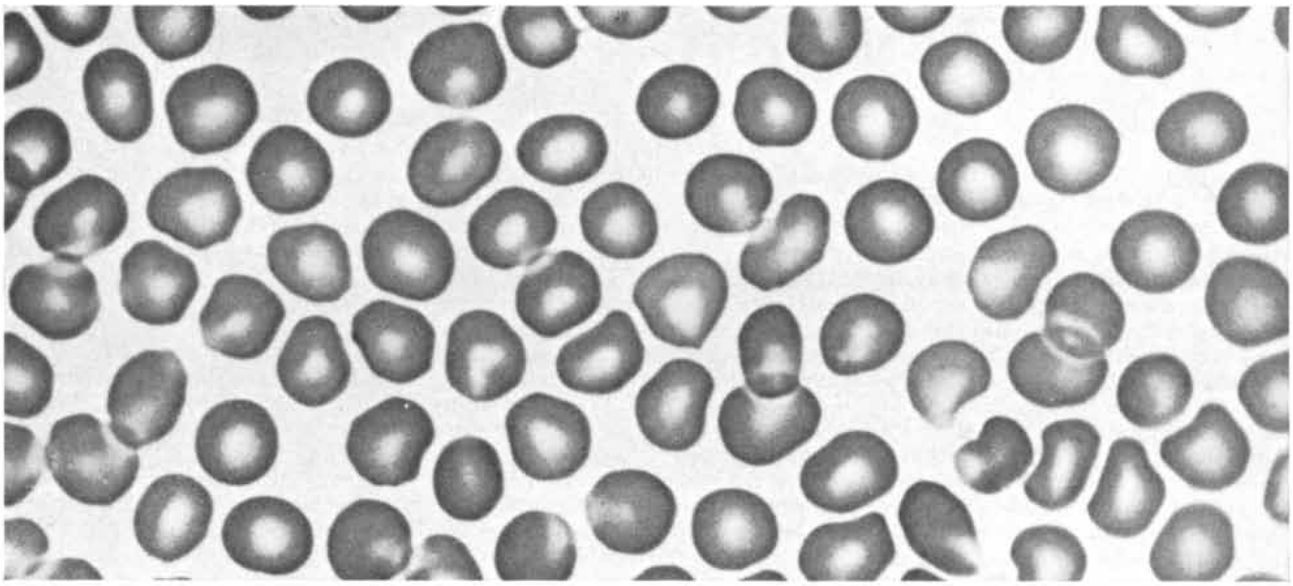
Another mysterious area concerns the relatively high number of modifications in the nucleotide sequences of the *D* loop as well as in those of the variable loop. Why has nature gone to such trouble to vary the nucleotides that project from the surface of the molecule? It is generally believed these sequences are not required for the specificity of protein synthesis; instead they may be involved in the regulatory functions of tRNA molecules, since the variable regions could provide sites for specific recognition by other molecules.

Finally, tRNA is associated not only with the synthesis of polypeptide chains but also with that of polynucleotide chains. This synthesis is carried out by special enzymes such as reverse transcriptase, which was discovered a few years ago as a constituent of several tumor viruses. Reverse transcriptase synthesizes a strand of DNA from a template of single-strand RNA, a direction of information flow that is the reverse of the normal one. Surprisingly, a specific type of tRNA first binds to the enzyme and to the viral RNA and signals the synthesis of the DNA copy to begin. Why a tRNA serves this purpose is completely unknown.

It is probable that tRNA-like molecules were an essential component of the earliest living systems. Once these molecules were formed their unusual stability may have resulted in their gradually being utilized to serve purposes other than their main function in protein synthesis. Although the elucidation of the three-dimensional structure of tRNA has been an important step forward, a great deal remains to be learned about this versatile molecule and its many roles in the living cell.




**ORIENTATION OF BASES in the transfer-RNA molecule is shown in this schematic view. The polynucleotide backbone is reduced to a thin line, with the short boardlike structures representing unpaired bases and the longer boards representing base pairs. The letters refer to the molecular diagrams on page 60. Note the presence of tertiary interactions between three bases in the core region of the molecule below the corner of the *L*. Overall the molecule is composed of two stacked columns of bases at right angles to each other. The stacking interactions between the parallel bases in the interior of the molecule provide a major stabilizing force.**



# ERYTHROCYTES

## (Color them red, or whatever.)

### About how biological stains are "certified"

Though we believe with all our heart in free enterprise, we cheer for the Biological Stain Commission and all it stands for in inhibiting us from even hinting at intrinsic superiority of any given stain bearing our label, as long as it carries the Commission's label. Preference for our product would have to be founded on such secondary factors as emotional bias or, better yet, convenience in ordering along with other EASTMAN Organic Chemicals from your lab supplier.

The Commission originated under auspices of the National Research Council in 1921 to cope with vagaries in performance encountered when snatching a few drops for the microscopical lab from the torrents of industrial dyestuffs. The Commission enjoys the backing of 11 scientific societies. It sells its label to cooperating lab supply houses. The revenue maintains a certifying laboratory at the University of Rochester School of Medicine and Dentistry. The Commission's label signifies that:

1. It has tested a sample of the particular batch of dye and has a sample of that batch on permanent file.
2. The identity of the dye has been confirmed spectrophotometrically.
3. The dye content meets Commission specifications.
4. The dye has been tested to perform satisfactorily the tests indicated on the label.

Each label also contains a certification number corresponding to the batch tested. The label requests that the user report any unsatisfactory results to the Commission.

*The catalog "EASTMAN Biological Stains" (Kodak Publication No. JJ-281) is available on request from Dept. 412-L, Kodak, Rochester, N.Y. 14650. It gives synonyms and early literature references for the certified stains we offer. It also tells about other Eastman products for microtechnique, quite apart from photography.*

### . . . and about the economy of black-and-white for illustrating

Uncommunicated science is hardly science at all. Therefore some of the cost of operating a laboratory must go for communicating by photographic illustrations. Color punches through generally better than black-and-white, of course, particularly when vividness counts, as in engaging those who hadn't known they were interested. But when a shrinking dollar must be sliced too many ways, black-and-white prints communicate better than words alone.

The economy in black-and-white stems from mechanized processing at appropriate levels of production volume. To get a better handle on the economics involved so that more reliable answers may be given, we have devised a questionnaire on photographic printing and processing practices and experiences in the laboratory environment.

In compensation for the kindness and postage in sending for the questionnaire on black-and-white printmaking, completing it, and returning it, we will send your laboratory a portfolio of 14 rather magnificent posters in photomicrographic *color*. Each bears a few words to help the color grab visitors and young scholars who may not have realized they were interested in that sort of thing.

*Please write on your institution's letterhead indicating very briefly how you are involved in its photography. Mail to J. E. Brown, Dept. 412L-97, Kodak, Rochester, N.Y. 14650. Offer expires March 31, 1978.*

Kodak

# SCIENCE AND THE CITIZEN

## *The Coal Option*

It has been estimated that coal represents as much as 90 percent of the total fossil-fuel reserves of the U.S., yet it currently supplies only about 18 percent of the nation's energy needs. In view of the dwindling domestic supplies of oil and natural gas, the growing dependence on imported fossil fuels and the unlikelihood that alternative energy sources can contribute significantly to the U.S. energy budget over the next 25 years, attention has focused on coal as a major component of the nation's future energy strategy. In accordance with this view the Carter Administration has included in its National Energy Plan, now before Congress, a proposal that would double domestic coal production in the next decade or so: from 665 million tons in 1976 to 1.2 billion tons by 1985. What is the probability that the U.S. can achieve such an expansion in its annual rate of coal production?

According to a report to Congress prepared by the General Accounting Office, the prospects are not good. The book-length GAO study, titled *U.S. Coal Development — Promises, Uncertainties,* concludes that, "in fact, it will be very difficult to achieve one billion tons annually by 1985." The detailed analyses that support this conclusion are set forth in the report, which reviews "the status, prospects and major issues in U.S. coal development from the standpoints of demand, supply, production, transportation, environmental and socioeconomic impacts, and America's position in the world coal market."

The decline of coal from its preeminent position of 50 years ago, when it supplied 80 percent of the energy consumed in the U.S., has apparently been checked by the recent renewal of interest in its prospects. Nevertheless, the GAO report makes clear, "the renewed interest in coal as an energy source is a matter of necessity rather than choice. If it were strictly a matter of choice, coal's decline relative to other fuels would continue. Coal would not be chosen over oil and gas for several basic reasons: coal is mined rather than pumped and therefore is more dangerous and difficult to extract from the earth; it is bulkier and therefore more difficult to transport and to handle; and it is dirtier and therefore causes more pollution when burned." In fact, the best one can say about coal, it seems, is that "there is a lot of it," particularly in relation to the nation's domestic oil and gas resources.

Recent coal prices have nonetheless not been as attractive as those of other energy resources. Among the reasons for this situation the GAO lists: (1) uncertainties about environmental standards, (2) possible increases in capital and operating costs owing to environmental-control requirements, (3) transportation and storage problems, and (4) the relative cost advantages of nuclear power.

The Administration's energy plan deals with some of the constraints on increased coal use, but in the view of the GAO analysts it neglects others, principally those having to do with transportation and productivity. In arriving at its own version of the physical and economic limits of "the coal solution" the GAO is led to the conclusion that "so many interrelated elements would have to work to double coal production by 1985 that GAO does not believe it could happen; to name only two, mining-equipment manufacturers would have to fill orders promptly and mining companies must have the foresight and capital to be able to open new mines when the added output is needed."

During the period up to 1985, the report continues, "coal is not only supply-constrained but also demand-constrained, in the sense that utility and industrial users are not going to buy coal if they cannot use it. There are long lead times involved just in building and installing boilers at existing plants, not to mention the lead times involved in planning and building completely new coal-burning plants."

In the medium-term future (from 1985 to 2000), the GAO report adds, coal will continue to be demand-constrained. "The possibilities of direct substitution for oil and gas are very limited on an economy-wide basis. The prospect for indirect substitution by coal-generated electricity, while more promising, is limited too by economics and the current state of industrial and transportation technology."

Over the longer term (beyond the year 2000) it appears to the GAO analysts that coal will again be both supply-constrained (particularly with regard to low-sulfur and metallurgical coal) and demand-constrained. The very-long-term prospects for increased coal demand, they feel, "ride on the hope of coal gases and liquids becoming environmentally safe and economical."

There is no question, according to the GAO report, that "coal will supply a large part of the nation's energy future." The required "trade-offs" will be costly, however, "particularly in terms of human life and disease." If in the final analysis it is decided that the costs of coal use beyond a certain level are too high and increased oil importation is not a tenable alternative, then, the report states, the U.S. has only two major alternatives open to it between now and the year 2000: (1) to accelerate the expansion of conventional nuclear power so that nuclear-generated electricity substitutes for oil and gas wherever possible, or (2) to improve energy conservation, through both increased efficiency and decreased consumption.

## *Shifting Sources*

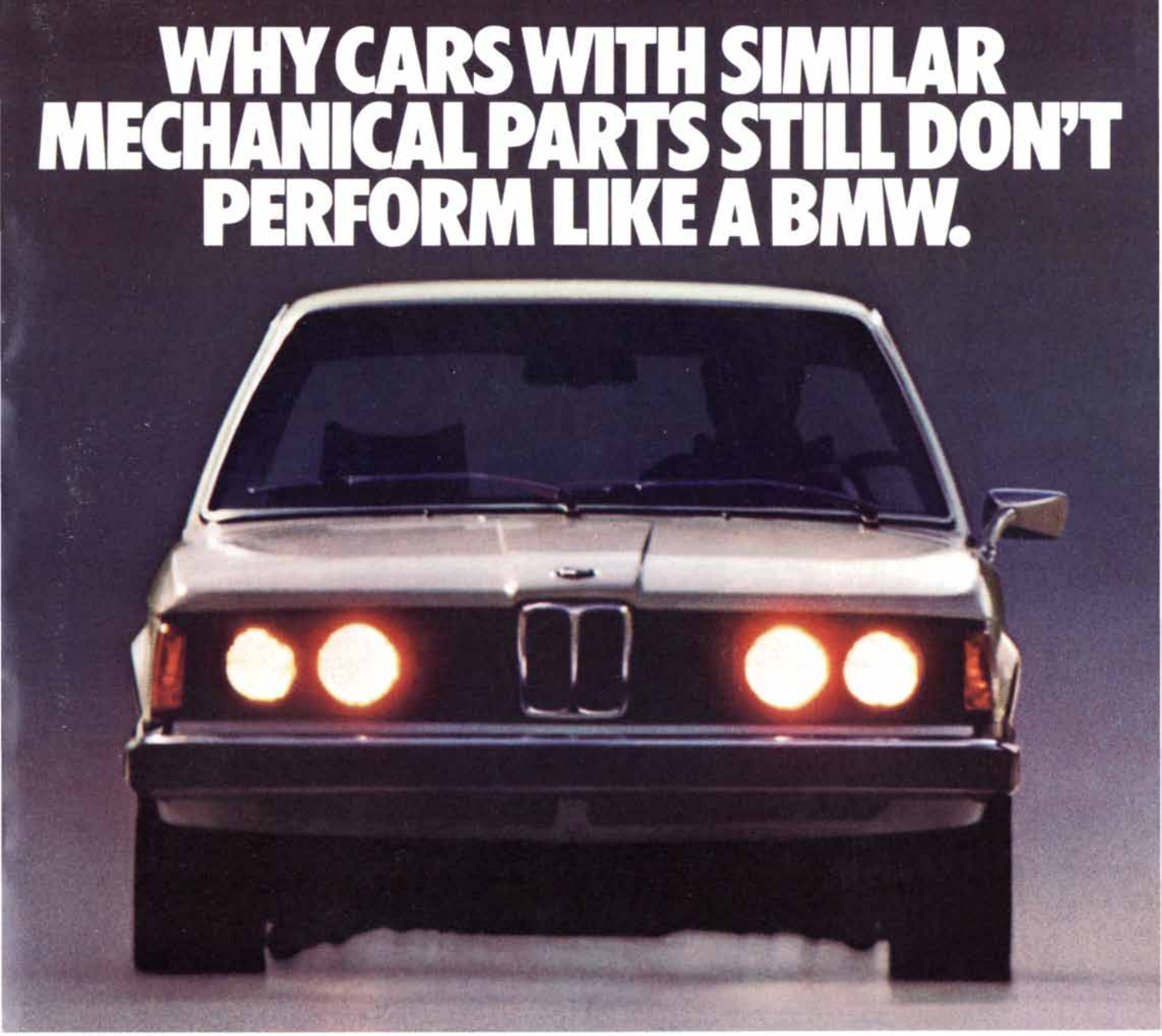
Nuclear power plants currently represent about 9 percent of the total electric-generating capacity of the U.S. During the first half of last year, however, nuclear plants actually produced 12 percent of the nation's electrical output, thus surpassing for the first time the contribution of hydroelectric power plants. According to figures from the Federal Power Commission, the leading sources of electricity for the first six months of 1977 were coal-fired plants, 476,574,723 megawatt-hours (46.1 percent); oil-fired plants, 183,715,160 megawatt-hours (17.8 percent); gas-fired plants, 136,587,365 megawatt-hours (13.2 percent); nuclear plants, 123,792,249 megawatt-hours (12 percent); hydroelectric plants, 110,371,137 megawatt-hours (10.7 percent); other sources, 2,032,366 megawatt-hours (.2 percent).

Meanwhile in the United Kingdom the latest annual report of the Central Electricity Generating Board, which serves England and Wales, indicates that on some days in recent months the nine "Magnox" nuclear power stations operated by the Central Generating Board generated more electricity than all the utility's 21 oil-fired plants combined. Nuclear energy currently accounts for approximately 10 percent of the electric-generating capacity and 13 percent of the actual electric output of the U.K.

## *HLA and Disease*

The individuality of tissue from different human beings is asserted primarily by the protein products of genes at four loci on the short arm of chromosome No. 6: the major histocompatibility complex, designated *HLA*. Genes *A, B* and *C* code for antigens inserted in the cell membrane, and gene *D* is apparently involved in the manufacture of antibodies. The four genes are highly polymorphic, that is, there are many different alleles, or alternate forms, of each; it is the resulting difference in *A, B* and *C* antigens that causes foreign tissue to be recognized as foreign and rejected by the immune system. The polymorphism of the complex, whose products in different individuals can be differentiated by laboratory tests, has provided human geneticists with an important tool for studying the relation of particular genes to particular diseases and the pattern in

# WHY CARS WITH SIMILAR MECHANICAL PARTS STILL DON'T PERFORM LIKE A BMW.



All expensive, imported automobiles feature an impressive list of sophisticated mechanical refinements.

Most have some sort of independent suspension system. Most have fuel-injected engines.

All are designed with more than a perfunctory nod to aerodynamics and functionality.

Yet, write the editors of Motor Trend magazine, "...once a knowledgeable and experienced driver has driven a BMW, any BMW, nothing else feels quite as good as it did before."

The explanation? Quite simple

The BMW 320i is far more than a collection of gears and axles and random parts.

It is a finely tuned, evolutionary machine. A practical sedan built by racing engineers and perfected in places like Le Mans, Monte Carlo and the Nürburgring, where precision is crucial and agility and durability are more than just matters of theoretical speculation.

PERFECTED ON THE RACE TRACK, NOT MERELY THE TEST TRACK.

While it is, of course, feasible to develop an acceptable automobile in the relative vacuum of the test track and the laboratory, it is virtually impossible to simulate the perfection demanded by motor racing.

When you press the 320i accelerator, the two-liter, K-Jetronic, fuel-injected engine—the same basic engine that powers a majority of the world's Formula Two race cars—responds without lag.

Its suspension, independent on all four wheels—with McPherson struts and coil springs in front, semi-trailing arms and coil springs in the rear—is quick and clean through the corners.

Its rack and pinion steering is sharp and accurate.

"All told," say the editors of Car and Driver magazine, "the 320i stands as eloquent rebuttal to all those who'd have us believe that small economical cars must be dull—and that automotive performance for the late 1970's is best achieved with decals..."

If the thought of owning such a car intrigues you, call us anytime, toll-free, at 800-243-6000 (Conn. 1-800-882-6500) and we'll arrange a thorough test drive for you.

**THE ULTIMATE DRIVING MACHINE.**

Bavarian Motor Works, Munich, Germany.

which diseases that have a genetic component are inherited. Recent results of studies associating the *HLA* complex with susceptibility to two diseases have appeared in *The New England Journal of Medicine,* along with editorials assessing progress in this field of investigation.

More than 100 diseases have been investigated in population studies for evidence of association with specific *HLA* antigens, and such evidence has been reported for more than half of them. As Leon E. Rosenberg and Kenneth K. Kidd of the Yale University School of Medicine and James V. Neel of the University of Michigan Medical School point out in their *New England Journal* editorials, such association can be explained in three ways. Particular *HLA* antigens could be directly implicated in the causation of the disease, the disease could be common in an ethnic subgroup of the study population in which particular *HLA* alleles are also common or the explanation could be "linkage disequilibrium": the gene that increases susceptibility to the disease could be genetically linked disproportionately often with specific alleles at the *HLA* loci. The last explanation is considered to be the most likely one for most of the associations. An association suggested by population studies can be confirmed by studies of affected families, and family studies can also indicate the means of inheritance.

Hemochromatosis, a disorder of iron metabolism characterized by the excessive deposition of iron in the tissues, has been associated with two specific *HLA* antigens: *A*3 and *B*14. In other words, population studies have shown that individuals carrying those particular antigens (and the alleles that code for them) are more likely than other individuals to have hemochromatosis. A report from four investigators at the Pontchaillou Hospital at Rennes in France provides support for that finding and for the recessive inheritance of susceptibility to hemochromatosis. Marcel Simon and his colleagues determined the haplotypes of siblings in families in which two or more siblings showed convincing symptoms of the disease. (An individual can have two different alleles at each locus, in this case the *HLA-A* and *HLA-B* loci; one allele is inherited from each parent. The alleles inherited from one parent constitute the haplotype and two haplotypes together constitute the genotype. Both haplotypes are expressed in the *HLA* genes: the genes are "codominant.")

The investigators found that a significant proportion of the affected siblings had both haplotypes in common, most often *A*3 and *B*14 but not always; very few of the unaffected siblings in these families had the identical genotype as an affected sibling. Such results imply that overt hemochromatosis depends on the presence of two specific homologous chromosomes—that its mode of transmission is recessive (as opposed to dominant transmission, in which one gene for a trait is enough for the expression of that trait). Such a conclusion is supported by studies of families with one affected sibling (in most cases unaffected members had different haplotypes) and by statistical analysis of a group of unrelated individuals with hemochromatosis. The authors suggest that *HLA* typing in families with hemochromatosis could detect the disease before symptoms appear and make preventive treatment possible.

One disease whose genetic basis has long been debated is diabetes mellitus. It now appears that the juvenile-onset form of the disease (which seems to be clinically and genetically distinct from the less serious adult-onset form) is caused by an autoimmune response to a viral infection, and that susceptibility to the response is influenced by a gene or genes close to the *HLA* loci. In an article in the *New England Journal* Pablo Rubinstein of the New York Blood Center and Nicole Suciu-Foca and John F. Nicholson of the Columbia University College of Physicians and Surgeons argue on the basis of family studies that a gene predisposing to juvenile diabetes is associated specifically with the *D* locus of the *HLA* complex and that inheritance is recessive—that two such genes are required for the disease to manifest itself. Neel points out in his editorial that such a conclusion should be considered tentative. One reason is the possible heterogeneity of the disease: the lack of insulin that characterizes diabetes may be caused by multiple genetic defects that have varying degrees of penetrance, or expression, and that vary in incidence within different populations.

## *Traumatic Birth*

In modern astronomy the sky is full of fireworks: supernovas flare and fade, neutron stars spin like pinwheels, binary-star systems erupt with X rays and possibly gamma rays. There is now evidence for an explosive event close to home. It appears that the formation of the solar system may have been triggered by the concussion of a supernova.

The evidence derives from small grains found in meteorites, the grains being made up of elements with an isotopic composition different from that of the other matter in the solar system. The grains can be interpreted as debris of a supernova explosion that could not have happened very long before the meteorites formed. The isotopic anomalies have been detected by two main groups of investigators, one led by Robert N. Clayton of the University of Chicago, the other by Gerald J. Wasserburg of the California Institute of Technology.

It has been apparent for some time that many supernovas in the galaxy must have preceded the creation of the solar system. Elements heavier than iron can be formed only under the extreme conditions that prevail in a supernova explosion, and the abundance of these heavy elements in the earth therefore implies that the solar system gathered up the ashes of many supernovas. It has been assumed that the material received from all the stellar explosions was thoroughly mixed in the cloud of dust and gas that ultimately gave rise to the solar system.

Several processes can lead to the segregation of one element from another in a developing star and planetary system. For example, because of differences in the density of the elements, the sun has accumulated most of the hydrogen in the solar system, whereas the earth is made up mainly of heavy elements. Isotopes of a single element, however, are very similar in all their chemical and physical properties; they are made up of atoms that have the same number of protons but different numbers of neutrons. The isotopes should therefore be distributed uniformly throughout the solar system. Wherever an element is found its various isotopes should be in the same ratio, whether their abundance is measured in earth rocks or in meteorites, or even if it is deduced from the solar spectrum.

The grains studied by Clayton and by Wasserburg are microscopic inclusions in carbonaceous chondrites, meteorites that are thought to represent the oldest solid matter in the solar system. Elements within the grains have been found to have anomalous ratios for isotopes of several elements. The Chicago group, for example, has detected ratios different from those found in earth rocks for magnesium, carbon, nitrogen, oxygen, neon and xenon. One notable anomaly is in the abundance of the oxygen isotope with an atomic mass of 16. In some samples as much as 5 percent of the oxygen is pure oxygen 16, with no admixture of the isotopes oxygen 17 and oxygen 18, which are usually present.

There is no known mechanism for converting one isotope into another under the conditions experienced by meteorites (except for radioactive isotopes, which decay spontaneously through a known sequence). Chemical fractionation might create small disturbances in isotopic ratios, but the discrepancy would be proportional to atomic mass, a relation that is not observed. Hence it is concluded that the grains may have had a different origin from the rest of the matter in the solar system.

Particularly strong support for the hypothesis that the grains came from a nearby supernova is provided by a study initiated by Typhoon Lee of Wasserburg's group of the isotope magnesium 26. He found large enrichments of the isotope in grains containing little magnesium overall but substantial amounts of aluminum. Magnesium 26 can be

# At C-E, we help the lights go on all over the world.

We help provide electricity. For homes, cities, industry, transportation, growth. In 77 countries, on six continents.

At Combustion Engineering, we build steam supply systems that electric utilities use to generate power. These include nuclear systems as well as those fueled by oil, gas or coal.

Significantly, since 1950, more than 645 coal-fired boilers of C-E design have been purchased around the world—representing a total capacity of 185 million kilowatts.

In fact, steam generators of C-E design produce about 40% of the thermal electric power in those areas of the world where statistics are available.

**We help the lights stay on.**

C-E Power Systems specialists are at work around the world engineering, building and maintaining power plants for utilities and industry. And C-E also helps convert power plants so they can burn much more abundant fuels, such as coal or solid waste.

**Our business is energy.**

For 65 years, we've been supplying equipment, technology and research to help produce energy, conserve it, squeeze more use out of it, even convert it to new forms.

For more information about our worldwide activities, write Combustion Engineering, Inc., Dept. 7008-34, 902 Long Ridge Road, Stamford, Connecticut 06902.

COMBUSTION ENGINEERING

The Energy Systems Company

produced by the decay of the radioactive isotope aluminum 26, and measurements of the amounts present suggest that this decay sequence was the source of the meteoritic magnesium. Aluminum 26 decays with a half-life of less than a million years, and so the supernova that produced it could not have exploded more than a few million years before the meteorites condensed.

The supernova hypothesis is attractive on other grounds. It is not well understood how a cloud of gas and dust can spontaneously condense to form a star of the sun's size; a critical density must be reached before gravitational attraction can overcome the internal pressure of the gas. The shock wave from a nearby supernova might provide the necessary initial compression. Thus it would not be by accident that the solar system formed in the neighborhood of a supernova; the solar system may be here because of the supernova.

## *The Cytoskeleton*

The living cell used to be regarded as being largely a tiny bag filled with fluid in which enzymes and other substances interacted freely. Over the years the interior of the cell was perceived to be much more rigorously organized into organelles and membranous compartments. Now it is becoming increasingly apparent that cells also possess an intricate internal skeleton made up of networks of fibrous protein. Although various types of subcellular fibers had been previously characterized with the electron microscope, their overall pattern was clearly revealed for the first time when specific antibodies were utilized to label the fibers with fluorescent markers so that they could be visualized with the light microscope. Such experiments showed that the three major types of subcellular fibers—the microtubules, the microfilaments and the intermediate fibers—are organized into networks that extend throughout the cell's interior.

The microtubules, about 250 angstroms in diameter, radiate outward from the cell nucleus along straight or curved paths and terminate near the cell surface. They help to maintain the shape of the cell and also guide the movement of organelles such as mitochondria and lysosomes from one part of the cell to another. At the beginning of cell division the microtubular system is organized into the mitotic spindle, which is involved in separating the duplicated chromosomes as the cell divides. The microfilaments, about 60 angstroms in diameter, are usually associated with the inside of the cell membrane. Composed primarily of the contractile protein actin, together with certain other proteins also found in muscle fibers, they are organized into a meshwork or into bundles, both of which are important for cell motility. The intermediate filaments are about 100 angstroms in diameter and form a wavy, interwoven network inside the cell that is distinct from the other fiber systems. All three types of fibers are dynamic structures that are constantly being assembled, disassembled and rearranged in actively moving cells.

In order to examine the structure of the cytoplasmic matrix in greater detail, Keith R. Porter and his colleagues at the University of Colorado at Boulder made use of one of the two high-voltage electron microscopes currently available for biological research in the U.S. This instrument accelerates electrons across a potential drop of a million volts, giving them sufficient kinetic energy to penetrate very flat intact cells. The resulting image is analogous to an X-ray plate, and it reveals the cell's internal structure without the cell's having to be sliced into thin sections, as must be done for electron-microscope examination at lower voltages. As a result one can visualize structures such as thin fibers that would be difficult or impossible to identify in section.

When Porter examined cultured human fetal lung cells (which are extremely flat) under the high-voltage electron microscope, he found that the entire cytoplasm of the cell was filled with a lacy network of fibers that had not been seen before: a fourth intracellular fiber system. The individual fibers were thinner than microfilaments—about 30 to 60 angstroms in diameter. Because the fiber network reminded Porter of the trabecular structure of spongy bone, he named it the "microtrabecular system." The lattice appears to provide a kind of intracellular scaffolding to which the various organelles, as well as the three major fiber systems, are anchored. Porter believes that it functions as a kind of cellular musculature, constantly redistributing and reorienting the organelles as the cell goes about its activities, and that it coordinates the functions of the three major fiber systems. The movement of cellular components through the lattice appears to be achieved by the extension and contraction of the individual microtrabecular filaments, with energy being stored in the extended state.

## *A Protoscientific Revolution*

When did philosophers begin to conceive of the universe as something natural that could be comprehended by systematic, rational investigation instead of something supernatural to be regarded only with passive awe? Although the appearance of rational thinking is commonly associated with the Copernican revolution and the Renaissance, in a recent article in the historical journal *Isis* Tina Stiefel discusses evidence from the early 12th century that, "four centuries before the beginning of modern science, a small group of Western Europeans... devised a concept of science as a separate discipline and went some distance toward formulating a reliable methodology for its practice."

The principal evidence lies in the writings of several early 12th-century cosmologists, notably Adelard of Bath, Thierry of Chartres and William of Conches. All were students of Plato, of works by Arab authors and of Aristotle's logic (which before A.D. 1150 was the only part of the Aristotelian scientific corpus that was available in Europe). "It has perhaps not been sufficiently realized to what extent practice in Aristotelian logic transformed men's belief in their rational faculty," Stiefel writes. "Deeply imbued with confidence in their training in the new dialectic, [the 12th-century cosmologists] were determined to extend this trust to include the power of human reason to penetrate nature's secrets."

The 12th-century cosmologists firmly expressed their belief that the universe was basically rational. Adelard of Bath wrote in his book *Quaestiones naturales,* "We must assume that all nature is based on a sure and logical foundation"; everything that exists has a logical reason for existing and man, part of creation, is equipped to understand those reasons. That understanding, however, can only come from cultivating and training one's mind in a logical way. William of Conches attempted to define science as that part of knowledge based on corporeal things, as opposed to metaphysics and theology, which were concerned with questions of essence and substance not visible to the eye. Stiefel observes: "To define science was less easy than we might imagine, and in the face of a commonly felt reluctance to fix limits upon the domain of theology, it took some courage to hazard a definition at all."

These 12th-century cosmologists also grasped the concept of a theoretical scientific explanation; although they did not completely formulate a systematic methodology for arriving at such explanation, they did recognize its component parts. Adelard of Bath and Thierry of Chartres advocated empirical techniques for studying the universe; nature was to be studied by means of a rational examination of evidence presented by the senses. John of Salisbury, a student of William of Conches, discussed in his *Metalogicon* the function of inductive reasoning in formulating hypotheses and the role probability plays in descriptions of natural events. Adelard and others were confident of the power of mathematical reasoning in deducing knowledge about nature.

The cosmologists also thought it was essential for the natural philosopher to approach scientific study in a specific frame of mind: he must strive to achieve a state of detached objectivity in order

to free his capacity for constructing useful hypotheses from the tyranny of unquestioned assumptions and accepted opinions about nature. In Stiefel's view, "the notion of making a habit of methodical doubt is perhaps the most sophisticated of any of the ideas... for it implies a dangerous liberty of the intellect."

The conceptual revolution represented by the work of Adelard, Thierry and William has been largely overlooked, according to Stiefel, because in the 12th century it had a very limited influence. Its promulgation coincided with the beginning of the Scholastic movement, which was concerned with applying Aristotelian logic to Christian beliefs. Overshadowed by scholasticism, the work of the protoscientists was ignored or treated as heresy, and its proponents endured ridicule and some persecution. In the following century, however, these concepts influenced such scholars as Robert Grosseteste and Roger Bacon, who had a more direct influence on the modern scientific revolution.

Stiefel concludes: "That the cosmologists went some way in the direction of formulating a new discipline and mapped out workable methods of approach to science is an accomplishment deserving more attention by historians. What has been largely overlooked is the imaginative leap that this work required, for it is easy to miss the sheer nerve it took to make that leap: from a cowed acceptance of nature as necessarily inscrutable to a spirit of inquiry, at once optimistic and trusting.... For their courage and originality the cosmologists deserve notice, and their efforts to construct the discipline of science are worthy of attention and respect."

## *Getting the Axe*

What gave the Neolithic period its name was the recognition that at that stage in human history a new kind of tool had emerged: a stone axe with a smooth or even polished surface. Many thousands of such axes have come to light all over the Old World at sites where excavated materials show that the people who used the axes were not nomadic hunters but were growing crops, tending animals and leading settled lives. The areas where the axes are found, however, are often devoid of the kinds of stone the axes are made of. A study of some 3,300 stone axes from Neolithic sites in Britain is now revealing, among other things, the remarkable lengths to which the island's early farmers went to get good stone and bring it home.

The study had its beginnings in the late 1920's, when an enthusiastic amateur archaeologist, Alexander Keiller, the heir to a marmalade fortune, excavated Windmill Hill, a major Neolithic site in Wiltshire, and unearthed large numbers of stone axes. Determined to learn more about the tools, Keiller turned for help to H. H. Thomas, the geologist who had traced the "blue" stones of Stonehenge to their quarry 140 miles away in Wales. Thomas began by removing thin sections suitable for petrological analysis from Keiller's axes, and what started as a private study developed over the next few years into a nationwide project. A succession of geologists analyzed stone axes from controlled excavations throughout Britain for decade after decade, until now, some 40 years after the work began, the petrological collection is representative of all parts of the country except Cumberland and Northumberland.

At a recent symposium convened at the University of Nottingham by the Council for British Archaeology it was noted that whereas British axes could be traced to 20 separate quarry areas, these sources of stone were not all equally popular. More than 900 axes, or 27 percent of those analyzed so far, are made of a hornstone available only from outcroppings close to Scafell Pike, a mountain in the Lake District of northwestern England that is the highest in the country. The next-largest group of axes, representing more than 9 percent of the total, are of a greenstone found only near Penzance in Cornwall. A further 9 percent are of a stone found only at Graig Llwyd in northern Wales. None of the 17 other quarries provided stone for more than 2 percent of the collection.

The distribution of the axes from Cornwall and the Lake District is anomalous. Greenstone axes are common in western Cornwall, but even more of them have been found in Essex, some 300 miles to the northeast, and they are rather uncommon at coastal sites in between. Commenting on this point, one of the project's leaders, William A. Cummins of the University of Nottingham, believes the greenstone axes moved by sea from Lands End at the southwestern tip of Britain to the Thames estuary on the eastern side of the island. As for the hornstone axes quarried near 3,210-foot Scafell Pike, they are found in the greatest number along the Humber River some 100 miles across the island; more than 50 percent of all the hornstone axes come from sites within a 70-mile radius of the mouth of the river. Here there seems to be no question of any movement by sea; such a voyage would have required a virtual circumnavigation of Britain. Discussing the question in *Current Archaeology,* one of the editors of the journal, Andrew Selkirk, speculatively visualizes Neolithic youths in Yorkshire, ready for initiation into adulthood, setting off overland for the Lake District on a ritual quest, determined to collect in the shadow of Scafell Pike stone for the axes symbolic of man's estate.

# The Johnsons thought their fuel bills were going through the roof.

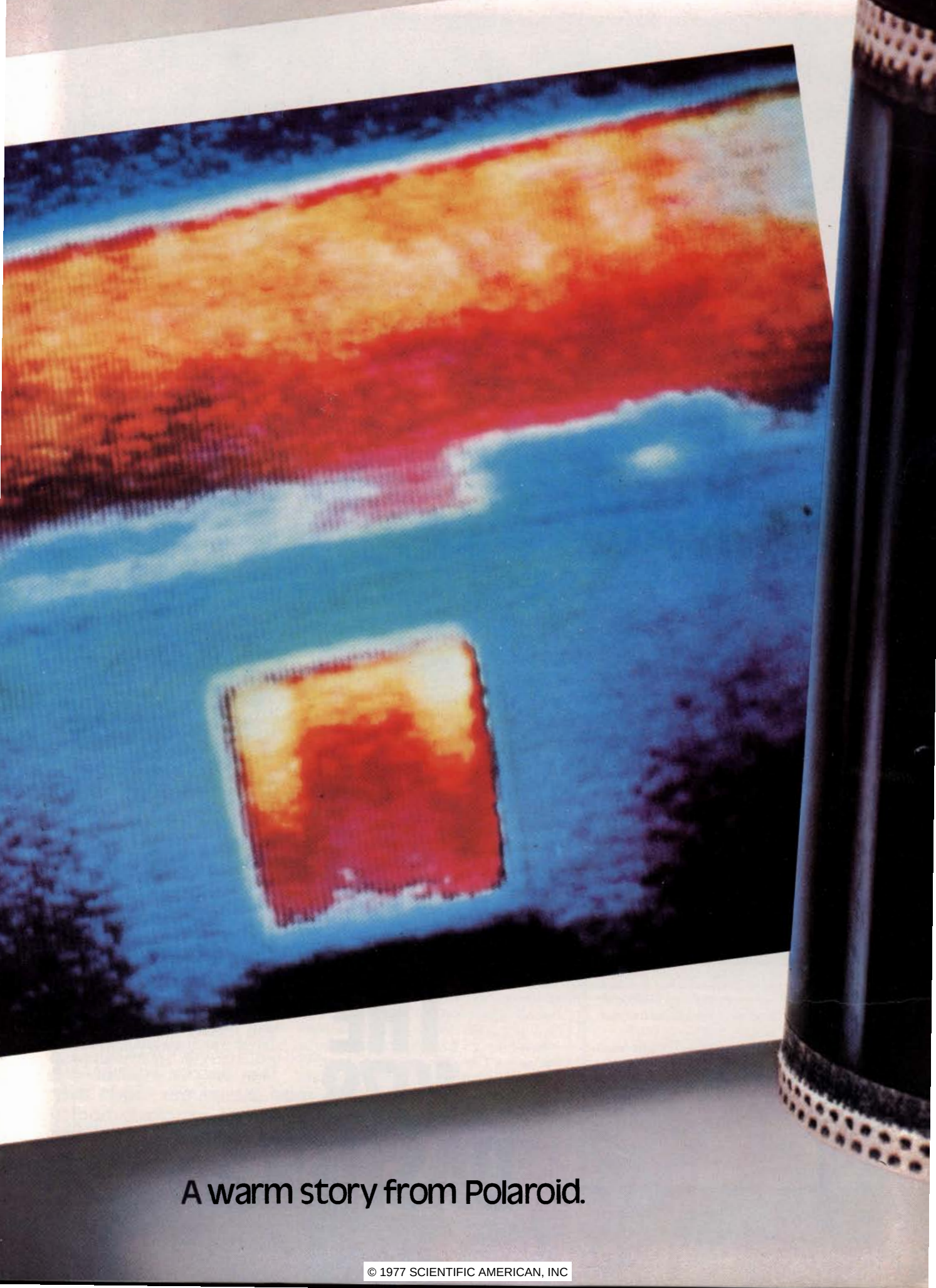
# They weren't too far wrong.

This Polaroid instant picture shows the Johnsons' beloved homestead caught in the act of wasting their hard-earned heating dollars. More specifically, it shows a dramatic heat loss through the roof. And significant losses through the walls and windows. It also shows the Johnsons exactly where to put extra insulation to save money on their fuel bills. (It even tells them where extra insulation *isn't* needed.)

This helpful photograph was shot from an infrared scanner display during a survey conducted by CertainTeed Corporation. ( They're one of a growing number of companies in the insulation field that make use of this technique.) The scanner translates the exterior heat values of a building into a range of colors (warm: white through red; cool: green through black) which pinpoint heat loss areas, and which form the basis for extremely accurate temperature readings. It's important to make a faithful, long-lasting record of these colors for future reference by the building owner. And it's easy. The survey operator simply records the image from the infrared scanner on Type 668 Polacolor 2 (one of Polaroid's 15 professional instant films). And since he gets his finished print in seconds, he can check its color fidelity against the scanner before he leaves the site.

This photograph may look odd to some people, but the Johnsons think it's the best picture ever taken of their home.

"Polaroid" and "Polacolor"®



A warm story from Polaroid.

†Based on 1977 EPA Gas Mileage Guide. 1978 data not available at press time.
*1977 Mercedes Benz acceleration (0-50 mph) from Car and Driver, March 1977. Comparison based on std. equipment cars. © Volkswagen of America.

# THE 1978 DASHER.

Imagine a Volkswagen limousine. Got it? Good!

You have now grasped the concept of the VW Dasher.

The Dasher Sedans are our most sumptuous cars; they outclass some very classy machines in some very surprising ways.

If you own a Mercedes-Benz 280E, you will be depressed to discover that

the Dasher 2-door Sedan not only goes from 0 to 50 miles per hour quicker than the Mercedes,* but carries more in its trunk.†

If you're about to spring for a Rolls-Royce, stop. The Dasher holds more in its trunk than the Rolls, too.†

Fine as they may be, neither the Mercedes nor the Rolls has front-wheel drive. The Dasher does, and it makes all the difference in poor driving conditions.

The Mercedes, the Rolls and the Dasher do have their similarities.

All 3 have dignified interiors, with handsome, thoughtful appointments like reclining bucket seats, remote control outside mirrors and quartz electric clocks.

But only the Dasher has a VW right there up front.

# VOLKSWAGEN DOES IT AGAIN

# The Structure of the Interstellar Medium

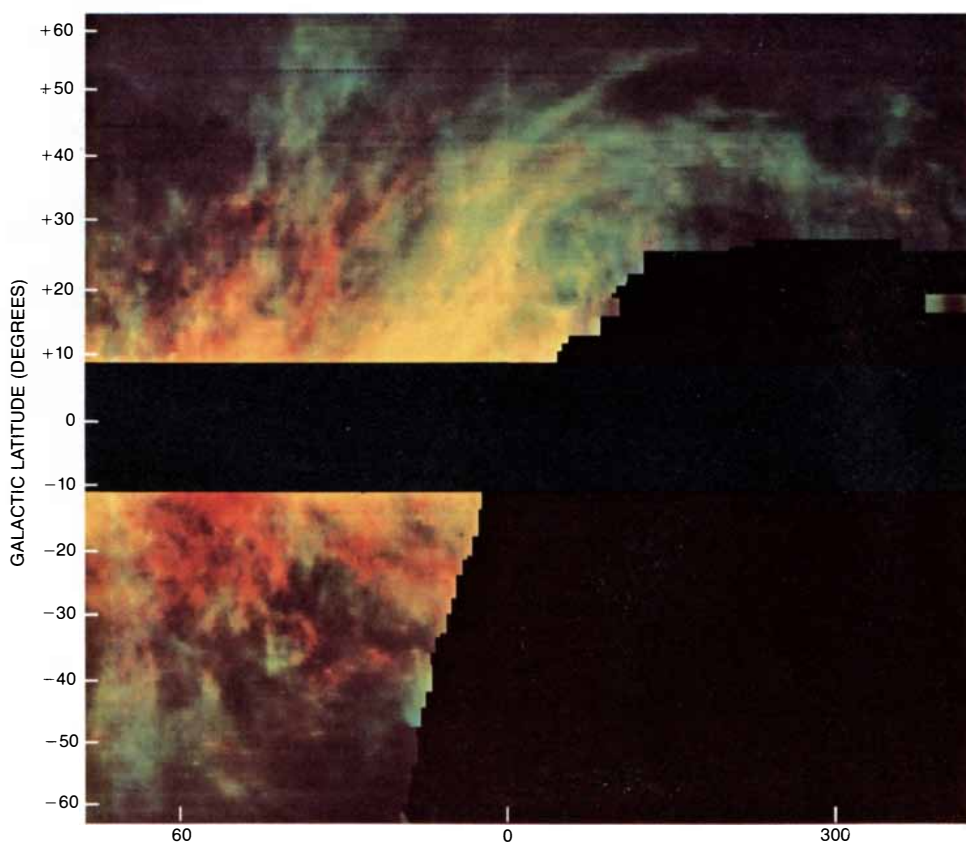
*It used to be thought that the gas between the stars probably formed clouds. A pictorial analysis of radio signals from the gas now indicates that it is largely marshaled in huge shells*

by Carl Heiles

Half a century ago spectroscopic observations of distant stars revealed that their light is reddened because it travels through gas and dust permeating interstellar space. The discovery of this interstellar material raised a number of questions that still remain largely unanswered. One of the most basic was: What is the structure of the interstellar medium? In a few regions of space near very hot stars the atoms of the interstellar gas are ionized, or stripped of one or more of their electrons, and they glow with a characteristic light of their own. The first four decades of observations of such ionized regions suggested that the interstellar material was not distributed uniformly through space but was clumped into large, roughly spherical clouds scattered randomly throughout the galaxy. These observations gave rise to a theoretical model of the interstellar medium known as the cloud model. According to this model, the clouds of interstellar gas were about 14 parsecs in diameter and an average of some 30 parsecs apart. (A parsec is 3.26 light-years.)

The great bulk of the interstellar material, however, is far too cold to emit energy at the wavelengths of visible light, and thus it remains invisible to light-gathering telescopes. Over the past decade or so astronomers have been able to extend their observations of the interstellar medium to other regions of the spectrum, including the X-ray, the ultraviolet, the infrared and the radio regions. These observations, particularly detailed maps of the distribution of atomic hydrogen in space compiled at the radio wavelength of 21 centimeters, have radically altered our views of the nature and the structure of the gas between the stars.

A hydrogen atom consists of a proton and an electron, each of which is spinning on its own axis. The spins may be parallel to each other or they may be antiparallel. If they are parallel, the electron may flip so that its spin is antiparallel to the spin of the proton. When the electron's spin flips, the atom emits radiation at the frequency of 1,420.4057 megahertz, which corresponds to a radio wavelength of about 21 centimeters. Hydrogen, the most abundant element in the universe, constitutes some 70 percent of the mass of the interstellar gas. More than half of the hydrogen is in the form of molecules ($H_2$) or is ionized so that the proton is separated from the electron. Molecular hydrogen and ionized hydrogen, however, are found in only a tiny fraction of the volume of




**SYNTHETIC PHOTOGRAPH OF THE INTERSTELLAR MEDIUM reveals that the gas of the medium loops through interstellar space in long filaments. The picture was artificially synthesized from data obtained at the radio wavelength of 21 centimeters with the 26-meter radio telescope at the Hat Creek Radio Observatory of the University of California. The author and his colleagues surveyed the interstellar medium in the entire sky visible from California, excluding a broad strip centered on the plane of the galaxy where there is so much gas that individual features cannot be distinguished. The large black parabola at the left is the portion of southern sky that is invisible from the latitude of the observatory; the small black rectangle**

interstellar space where the medium is relatively dense. The hydrogen that permeates the rest of the volume is in the form of individual atoms.

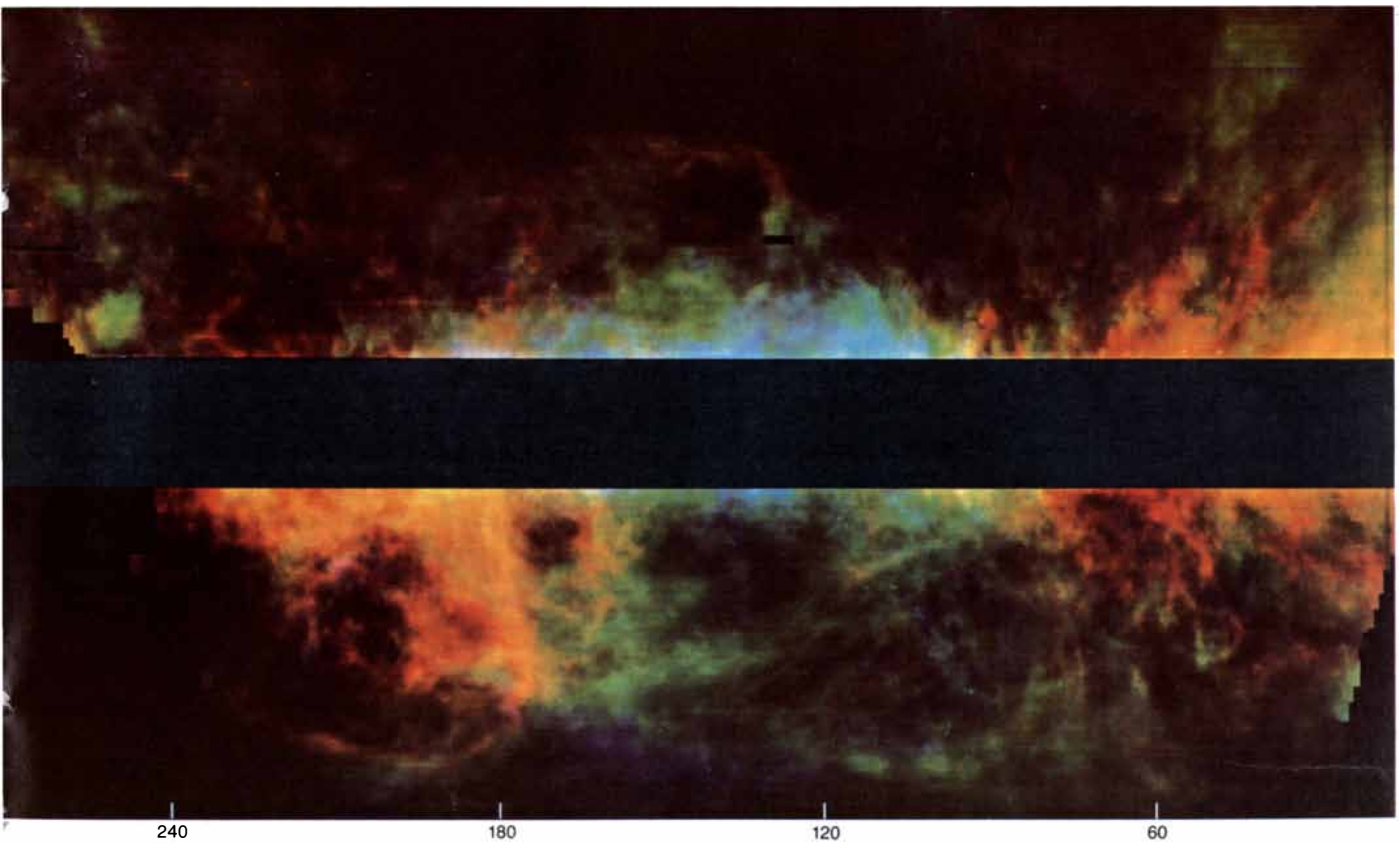
Observations of atomic hydrogen at radio wavelengths have a crucial advantage over observations at visible wavelengths. At visible wavelengths the interstellar hydrogen is studied by observing the way the gas absorbs starlight; hence observations of the gas at those wavelengths are limited to regions of the sky where there is a suitable star behind the gas. At radio wavelengths the interstellar hydrogen is studied by observing the radiation emitted by the gas itself; no background source of radiation is needed. Therefore observations at radio wavelengths can be made in any region, and they can reveal the distribution of the gas over the entire sky.

## Early Sky Surveys

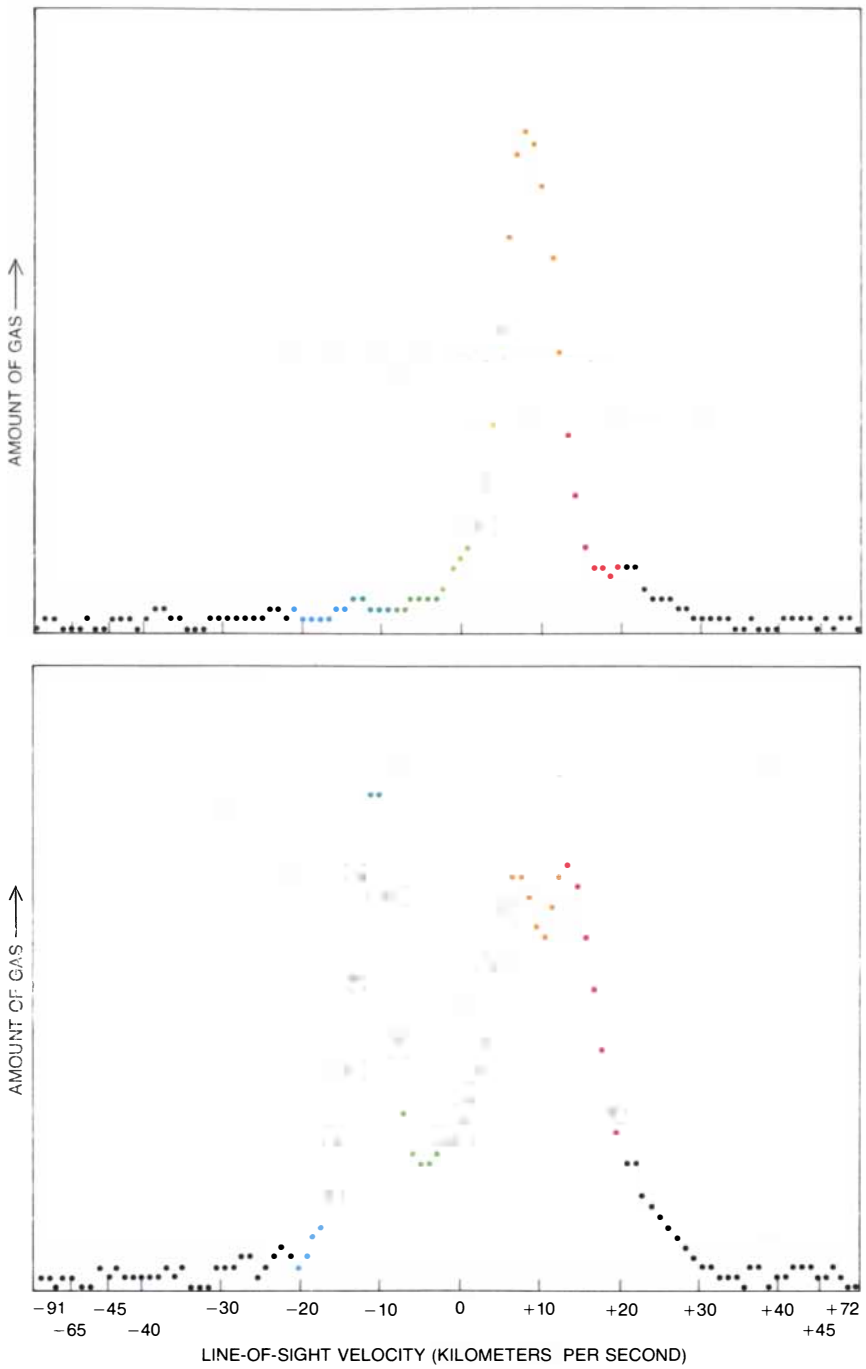
The first attempts to map the distribution of hydrogen were confined to observations of the gas in the central plane of the galaxy. It soon became apparent that there was so much gas in the galactic plane that the gas in any one structure could not be distinguished from all the other gas along the line of sight. Later surveys concentrated more on regions away from the galactic plane, mapping areas of the sky that were between 10 and 20 degrees across. According to the cloud model of the interstellar medium, in an area 10 to 20 degrees in angular diameter dozens or perhaps hundreds of gas clouds should have been discernible. The first surveys of the interstellar medium away from the galactic plane, however, revealed a surprising thing: the gaseous structures observed were so large in angular scale that they extended beyond the borders of the areas surveyed. Although there also seemed to be features rather smaller in scale, they appeared to be contained within the larger structures. It was clear that the smaller features could not be understood until the larger structures were.

Since the full angular extent of the large structures was not known, it was evident that larger areas of the sky would have to be mapped. Moreover, astronomy is by its very nature a statistical discipline: since astronomers cannot modify conditions in space in order to conduct experiments on how celestial systems behave, they must be content to observe objects in the variety of conditions already provided by nature. To observe a statistically significant sample of the large structures in the interstellar medium, one would have to map very large areas indeed. The choice my colleagues at the University of California at Berkeley and I made was to survey the entire sky, or at least as much of it as we could observe from the university's Hat Creek Radio Observatory, 170 miles north of Sacramento near Lassen Volcanic National Park.

With a radio telescope 26 meters in diameter we surveyed the sky visible from Hat Creek, excluding a strip 20 degrees across centered on the galactic plane where there is so much gas that individual structures cannot be distinguished. To cover so much of the sky it was necessary to point the telescope at some 140,000 distinct positions. The sensitivity of the observations increases with the amount of time devoted to each position; we had to spend about two minutes per position in order to be able




**near the galactic longitude of 120 degrees and the galactic latitude of +30 degrees is the north celestial pole, where the telescope cannot point owing to details of its design. Note that a portion of the galactic longitude repeats itself, so that the continuity of structural features across the edges of the map can be seen. The brightness of each filament in the synthetic photograph represents the amount of gas in that region. The color of each feature represents the velocity of the gas and the direction in which it is moving. There is a smooth progression of color through the spectrum with velocity: red represents gas receding from the solar system at 20 kilometers per second, yellow-green represents gas having no velocity in the line of sight, and blue represents gas approaching the solar system at 20 kilometers per second. Intermediate colors represent intermediate velocities. Near the plane of the galaxy it can be seen that in general the color of the gas alternates from blue to red with galactic longitude. The alternation in color results from the fact that the galaxy does not rotate as a solid disk.**




**VELOCITY PROFILES of the interstellar medium were measured with the Hat Creek radio telescope at each of 140,000 positions in the sky. Each profile is a graph of the amount of gas moving at different discrete velocities in the observer's line of sight. All the profiles were combined to yield the synthetic photographs. Positive velocities indicate gas receding from the solar system; negative velocities indicate gas approaching it. Each dot represents a measurement of the amount of gas in each of 100 narrow velocity intervals. Because very little gas in the interstellar medium is moving at very high velocity with respect to the solar system, for velocities higher than plus or minus 45 kilometers per second the width of the velocity interval is larger. The shape of the velocity profile of the interstellar gas is unique for every position in the sky. In constructing the color synthetic photograph it was imagined that the wavelength of visible light increased from left to right across the radio velocity profile, with pure blue at minus 20 kilometers per second and pure red at plus 20 kilometers per second. The velocity profile at the top was measured for a small region near the galactic longitude of 181.7 degrees and the galactic latitude of −30.4 degrees. At that position most of the gas is receding from the solar system at nearly the same velocity. The velocity profile is narrow and is represented by a color that is relatively pure; in the synthetic photograph on the preceding pages the region is reddish orange. The velocity profile at the bottom was measured for a small region near the galactic longitude of 205.7 degrees and the galactic latitude of −22.6 degrees. At that position in the sky some gas is approaching the solar system (*spike at left*) and somewhat more gas is receding from it (*peaks at right*). In the synthetic photograph on the preceding pages the region is whitish red.**

to obtain data at the required sensitivity. The observations took a total of seven months of telescope time, which was spread over an interval of two years.

## A Wealth of New Data

At each position we gathered information about both the total amount of the gas and its state of motion. Although each atom of hydrogen is emitting energy at the frequency of 1,420.4057 megahertz, the frequency at which the radiation is received by the telescope is Doppler-shifted by the motion of the gas with respect to the observer. The amount by which the frequency is shifted depends only on the direction of motion and velocity of the gas in the line of sight: gas moving away from the observer is seen at a lower frequency and gas moving toward the observer is seen at a higher one. By convention a source receding from the solar system is given a positive velocity and a source approaching the solar system is given a negative velocity.

We measured the intensity of the radiation from the gas at 100 separate frequencies, which corresponded to 100 separate velocity intervals. The amount of gas moving at each velocity is proportional to the intensity of the radiation from the gas moving at that velocity. In many instances there is more than one individual gas structure along the line of sight, with the structures located at different distances from the solar system. Usually each structure will have a different Doppler shift and therefore will appear in a different velocity interval.

Any survey of the entire sky generates large amounts of data, which is a problem in itself. In our case the problem was particularly acute. Since we measured the amount of gas in 100 velocity intervals at each of 140,000 positions in the sky, we obtained a total of 14 million data points, each one located in a three-dimensional space of galactic longitude, galactic latitude and velocity. It is a real problem to organize so many data points in such a way that they can be readily interpreted by the human brain. The standard format for presenting data in radio astronomy is a map in which the amount of gas can be represented by contours of constant intensity in exactly the same way that on a topographic map the heights of mountains are represented by contours of constant elevation. Initially we compiled such a map for our data, but on the map the large-scale structures we sought to study were obscured by a tangled maze of contour lines for the smaller features located within them. Clearly a better means of presenting the information was needed.

Astronomers observing the sky at visible wavelengths routinely gather information about large objects that contain small-scale structures by recording them on photographic film. What my

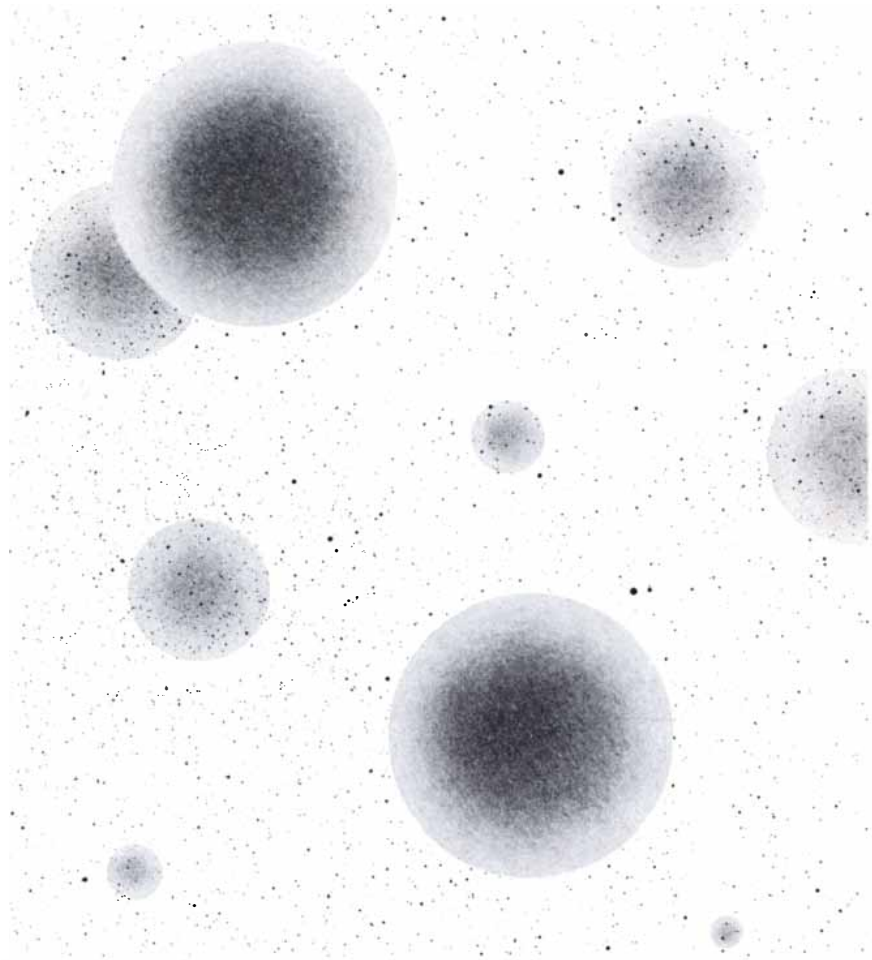
colleagues and I needed was a machine that could synthesize a "photograph" from the digital data we had gathered and stored on magnetic tape. Fortunately such a machine had been developed for other purposes in the department of astronomy at Princeton University. Furthermore, several Princeton astronomers are also working on the structure of the interstellar medium. Collaboration followed naturally.

## Synthesizing Radio Photographs

Working with Edward B. Jenkins of Princeton, we first ignored the information in the individual velocity intervals and simply made a synthetic photograph of the total amount of gas contained in the large velocity interval between minus 20 kilometers per second and plus 20 kilometers per second. Nearly all the gas is contained within that large interval [*see illustration on next two pages*]. The first thing we noticed in the synthetic photograph was that the gas tends to be organized not in individual clouds but in long filamentary structures. Many of the filaments are extremely long in their angular dimensions, some of them extending a quarter of the way across the sky. One cannot determine their actual length in parsecs, however, without knowing how far they are from the solar system.

The radio observations of the gas alone provide no information about the distance of the filaments from the solar system. To obtain that information it is necessary to relate observations of the gas to observations of stars, whose distance can be independently determined. The most straightforward and direct way to relate the two types of observation is to study absorption lines in the spectrum of a star produced by the interstellar gas lying between the star and the solar system. The absorption lines result from the fact that the cool interstellar gas absorbs radiation from the hotter star behind it at the same wavelengths the gas itself would emit if it were heated to a high enough temperature.

Jenkins and Blair D. Savage of the University of Wisconsin have studied the absorption lines in the spectra of stars produced by the interstellar atomic hydrogen. Because both the absorption lines superposed on the star's spectrum and the radio emission lines at 21 centimeters are produced by atomic hydrogen, a comparison of the intensity of the absorption lines and that of the emission lines will reveal whether all the hydrogen is in front of the star or whether some of it is behind the star, where it would not affect the star's spectrum. Then by determining the distance to the star by standard astronomical techniques one can be sure that the gas in front of the star can be no farther from the solar system than the star is.



**CLOUD MODEL of the structure of the interstellar medium was proposed on the basis of observations at visible wavelengths. Model proposed that gas was clumped into individual clouds (*gray*) seven parsecs in radius and distributed throughout space an average of 30 parsecs apart. (One parsec is 3.26 light-years.) The average distance between stars (*black dots*) is .4 parsec.**

The absorption lines of interstellar atomic hydrogen are all located in the ultraviolet region of the spectrum, which cannot be observed from the ground. A telescope aboard the *Copernicus* satellite in orbit around the earth, however, is well suited for observing the proper wavelengths, although it is capable of observing only hot stars that are very bright at ultraviolet wavelengths. Such stars are rare, but those that exist provide crucial information that is otherwise unobtainable.

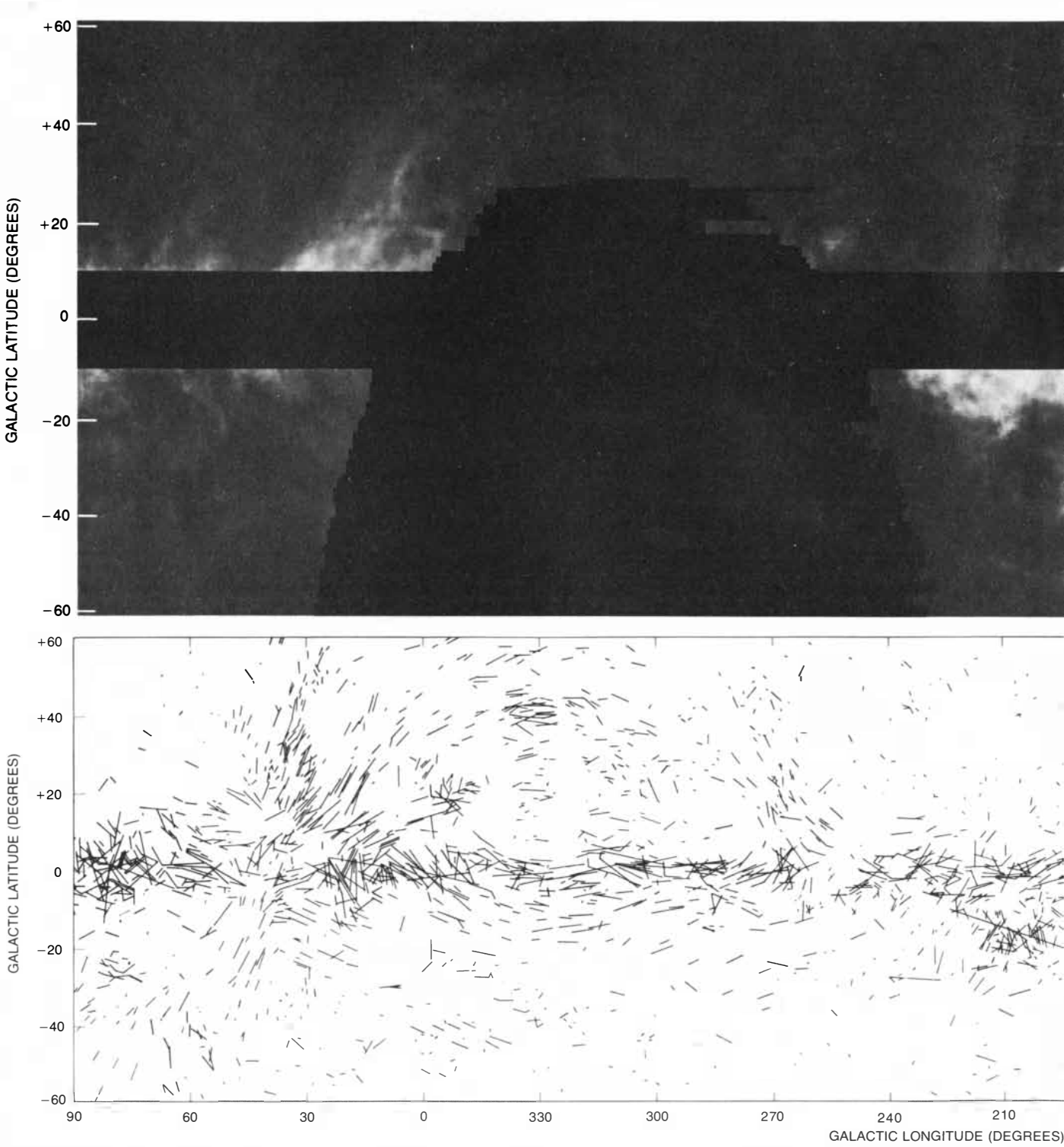
One area of the sky populated with a number of hot stars is a large region nearly 30 degrees across in the constellation Orion that also contains the Great Nebula in Orion. The region is near the galactic longitude of 210 degrees and the galactic latitude of $-20$ degrees. The stars in that region are all about 500 parsecs from the solar system. The ultraviolet spectra of the stars show that in this region there is little gas in front of the stars. The radio-wavelength observations, however, show that in the same region there is plenty of interstellar gas, and that the gas is all part of one large structural feature. These observations mean that there is a fairly large area of the sky in which there is little gas closer than 500 parsecs to the solar system. The observations contradict the prediction of the cloud model that the average distance between interstellar clouds is 30 parsecs.

## The Galactic Magnetic Field

A less direct way to associate observations of stars with observations of the interstellar medium was inadvertently discovered when we made the first synthetic photograph. Some filaments of the gas appeared to be aligned with the lines of force in the magnetic field of the galaxy. The magnetic field aligns elongated grains of interstellar dust, which are associated with the gas, in such a way that the dust acts as a filter that polarizes the starlight passing through it. It is believed the direction of the polarization is parallel to the lines of force in the magnetic field; therefore a measurement of the direction in which starlight is polarized is in effect a measurement of the direction of the galactic magnetic field, averaged over the indi-

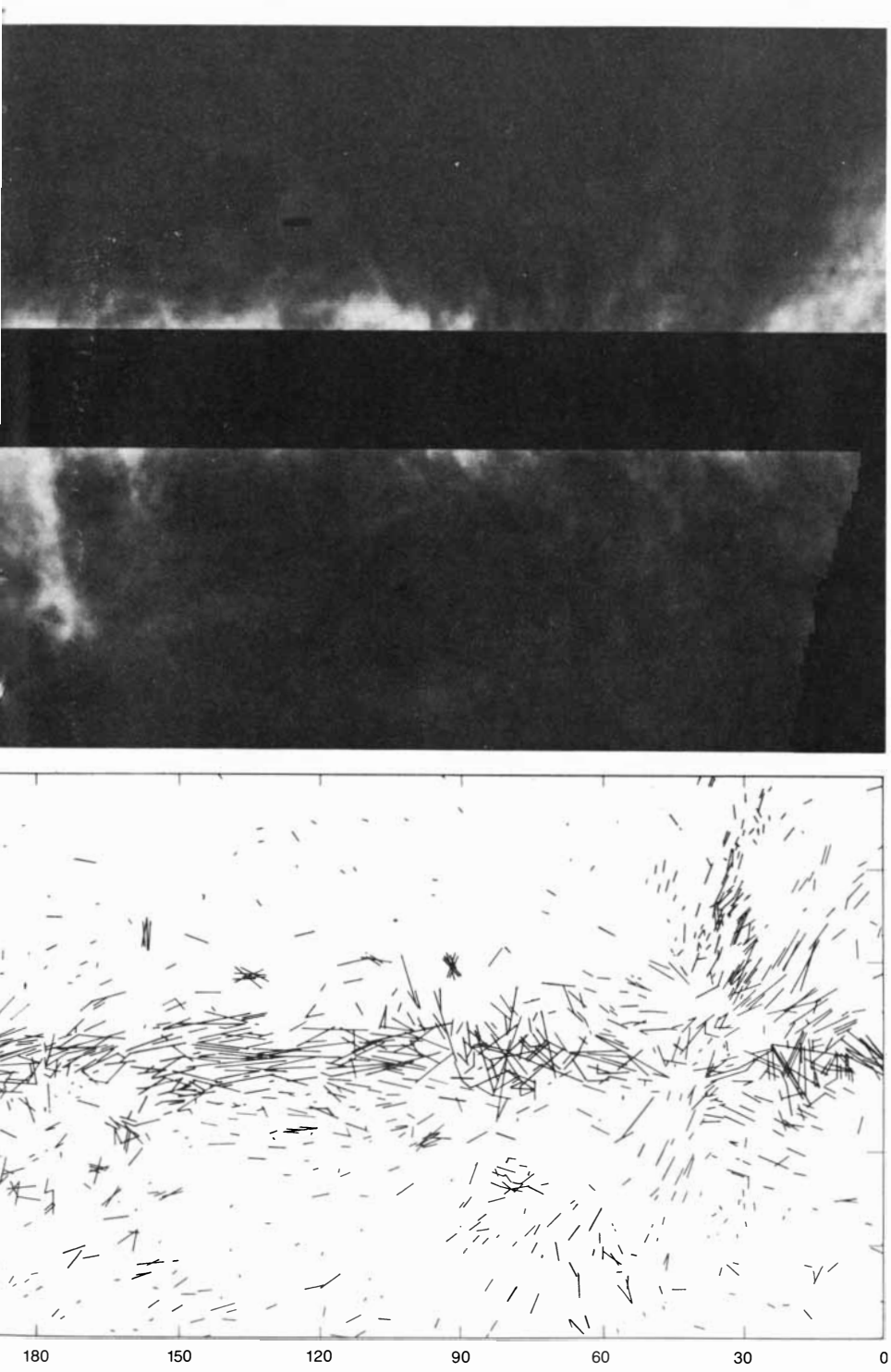
vidual concentrations of interstellar material between the solar system and the star emitting the light. If there is only one concentration of interstellar material between the solar system and the star, the direction in which the starlight is polarized unambiguously indicates the direction of the magnetic field. The polarization of starlight has been measured for a large number of stars scattered over the entire sky. The measurements have been compiled and organized into a map by D. S. Mathewson and K. L. Ford of the Mount Stromlo and Siding Spring Observatories of the Australian National University [*see illustration below*].

Even with the large number of stars represented on the polarization map, there remain regions of the sky that are inadequately covered. There are also




**FILAMENTS ARE ALIGNED with the magnetic field of the galaxy, as can be seen in this comparison of a synthetic photograph of the interstellar medium (*top*) and a map of the polarization of starlight (*bottom*). Dust lying between a star and the solar system polarizes light from background stars. Each line represents a measurement of the amount by which the starlight is polarized; the length of the line is proportional to the strength of the polarization, and the direction of the line indicates the direction of the polarization. Galactic**

many regions where the polarization appears to be chaotic, probably because there is more than one concentration of interstellar material along the line of sight. These problems make it difficult to advance general statements about the exact relation between the interstellar gas and the galactic magnetic field. There is one region, however, near positive galactic latitudes close to the galactic longitude of 30 degrees where the alignment of the gas with the magnetic field is striking. From this clear relation I believe we can conclude that the gas filaments are indeed aligned with the galactic magnetic field.




**magnetic field is parallel to direction of polarization. Well-defined arches in magnetic field at positive galactic latitudes near galactic longitude of 30 degrees are well aligned with filaments of gas in that vicinity. Areas where field seems to be chaotic and filaments are not well aligned may be regions where there are several concentrations of gas between star and solar system.**

The distances to the stars in this region indicate that the distance from the solar system to the gas there is about 100 parsecs. The long arcing filaments of gas extend some 50 degrees across the sky, meaning that they must be some 100 parsecs long. Their width is a few degrees, roughly equivalent to 14 parsecs, which was the diameter of a "standard cloud" predicted by the cloud model of the interstellar medium on the basis of the older visual observations. It seems highly unlikely that the agreement between the width of the filaments and the prediction of the cloud model is fortuitous. Since the visual observations of the interstellar gas depended on seeing the gas illuminated by background stars, given the sparse sampling of the interstellar gas that can be achieved with this technique it is unlikely one would see two or more stars lined up along a filament. Therefore the visual observations of the interstellar medium yielded a size for the gaseous features that is equivalent to the width of a filament rather than to its length.

## Tangles of Filaments

The first synthetic photograph of the total amount of the gas in interstellar space thus revealed much about the structure of the interstellar medium. My colleagues and I reasoned that it should be possible to learn even more about the gas by examining how its filaments are moving. We chose to represent different velocities of the gas by different colors. Following the convention that at visible wavelengths the radiation emitted by an object moving away from the solar system is Doppler-shifted toward the red region of the spectrum, we chose red to represent gas moving away from the solar system, yellow-green to represent gas having almost no Doppler shift and blue to represent gas moving toward the solar system. We displayed all the gas in the large velocity interval between minus 20 kilometers per second and plus 20 kilometers per second; the intensity of each color represented the amount of gas moving at velocities within that interval.

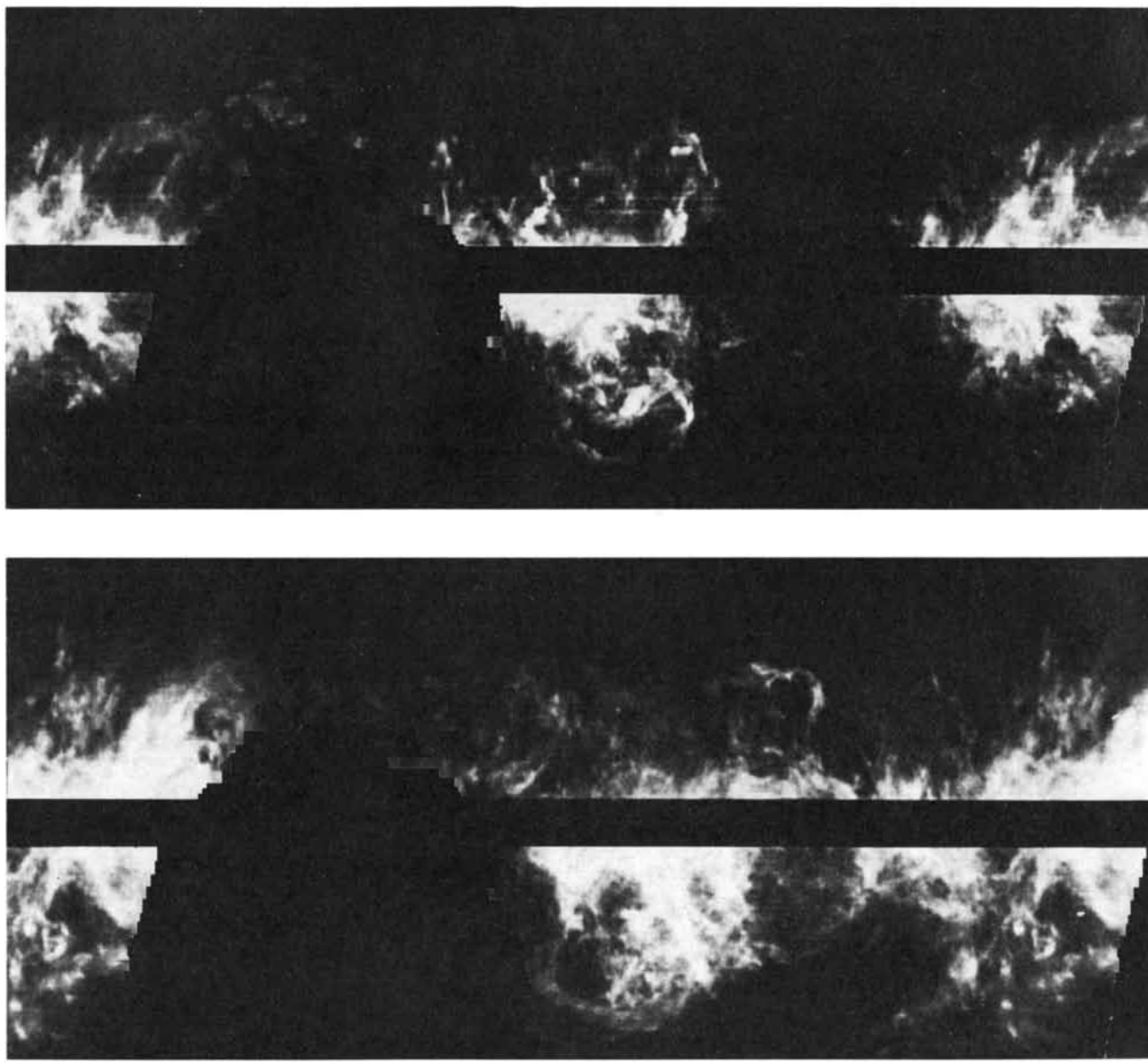
The most striking large-scale characteristic of this color synthetic photograph is the fact that at low galactic latitudes the color of the gas alternates from red to blue with increasing galactic longitude. The alternation in color results from the differential rotation of the galaxy. The galaxy does not rotate like a solid body such as a phonograph record; if it did, there would be no motion of one portion of the galaxy with respect to another. Instead the inner portions of the galaxy rotate faster than the outer ones. This differential rotation means stars that are far apart have a larger mo-

tion with respect to each other than stars that are close together. The observed large-scale differential velocities of stars with respect to the sun can be correlated with the average velocity of the interstellar gas and so can be employed to make crude estimates of the distance of the gas from the solar system. The estimates of the distance of the gas by this method agree with the independent estimates of the distance determined by the other techniques I have described.

## Individual Filaments Revealed

On a smaller scale the color synthetic photograph revealed that the gas has a velocity structure that is quite detailed. In general an individual filament is not characterized by a single velocity. In fact, to us there seemed to be no systematic pattern to how the velocity varies with the filament's length and width. Portions of the filament appeared to be curling or looping in such a way that its structure was characterized by seemingly random variations in velocity. Such variations provided much more information about the motions within the filament than we had had before, but it was information of a kind that is disappointing to someone seeking to understand the overall structure of the interstellar gas; it is difficult to arrive at a straightforward interpretation or explanation of a jumble of random numbers.

In an effort to make sense of the intricate structure of the interstellar medium we decided to examine the gas in smaller velocity intervals. We broke up the entire range of velocities from minus 20 kilometers per second to plus 20 kilometers per second into 20 velocity intervals of two kilometers per second and synthesized a photograph of the gas for each one. Given the results of the color picture over the entire range that revealed apparently random variations in velocity on a relatively small angular scale, we did not expect our series of narrow-velocity-interval photographs to yield fruitful results. This expectation was wrong. It seems that this step was the key to understanding the true nature of the structure of the interstellar gas and the kind of physical phenomena that mold that structure: it enabled us to

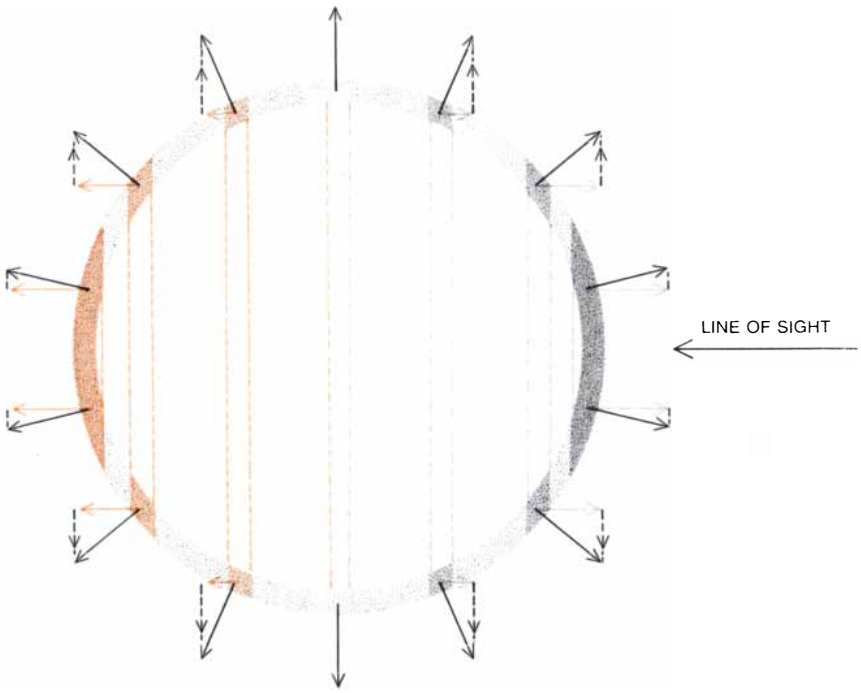


**ELLIPTICAL RING OF GAS** can be seen near the bottom center of both of these synthetic photographs, reproduced here as negative prints. The top picture is a synthetic photograph of the gas in the narrow velocity interval between plus 16.4 and plus 18.4 kilometers per second; the bottom picture is a synthetic photograph of the gas in the velocity interval between plus 3.7 and plus 5.7 kilometers per second. Note that the elliptical ring is somewhat larger in the bottom picture, which was made at the lesser velocity, than it is in the top picture.

view the gas in one structure without any interference from other gas along the line of sight.

It is not possible to show all 20 of the synthetic photographs here, and so I have chosen only the two that display an interesting feature that best illustrates my argument. The two synthetic photographs appear on the opposite page. The top one covers the interval between plus 16.4 and plus 18.4 kilometers per second; the bottom one covers the interval between plus 3.7 and plus 5.7 kilometers per second. Both show a large elliptical structure of gaseous filaments some 50 degrees across centered near the galactic longitude of 200 degrees and the galactic latitude of −40 degrees. Comparison of the two synthetic photographs reveals that the structure has a large diameter at the lesser velocity and a small diameter at the greater one. We were startled to find such a coherent structure over such a large angular region in the sky and large range of velocities. Even more surprising was that this type of structure has a very simple interpretation.

Consider a spherical shell of gas that is expanding radially outward from its center. To an observer outside the shell the Doppler shift of the moving gas depends on the projection of the motion of the gas in the line of sight. The gas on the nearest edge of the shell and the gas on the farthest edge (on the line of sight through the center) will be respectively directly approaching the observer and directly receding from him and will have the greatest Doppler shift. The gas on the limb of the shell at the greatest angular distance from the center, however, will be moving almost at right angles to the line of sight and will have the smallest Doppler shift. The gas between the center of the shell and the limb will have intermediate velocities in the line of sight and will have intermediate Doppler shifts. Thus in each narrow velocity interval the observer should see a ring of gas whose angular size decreases with increasing line-of-sight velocity, and indeed that is what is observed. The elliptical filamentary feature we see has its smallest diameter at the largest velocity and its largest diameter at almost zero velocity. The shell is expanding at a velocity of 23 kilometers per second.




**RADIALLY EXPANDING SHELL OF GAS in interstellar space would give rise to exactly such an elliptical or circular feature in the synthetic photographs on the opposite page. Here a spherical shell of gas, shown in cross section, is expanding radially in all directions (*black arrows*). The Doppler shift, by which the shell's velocity is measured by the observer, depends only on the component of the shell's velocity in the line of sight (*arrows in gray and in color*). When the distant observer examines such a shell in only a narrow interval of velocity, he will see a ring whose size depends on the line-of-sight velocity. The measured velocity will be at its highest for the nearest point (*darkest gray*) and the farthest point (*darkest color*) of the shell, which are almost directly approaching the observer or receding from him; at that high velocity the measured diameter of the shell will be small. Measured velocity will be at its lowest for areas near limb of shell (*lightest gray and lightest color*), which are traveling almost perpendicularly to the observer's line of sight; at that low velocity the measured diameter of shell will be large.**

### The Filaments Are Shells

In reality the shell observed in the synthetic photograph is not an ideal and simple structure. At any particular velocity the ring is not circular but is elliptical. The shell cannot be traced through the full range of velocities but is most easily visible at positive velocities that correspond to the speed at which the gas at the far side of the shell is receding. Moreover, in most of the velocity ranges only a portion of the shell can be seen. It must be remembered, however, that the interstellar medium is inhomogeneous and that therefore one cannot expect perfect agreement between an ideal model and reality. Departures from ideal models are so common in observations of the interstellar gas that in fact we feel fortunate in having observed such a good specimen with which to illustrate the model of a radially expanding shell.

If the distance to the elliptical shell could be determined, it would be possible to learn more about the shell's physical properties. Distance information for this particular feature is meager, but the data available indicate that a distance of 150 parsecs is probably accurate to within a factor of two. At this distance the angular size of the shell indicates that its diameter is 100 parsecs, and that in the course of its expansion through space it has swept up a mass of interstellar material equal to 73,000 times the mass of the sun.

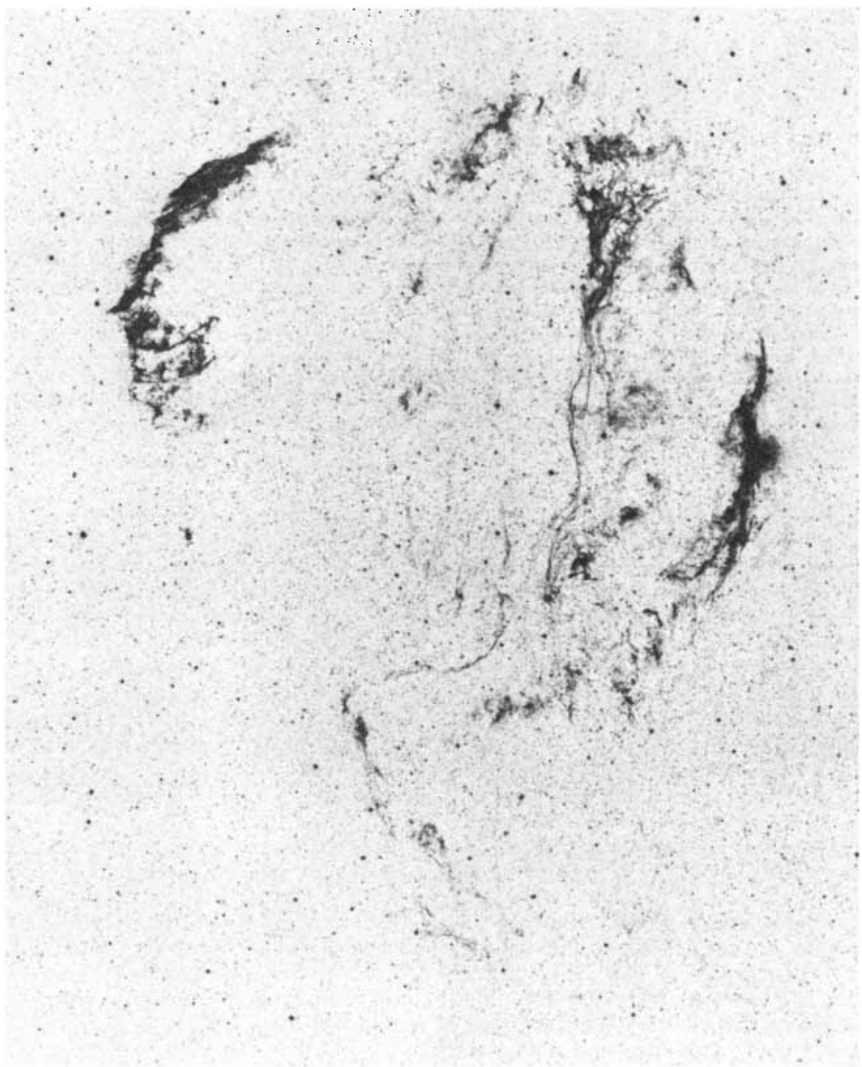
There are two known mechanisms that can give rise to such a large expanding shell of interstellar gas. One mechanism, recently suggested by John I. Castor, Richard McCray and Robert Weaver of the University of Colorado, is a transfer of energy from a highly energetic stellar wind: a rapid flow of material away from a star. Stellar winds have indeed been observed in certain types of stars. Such stars are not very common, and there is no star within the observed elliptical shell of gas that is known to have an energetic wind. The shell may have been created by such a star that has since evolved and become extinguished, but the energy of the shell is probably larger than could be provided by one such star or even several.

The other mechanism for producing an expanding shell of gas is a supernova explosion: the catastrophic near-obliteration of a massive star near the end of its life cycle. The Cygnus Loop, or the Veil Nebula in the constellation Cygnus, is an expanding shell created by a supernova that exploded some 45,000 years ago. It is expanding at a velocity of about 100 kilometers per second and has a radius of some 20 parsecs. It is a younger object than our newly discovered elliptical shell, which has swept up more interstellar matter and is moving more slowly. On the basis of theoretical models of old supernova shells calculated by Roger A. Chevalier of the Kitt Peak National Ob-

servatory and William C. Straka of Boston University it seems evident that our elliptical shell was thrown off by a supernova that exploded some 650,000 years ago and released a total energy of $4 \times 10^{51}$ ergs. This amount of energy is near the upper limit of what is generally accepted as the energy output of a supernova, and it is probably several times the amount released in the formation of the Cygnus Loop.

An expanding shell of interstellar gas, whether it is created by a stellar wind or by a supernova, should have a characteristic structure. Just behind the leading edge of the shell the gas is compressed and heated. As this heated gas is carried along with the shell, it slowly retreats from the leading edge toward the interior of the shell. As it does so it is cooled but is compressed even further. If the shell is the product of an energetic stellar wind, ultraviolet emission from the star will ionize the dense region behind the leading edge.

The shell consists of nearly all the matter that was originally contained in the volume the shell now encloses. The interior of the shell is not empty, however: it contains a small amount of the original interstellar gas, which is now very hot, rarefied and at high pressure. This rarefied gas is the source of the X-ray emission generally observed from the interior of supernova remnants [see "X Rays from Supernova Remnants," by Philip A. Charles and J. Leonard Culhane; SCIENTIFIC AMERICAN, December, 1975]. The rarefied gas is also the probable source of oxygen so highly ionized that it is stripped of five of its electrons. Such highly ionized oxygen, denoted O VI, is often seen in the ultraviolet spectra of hot stars observed with the telescope aboard the *Copernicus* satellite. Both the ultraviolet emission from O VI and X-ray emission have been observed in the interior of our elliptical shell.



**CYGNUS LOOP, photographed at visible wavelengths with the 48-inch Schmidt telescope on Palomar Mountain, is a shell similar to the elliptical shell shown in the synthetic photographs on page 80. As can be seen in this negative print, the Cygnus Loop is not perfectly circular because the surrounding interstellar medium is not homogeneous and smoothly distributed. Shell is expanding at about 100 kilometers per second and has a radius of some 20 parsecs. It was probably created by a supernova, a star that exploded catastrophically, some 45,000 years ago.**

## Structure of the Shells

The way matter is concentrated in such a shell suggests why the filaments of interstellar gas are aligned with the galactic magnetic field. Even the coldest gas in interstellar space is slightly ionized by starlight. Since ions are electrically charged particles, they will flow as an electric current in the presence of a magnetic field. The gas and the magnetic field are thus inextricably bound together by the slight electrical conductivity of the gas. As the expanding shock front of the shell sweeps up the interstellar gas, it also sweeps up the magnetic field. The magnetic field should be stronger in the shell than it is in the surrounding interstellar medium in the same ratio that the density of the gas is higher in the shell than it is in the surrounding medium. Moreover, the direction of the magnetic field will tend to lie along the surface of the shell.

How filaments form within a shell is not well understood. They may do so by the same process that creates filaments in the corona of the sun. In all the synthetic photographs of the interstellar medium at the narrow velocity intervals a multitude of long filaments can be seen; there are many more than are discernible on the photographs I have presented here. Most of the filaments show no change in structure with velocity. Nevertheless, my colleagues and I regard the filaments too as being portions of expanding shells, portions that have been slowed down by collisions with the ambient interstellar gas or with other shells.

Our interpretation is an extrapolation from the relatively reliable observations of the few shells that do have different sizes at different velocity intervals, but we think it is justified. One reason is that in the case of one shell showing a pronounced change in diameter with velocity and in the case of two other shells exhibiting little or no change in diameter with velocity there is independent evidence in the form of radio-wavelength synchrotron emission that the features resulted from a supernova explosion. Synchrotron emission is generated by highly energetic electrons spiraling in an interstellar magnetic field at nearly the speed of light. This type of emission is often associated with supernova shells; the electrons are believed to have been accelerated to high energies in the original supernova explosion.

The many filaments visible in the syn-

thetic photographs imply that there are many shells in interstellar space. The existence of such a large number of shells has interesting ramifications that were first explored theoretically by Donald P. Cox and Barham W. Smith of the University of Wisconsin. If there are many shells in interstellar space, it is likely that as they expand they will collide. After the collision the interiors of the shells are connected, and the hot gas inside the shells is free to flow rapidly from one to another. In a steady-state situation it is likely that a number of shells would join to form tunnels of hot, low-density gas surrounded by thin shells of cold, higher-density gas and magnetic fields. Occasionally a shell might burst like a bubble if there were not enough cold material to resist the pressure of the hot gas. The burst shell would provide an escape hatch for the confined hot gas, which could then flow into the regions outside the space bounded by the shells.

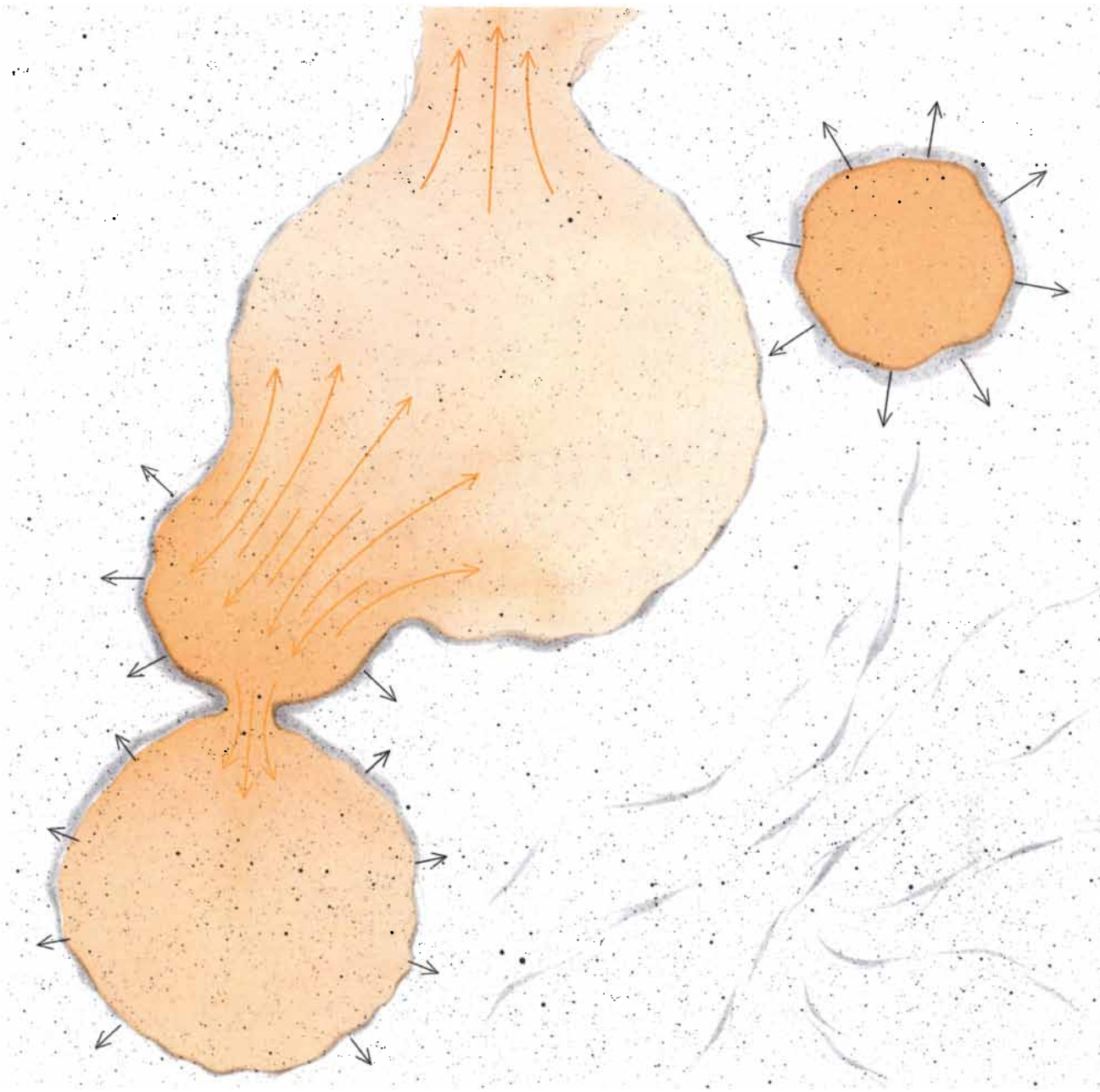
### The Shell Picture

The radio-wavelength survey we have conducted seems to be consistent with this picture of the structure of the interstellar medium; the picture also seems to be confirmed by the observations of X rays and of the ultraviolet emission of O VI in interstellar space. The shell picture of the structure of the interstellar medium is radically different from the cloud model, and it is not yet widely accepted. Important questions remain that must be answered before this description can be considered completely correct. Are observations at visible wavelengths consistent with such a picture or do they provide data on yet another component of the interstellar gas? Are the temperature and the density structure of the observed interstellar gas consistent with what would be expected behind the leading edge of a shell? Are the topology and the strength of the magnetic fields in interstellar space consistent with this interpretation?

New instruments just completed or now under construction will play a crucial role in efforts to attack such questions. In particular the 1,000-foot radio telescope at Arecibo in Puerto Rico, which has recently been resurfaced, and the Very Large Array of radio telescopes in New Mexico should provide much of the necessary data at radio wavelengths. Data at both X-ray and ultraviolet wavelengths are required to determine whether or not the hot gas is preferentially associated with the interior of shells. Such data can be obtained only from instruments in space. Meanwhile astronomers on the ground are observing the interstellar medium at visible wavelengths with many new devices that provide data of much higher quality that can be more readily compared

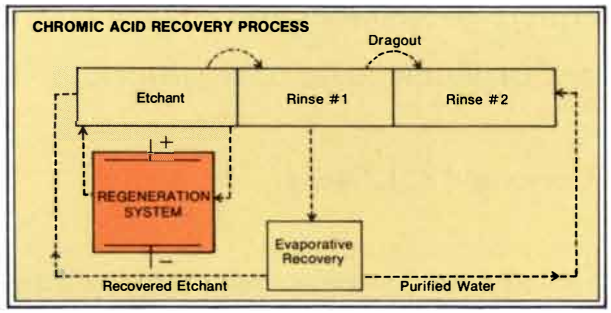
with the data obtained at radio wavelengths.

There are also many questions concerning the origin of the shells. The energy required to create such shells is great, possibly greater than the energy that can be supplied by stellar winds or supernova explosions. The number of shells seems to be large. Although no one has yet analyzed this multiplicity in detail, it is possible that there are too many shells compared with the generally accepted rates at which stars with energetic stellar winds are formed or at which supernovas explode. If that is the case, perhaps stars with energetic stellar winds or supernovas are commoner than we now suppose. Or perhaps there is still another kind of object, as yet unknown, that injects vast amounts of energy into the interstellar medium.



**SHELL PICTURE of the structure of the interstellar medium is proposed by the author to account for the filamentary structure of the interstellar medium. A shell of gas is produced either by an energetic stellar wind from a hot star or by the explosion of a supernova. A newly formed shell expands rapidly (*top right*), sweeping up gas in the surrounding cold interstellar medium (*light gray*) and compressing and heating it to form a dense outer rim of ionized gas such as is visible in the Cygnus Loop (*dark gray*). As the shell expands it leaves behind in its interior hot, rarefied gas (*color*) that emits X rays and also produces the absorption lines of highly ionized oxygen atoms in the ultraviolet spectra of hot stars. As a shell grows larger and older the gas is further compressed and cooled. Electrons recombine with ions to form atomic hydrogen, which emits radiation at the radio wavelength of 21 centimeters. Meanwhile the shell's expansion slows (*bottom left*). According to theoretical work by Donald P. Cox and Barham W. Smith of the University of Wisconsin, when many large expanding shells coexist, in time they will collide. When they do, their interiors are connected, so that the hot gas can pass from one shell to another. If enough shells are connected, the hot gas will be able to flow through tunnels in the interstellar medium formed by the cold, dense shells and magnetic fields. A very old shell may eventually become so thin that its weakest part will burst like a bubble (*top center*), enabling the hot gas to escape into surrounding space. Remnants of ancient, shapeless shells form wispy filaments (*bottom right*).**

In industry, chromic acid and oil-water machining emulsions usually live only once. They do their job; then they get discharged, sludged, trucked away, and buried. End of story.




Or is it?

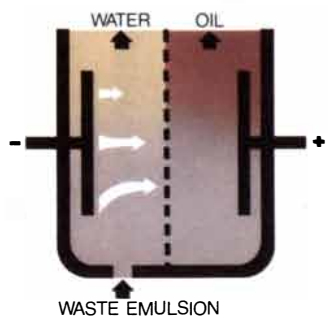
Recently, electrochemists here at the General Motors Research Laboratories have developed ways to infuse new life into these spent materials. In fact, GM's Guide Division is presently using a process of ours to recycle chromic acid.

The acid etches plastic components for plating. However, etching reduces the hexavalent chromium ions to a less effective trivalent form, and eventually the etchant loses its bite. At one time, that meant dumping the bath.

Now, our electrolytic process oxidizes the Cr (III) ions back to Cr (VI). This restores good etching and greatly extends chromic acid life. But first we had to:

- Find an anode material that would favor ion oxidation, rather than oxygen evolution.
- Determine the best anode/cathode ratio and current densities for efficient conversion.

We hope to extend oil life too, in this case by breaking up oil-water emulsions. It's all done in a multicelled electrophoretic unit. Direct current drives oil droplets to the anodes. There the droplets coalesce, and the oil floats to the surface to be collected.




In a successful pilot test, our method reclaimed nearly 90% of the oil without requiring chemicals or forming sludge.

Applying electrochemical technology: Part of the Labs' ongoing program to help conserve critical materials and protect the environment.

# We're recharging and reclaiming "waste" materials.

**General Motors Research Laboratories**
Warren, Michigan 48090

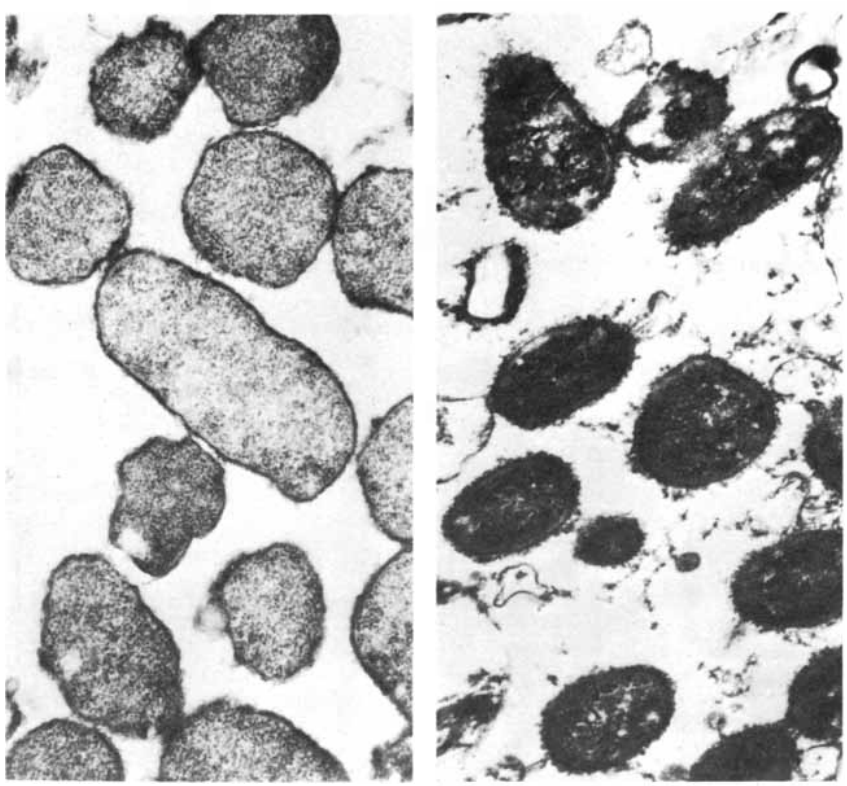
# How Bacteria Stick

*In nature (but not in laboratory cultures) bacteria are covered by a "glycocalyx" of fibers that adhere to surfaces and to other cells. Adhesion might be prevented by a new kind of antibiotic*

by J. W. Costerton, G. G. Geesey and K.-J. Cheng

Bacteria stick, tenaciously and often with exquisite specificity, to surfaces ranging from the human tooth or lung and the intestine of a cow to a rock submerged in a fast-moving stream. They do so by means of a mass of tangled fibers of polysaccharides, or branching sugar molecules, that extend from the bacterial surface and form a feltlike "glycocalyx" surrounding an individual cell or a colony of cells. The adhesion mediated by the glycocalyx determines particular locations of bacteria in most natural environments; more specifically, it is a major determinant in the initiation and progression of bacterial diseases ranging from dental caries to pneumonia.

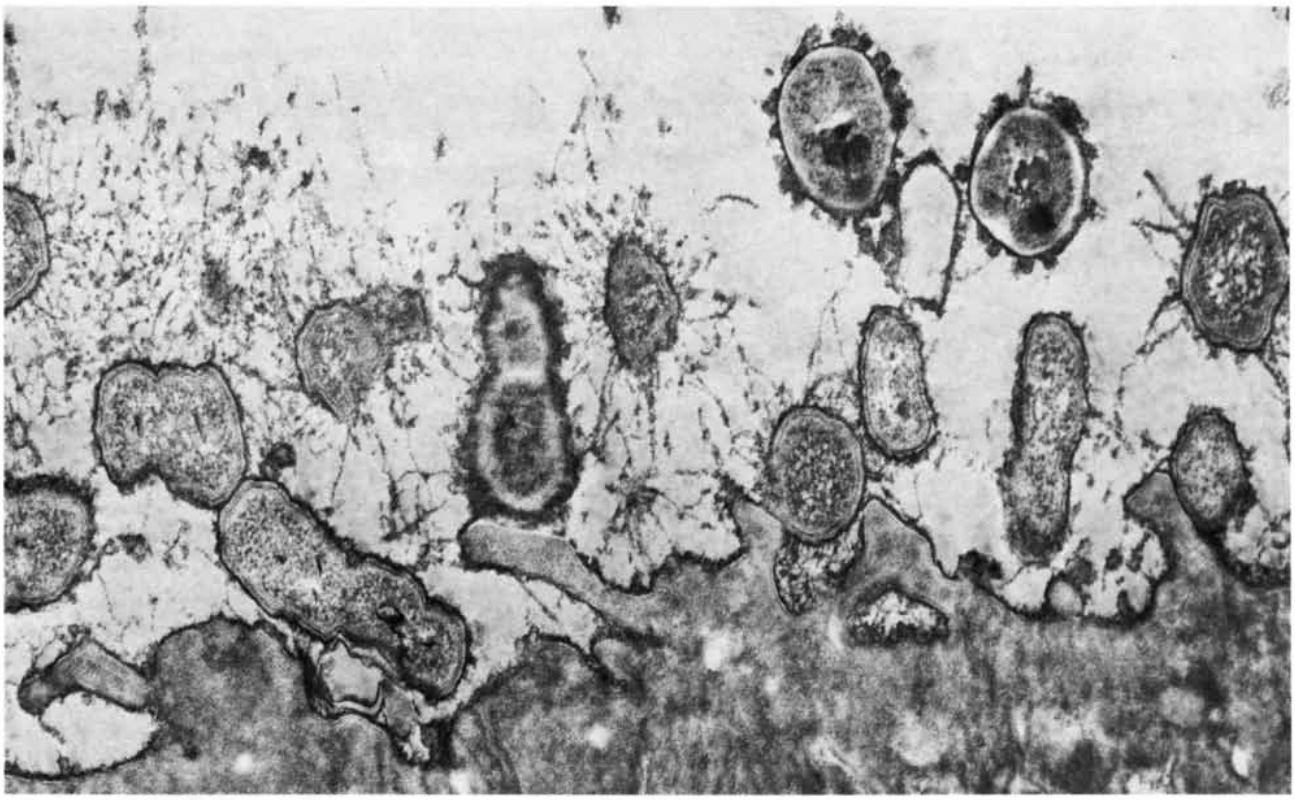
These major—and, with the benefit of hindsight, rather obvious—facts about the bacterial cell surface have become known only within the past decade. Ironically the main reason for the late discovery of the bacterial glycocalyx and its functions was the long reliance by microbiologists on an otherwise eminently effective investigative system: the pure laboratory culture of an individual bacterial strain. To generate and maintain a glycocalyx a bacterial cell must expend energy, and in the protected environment of a pure culture the glycocalyx is a metabolically expensive luxury conferring no selective advantage; cells that fabricate these elaborate coatings are usually eliminated from pure cultures by uncoated mutants that can devote more of their energy budget to proliferation. Microbiologists have largely studied such naked mutants.



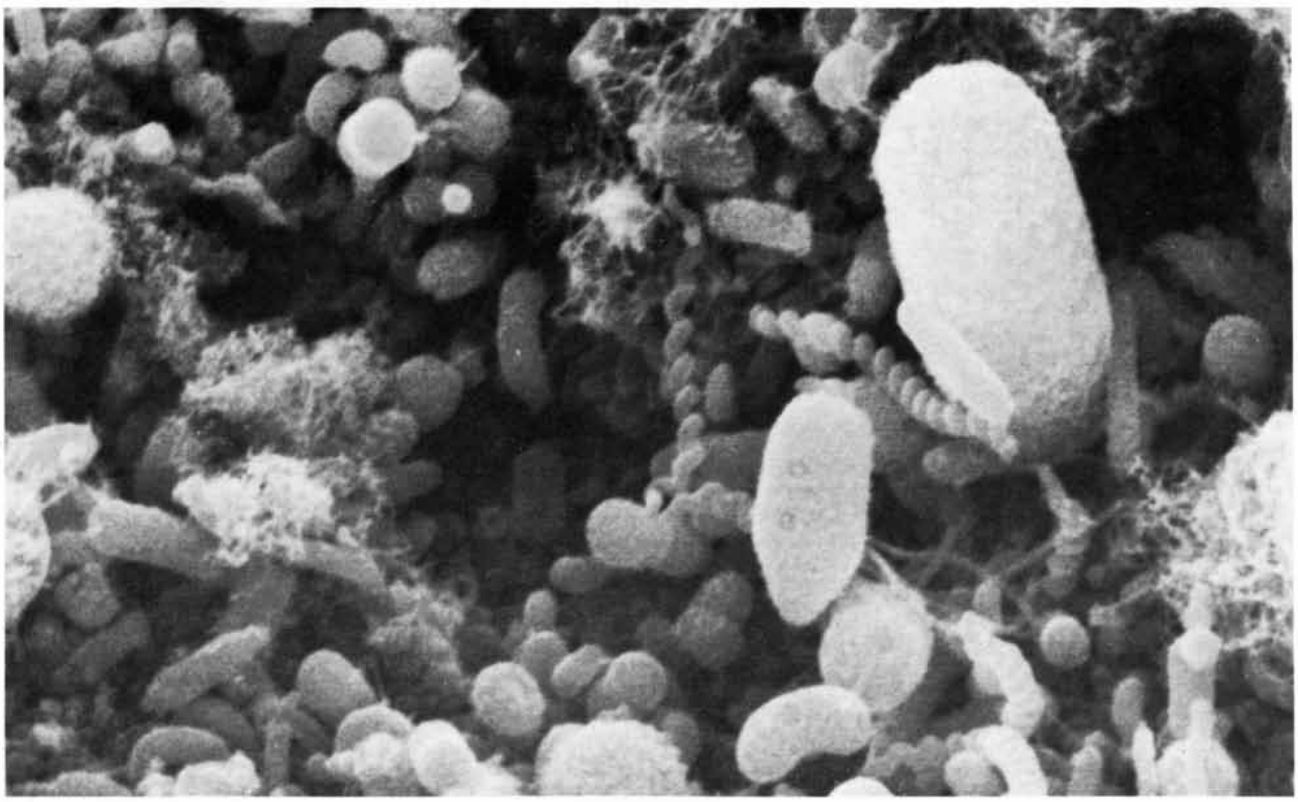
**NAKED BACTERIA (*left*) are from a typical pure laboratory culture of *Escherichia coli;* the glycocalyx-coated bacteria (*right*) are *Pseudomonas* cells from an infected human bladder. In both preparations the cells were stained with ruthenium red, which is taken up by any polysaccharide glycocalyx fibers that are present. The bacterial glycocalyx was ignored until recently because the familiar pure laboratory strains do not need it and therefore do not fabricate it.**

In a competitive natural environment populated by several kinds of bacteria, on the other hand, selection favors cells that are protected, and enabled to adhere to a desirable surface, by a glycocalyx. It was only in 1969 that Ivan L. Roth of the University of Georgia demonstrated carbohydrate fibers surrounding bacteria in an aquatic system and Ian W. Sutherland of the University of Edinburgh Medical School characterized the surface polysaccharides of bacteria taken from natural environments, thus drawing attention to the universality of what we now know as the glycocalyx. Since then studies in our laboratories at the University of Calgary and at the Agriculture Canada Research Station at Lethbridge in Alberta, and in a number of other laboratories, have made it clear that the glycocalyx is essential to the biological success of most bacteria in most of the varied natural environments in which they are observed.

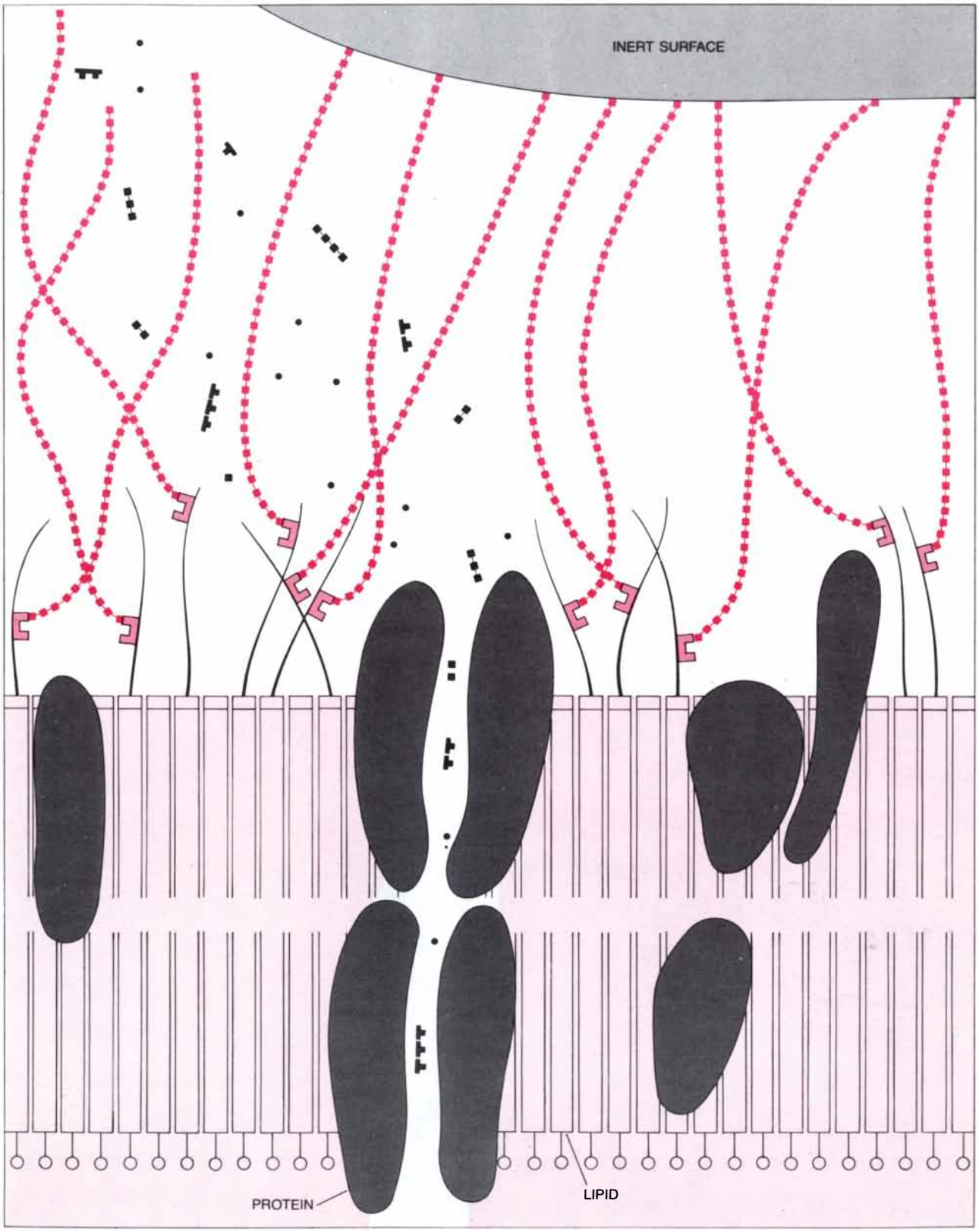
The polysaccharide-coated surface is not a peculiarity of the bacterial cell. The more rigid polysaccharide cell wall of higher plants was among the first microscopic structures described by Robert Hooke in 1665. The analogous surface of animal cells—a glycocalyx like that of bacteria—was described only in 1971, by Vincent T. Marchesi and his colleagues at the National Institute of Arthritis, Metabolism, and Digestive Diseases. They isolated and identified glycoproteins arrayed in the membrane of animal cells and showed that the polysaccharide fibers they bear extend outward from the membrane to form



**BACTERIAL GLYCOCALYX** is a network of fibers that extend from the bacterial surface. The fibers adhere to one another and to inert or animal-cell surfaces nearby. This electron micrograph made in the laboratory of one of the authors (Costerton) shows a mixed population of bacterial cells that have been attached by their glycocalyxes to one another and to the surface cells of a cow's rumen.



**MIXED POPULATION** of a number of different kinds of bacteria and some yeast cells adhering to the lining of a cow's rumen is seen in a scanning electron micrograph. Here the individual polysaccharide fibers by means of which the bacteria are held to the rumen cells and to one another are not resolved, but some of the glycocalyx fibers have coagulated and are seen as masses of branching foamy material.



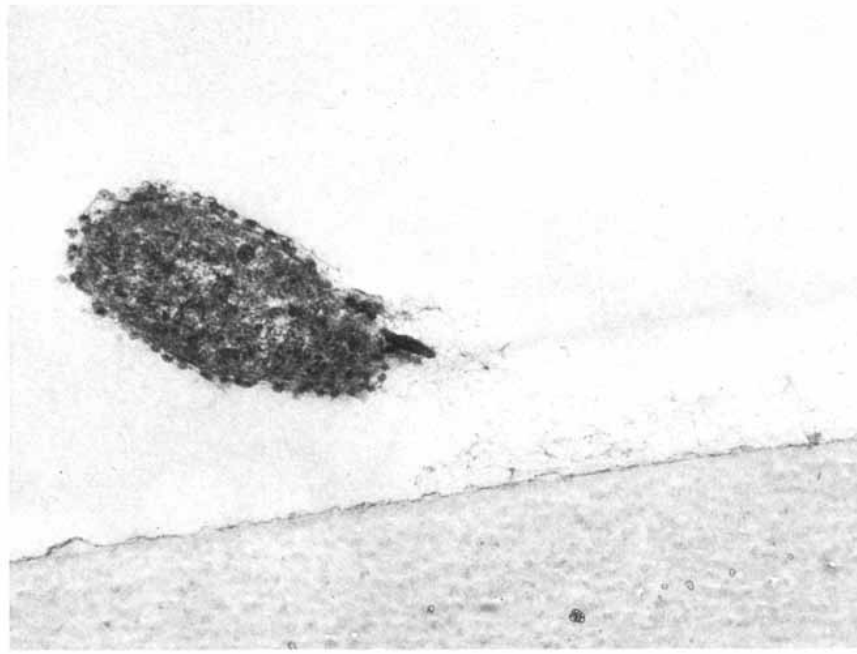

**GLYCOCALYX extends from the outer membrane of a bacterium as is indicated in this generalized and highly schematic diagram. The membrane is a bilayer of lipid molecules (*forked structures*) in which protein molecules (*gray shapes*) are embedded. Lipopolysaccharide molecules (*black hairlike structures*) extend from the membrane. The glycocalyx is a mass of long polysaccharide fibers (*chains of colored squares*). The fibers are chains of sugar molecules that are generated by bacterial enzymes called polymerases (C-*shaped structures*) affixed to the lipopolysaccharides. The glycocalyx fibers adhere to nearby surfaces, in this case an inert surface (*top right*). In addition to mediating bacterial adhesion the fibers channel toward the bacterium various nutrients such as sugars (*rectangles*), amino acids (T-*shaped objects*) and inorganic ions (*dots*), which enter the cell through channels in the membrane formed by arrays of proteins.**

not just projections but a matted glycocalyx that is itself a continuous surface, actually the functional surface of the cell. The precise chemical nature of this ubiquitous glycocalyx varies with the sugar composition of the fibers, so that it differs from cell to cell within the body of the animal and probably also at different stages in the life of the individual cell.

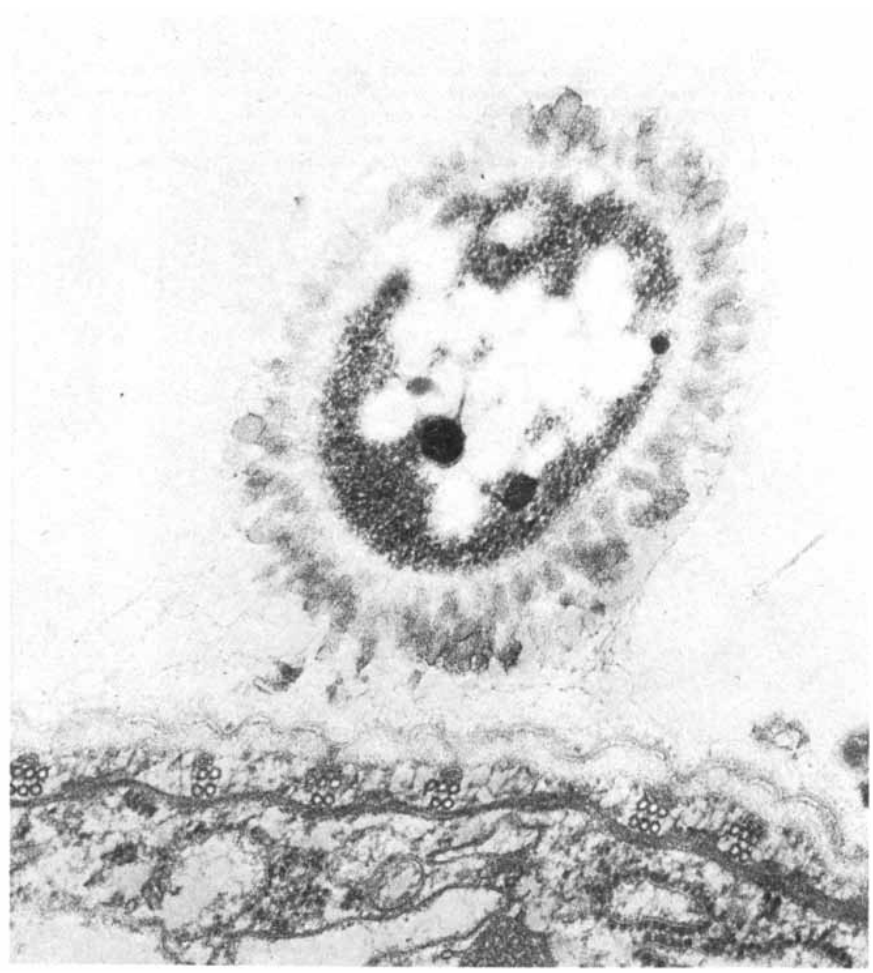
The striking conclusion to be drawn from these observations is that the real functional surface of all cells—bacterial, animal and higher-plant—is a tangled mat of polysaccharide fibers fabricated and oriented by the cell itself. A bacterial cell adheres to a plant cell, an animal cell or another bacterial cell by juxtaposing its own glycocalyx to the surface of the cell it adheres to. In many cases the link is supplied by the simple proteins called lectins, which bind very specifically to polysaccharides with a particular molecular structure [see "Lectins," by Nathan Sharon; SCIENTIFIC AMERICAN, June, 1977].

The pioneering investigation of glycocalyx formation in bacteria was actually a study of *Streptococcus mutans,* an organism that colonizes the human tooth, by Ronald J. Gibbons of the Forsyth Dental Center in Boston and other workers. Between 1960 and 1967 they reported that three enzymes at the bacterial surface deal in particular ways with the common sugar sucrose, which is composed of one molecule of glucose and one of fructose. The enzyme invertase splits the sucrose into its two components, both of which are released to become sources of energy for the cell. A second enzyme, glucosyltransferase, splits the sucrose and releases fructose as a nutrient but polymerizes the glucose into a long polysaccharide called a glucan, which is insoluble in water. A third enzyme, fructosyltransferase, builds the fructose into a similar but water-soluble polysaccharide and liberates glucose.

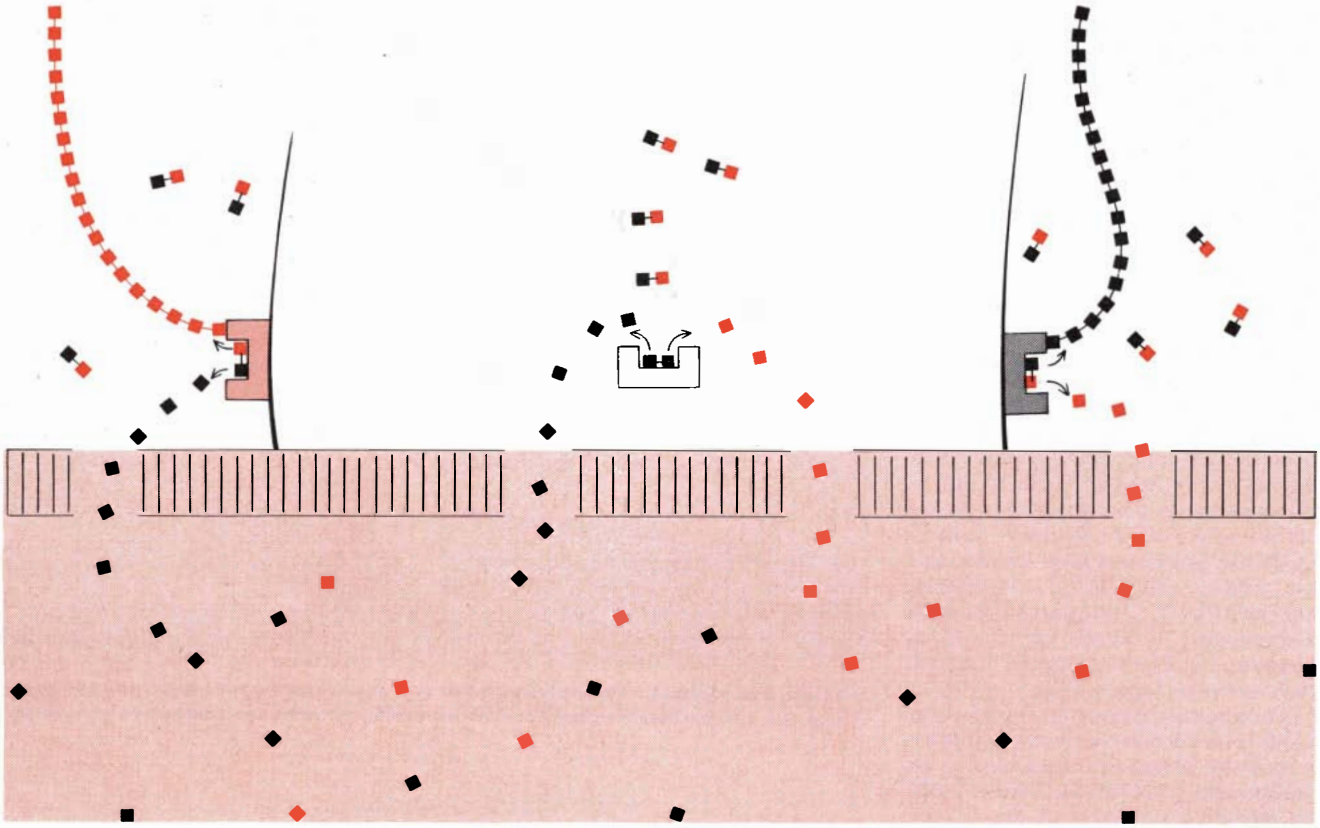
The important finding is that the glucan is somehow able to adhere to the inert enamel surface of a tooth and thus to attach the bacterial cell to the tooth. Since glucosyltransferase is present throughout the glucan network of the glycocalyx, the glucans it continues to produce keep thickening the glycocalyx, entrapping more bacterial cells of the same or different species and building up the yellowish film known as plaque. Another human oral bacterium, *Streptococcus salivarius,* colonizes not the tooth itself but the gum. It liberates free glucans that migrate to adjacent tooth surfaces and, with their accompanying glucosyltransferase, build a polysaccharide mat that traps a mixed population of bacteria to form plaque. It is within the plaque that the bacterial enzymes re-



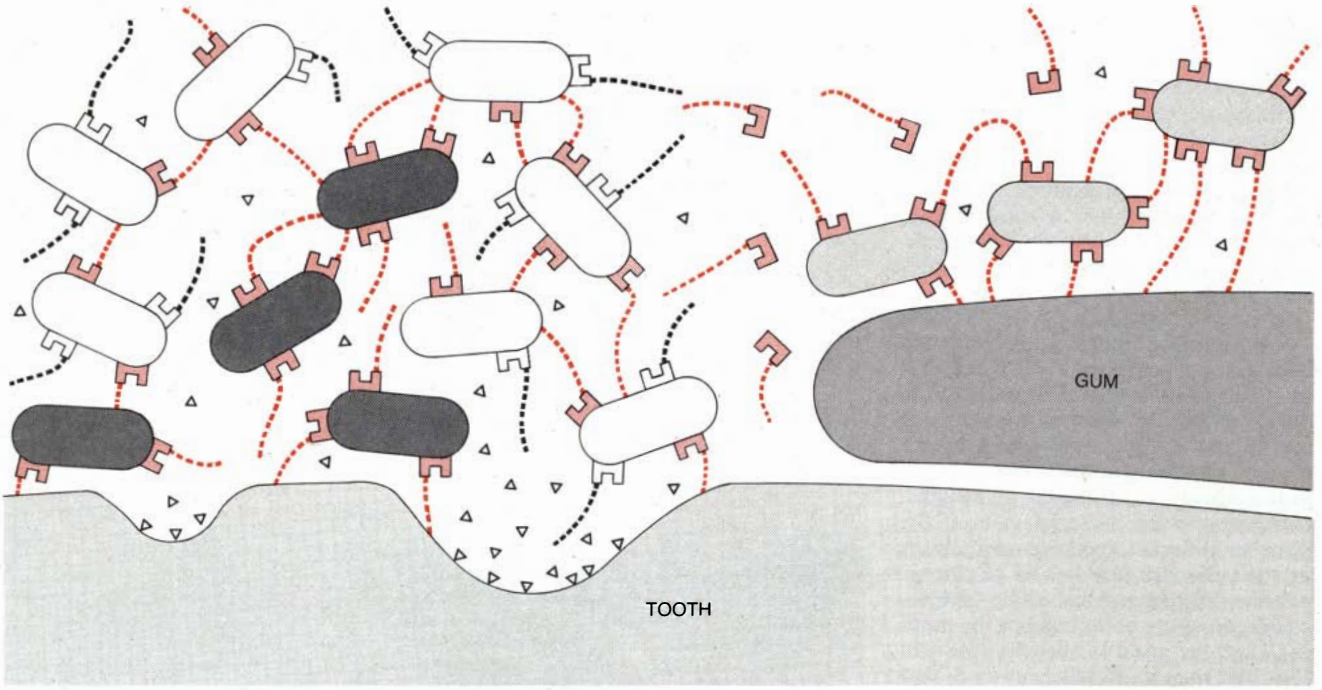
**SINGLE BACTERIAL CELL attached itself to a plastic surface submerged in a mountain stream and was anchored against the flow of rushing water by the fibers of its glycocalyx.**



**ELABORATE GLYCOCALYX attaches a single bacterium (*top*) to an animal-cell glycocalyx, a mat of fibers that is visible above the larger cell's double-layered plasma membrane.**



**DENTAL PATHOGEN** *Streptococcus mutans* **deals with the common sugar sucrose as is shown here, according to Ronald J. Gibbons of the Forsyth Dental Center in Boston. Sucrose is composed of one molecule of glucose (*colored square*) linked to one of fructose (*black square*). Tbe enzyme invertase (*middle*) splits sucrose, releasing the two simple molecules to enter the cell as nutrients. Glucosyltransferase (*left*) splits sucrose and releases the fructose, linking the glucose units to form a long polymer called a glucan, which is insoluble in water and is the major glycocalyx component. Fructosyltransferase (*right*) releases glucose and polymerizes fructose.**




**DENTAL PLAQUE, the yellowish film that forms on teeth, is composed of bacteria fixed in a network of glycocalyx material. Some oral bacteria (*dark gray cells at left*) fabricate their own adherent glucans. Others, such as *Streptococcus salivarius* (*light gray*), which colonizes the gum (*right*), make glucans and liberate some of the fibers, which travel, with their polymerizing enzymes, to a nearby tooth surface. There the free glucans can contribute to the formation of a mat that traps various nonspecific bacteria (*white*), building up plaque. Within the plaque the bacterial enzymes (*triangles*) that attack dental enamel are directed against the tooth, causing cavities.**

sponsible for tooth cavities are concentrated against the tooth enamel.

Gibbons' report on dental pathogens directed the attention of investigators to other natural bacterial environments. Roth and other workers looked at various surfaces submerged in water and found they were covered with a film of bacterial polysaccharide. We undertook to investigate the process that creates such a film by placing plastic disks in a rushing mountain stream. Examining the disks with the electron microscope, we found that a bacterium can anchor itself to the inert plastic by spinning a mat of polysaccharide fibers that withstands enormous shear forces. The initial colonization may be accomplished by either bacteria or algae (which have similar polysaccharide fibers), and in time a complex mixed population of cells builds up within a network of fibers. Often one sees within these populations microcolonies of cells of one type, the products of cell division within a common glycocalyx.

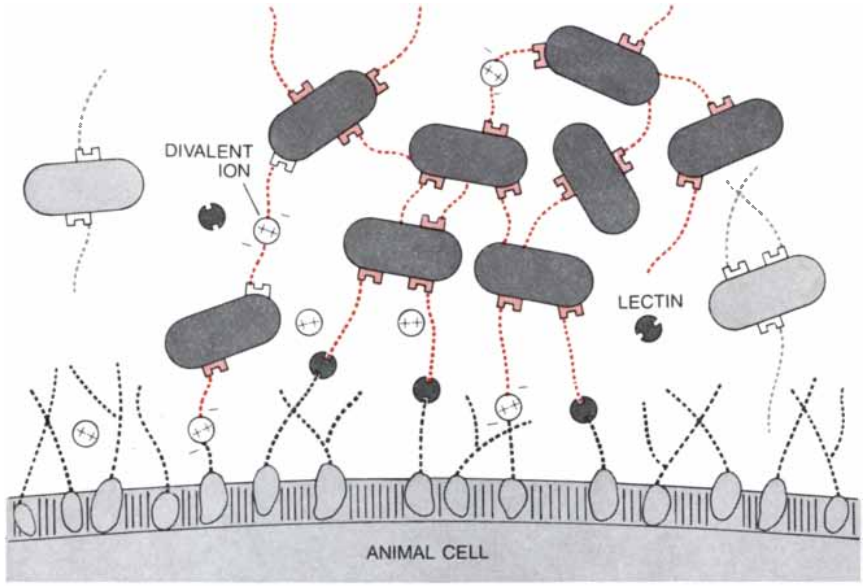
**DIVISION OF BACTERIA within a glycocalyx tends to produce a microcolony of sister cells of the same species. Here a microcolony is seen in an enveloping common glycocalyx that also contains remnants of dead cells whose contents have been recycled as nutrients for the cells that survive. A nonmember cell outside the microcolony has its own glycocalyx (*top right*).**

It is still not known by just what glue-like bond a glycocalyx fiber is linked to a rock or a plastic. In contrast to inert surfaces, the cells of plants and animals present to bacteria a cell wall or a glycocalyx that is chemically defined by the composition of its own particular polysaccharides. The polysaccharide fibers of the bacterial glycocalyx, which are for the most part negatively charged, can form a polar bond with a higher-cell polysaccharide by way of divalent positive ions in the medium. Lectins with a specific attraction for the glycocalyx fibers and for the higher-cell polysaccharides can also form a bridge between them. Bacteria whose fibers can bind neither directly to those of the higher cell nor to a suitable lectin present in the system simply do not adhere. As a result of the specificity of adherence a higher-cell surface may often be colonized by bacteria of a single species, which may then proliferate to form a microcolony enclosed and anchored by the fibers of its component cells. At that point other bacteria that are not specific for the higher cell may adhere to the glycocalyx of the initial colonizers, forming a mixed population.

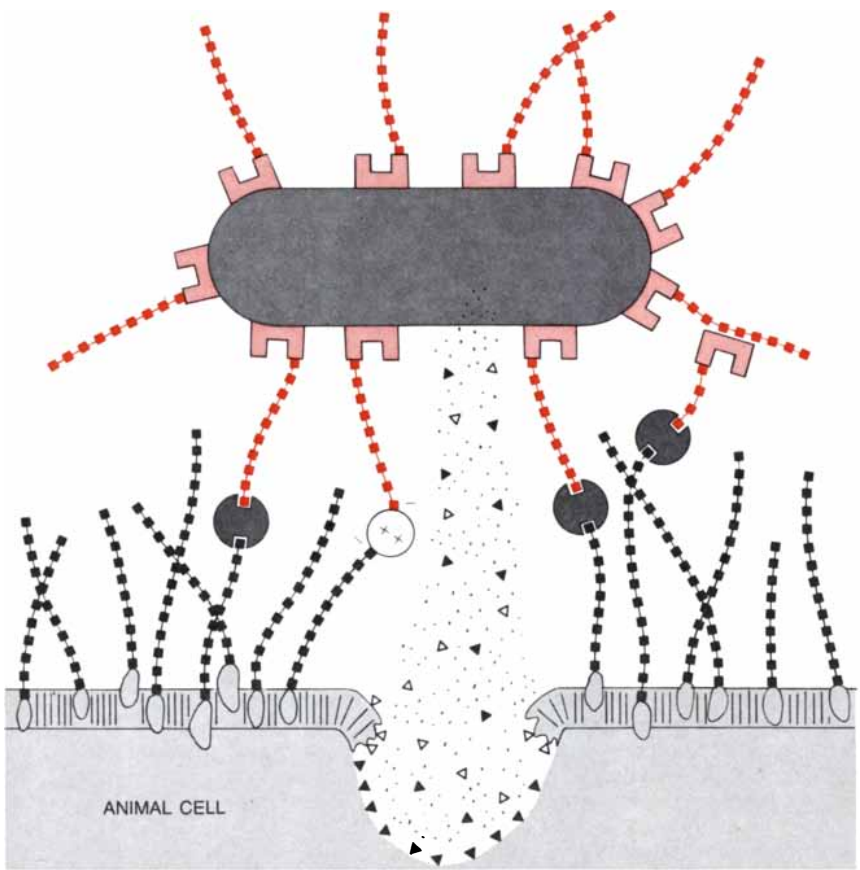
The extent to which a given bacterium will adhere to a particular tissue can be assessed by admitting the bacteria, in a suitable medium, to an experimental chamber with the tissue to be studied as the "floor." After a stated interval the tissue is examined to determine how many bacteria have adhered to the glycocalyx of the tissue cells. Andrew B. Onderdonk of the Veterans Administration Hospital in Boston and his colleagues showed that in such a situation cells of a *Bacteroides fragilis* strain with a well-developed glycocalyx adhere well to the peritoneum lining the body cavity of a rat and that mutant cells that do not produce a glycocalyx fail to adhere. Similar experiments suggest that the nature of the higher-cell glycocalyx changes as the cell ages, which may explain why some cells in the intestine of the cow may be highly colonized whereas other cells in the same tissue are not colonized at all. The recent finding that the glycocalyx is altered in cells that have been infected by a virus may explain the increase in susceptibility to bacterial disease that is often observed to follow viral infection.

If the bacteria native to a rushing stream were not adherent, the stream would be virtually sterile because the bacteria would be swept away much faster than they could swim against the current. We find that a square centimeter of a submerged surface may typically have as many as a million bacteria attached to it, whereas a cubic centimeter of the water flowing over that surface contains only 1,000 bacteria. The adaptive value of adherence in this situation is not hard to understand. The bacteria live on the organic molecules they extract from the passing water. Life in a stationary location with a continuous supply of organic nutrients, and with vigorous aeration and excellent waste removal also provided by the stream of water, clearly agrees with the bacteria: in streams enriched by pollution we have recorded populations as large as 10 billion attached cells per square centimeter. (Such a population forms a thick slimy layer on rocks in the streambed, but it also serves to remove significant amounts of organic pollutants from the water.)

Much as adhesion to a rock benefits the bacteria in a stream, adhesion to a particular tissue provides bacteria that colonize animal or plant cells with a constantly renewed supply of organic nutrients and with physical conditions conducive to growth. For example, Rolf Freter of the University of Michigan has shown that *Vibrio cholera* cells adhere specifically to the "brush border" of human intestinal cells. *Neisseria gonorrhoeae* adhere to the lining of the urethra, and other pathogen-host pairs also exhibit specific adhesion. Adhesion enables infecting bacteria to resist removal, particularly in a system that is nor-




**ANIMAL-CELL GLYCOCALYX is also a fringe of polysaccharide fibers, the branching sugar chains on the glycoprotein molecules of the cell membrane (*bottom*). The bacterial and animal-cell glycocalyxes interact as is shown here. The sugar units at the ends of the fibers can interact directly through polar attraction, the negatively charged sugars being linked by a divalent positive ion such as magnesium. The bridge may also be supplied by a lectin (*dark gray*), a protein that carries receptor sites for which the fibers' sugar units have an affinity. The interaction is specific: bacteria (*light gray*) with a different polysaccharide chemistry do not adhere.**




**BACTERIAL INFECTION may begin with the specific adhesion of a bacterium to a particular animal-tissue cell. The closely packed fibers of the two glycocalyxes provide an environment within which the bacterial toxins and enzymes (*triangles*) diffuse without loss to attack the animal-cell membrane and to digest the contents of the animal cell. The nutrient molecules (*black dots*) thus released from the animal cell are channeled back to the infecting bacterium.**

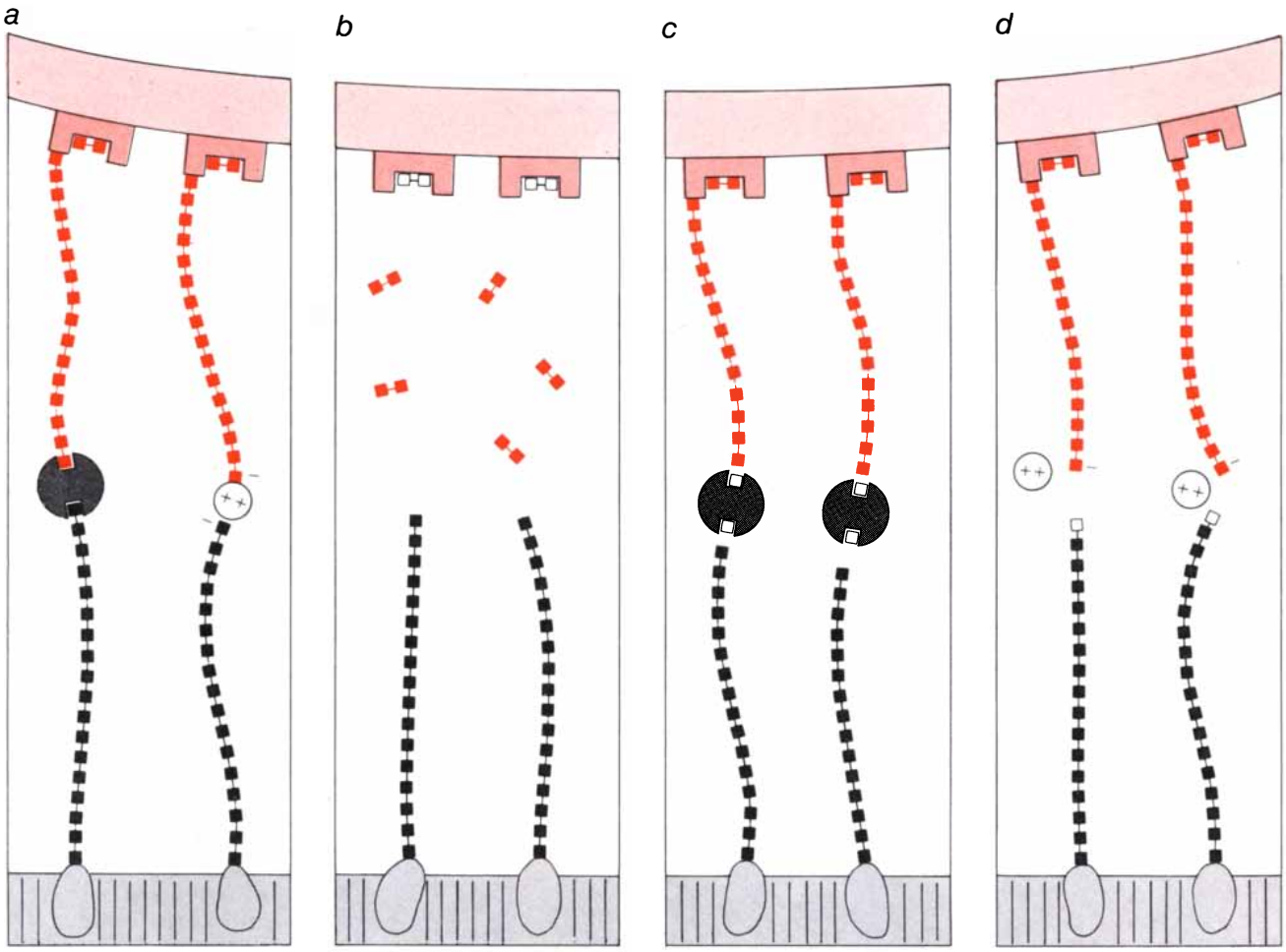
mally sterile. We find that persistent pathogens in the urinary tract are surrounded by well-developed glycocalyxes that keep the cells from being carried away in the urine. The pronounced specificity of some bacteria and viruses that attack only a particular host tissue (as *Salmonella typhimurium* attacks lymphoid patches in the intestine and as the poliovirus attacks nerve cells) may well be explained by the specificity of the glycocalyx of the host-tissue cells.

Among other examples of specificity there are starch-digesting bacteria in the rumen of the cow that adhere to grains of starch so specifically that the best way to separate such cells from the mixed bacterial population of the rumen is to recover the starch grains from the rumen fluid. Again, the luminescent bacteria that populate the light organs of deep-sea fishes such as *Photoblepharon* are selected from the vast population of marine bacteria by their ability to bind to the glycocalyx of the cells that line the cup-shaped luminous organ. In this symbiotic relation the bacteria provide a light source for the fish and in return enjoy a favorable nutritive situation, feeding on secretions of the fish.

The fibers of the glycocalyx may not only position bacteria but also conserve and concentrate the digestive enzymes released by the bacteria and direct them against the host cell. Our micrographs show that enzymes from a bacterium attached to a food source, such as cell-wall material from hay in the cow's rumen, may dig a cavity into which the bacterium slowly works its way. Before aging surface cells of the skin and similar tissues are sloughed off they may be invaded by adherent bacteria that digest pits in them. This suggests that the troublesome persistence of *Staphylococcus aureus* on the skin of health professionals could be due to the fact that these bacteria are hard to eliminate from the deep pits they dig in the skin cells from which they derive their nutrients.

A glycocalyx can also function as a food reservoir for bacteria. The polysaccharide fibers are for the most part negatively charged. Somewhat in the manner of an ion-exchange resin in the laboratory, they can bind nutrient ions and molecules that wander into the immediate environment or are produced by bacterial digestive enzymes, and thus keep them available to the cell.

In nature bacteria are subjected to many sources of stress, against most of which the glycocalyx offers protection. Attachment to a surface protects the bacteria from certain protozoans. The glycocalyx is a physical barrier against predatory bacteria and bacterial viruses, and its binding capacity traps even small harmful ions and molecules in the environment. The protective capacity of the




**PATHOGENIC ADHESION might be blocked, in order to prevent or treat infection, by a new kind of antibiotic. The adhesion of a bacterium (*top*) to an animal cell (*bottom*) by means of a polar bond or a lectin (*a*) might be disrupted in one of three ways. An analogue (*white squares*) of the units that are polymerized to form the bacterial glycocalyx might be supplied, occupying the active sites of the polymerizing enzyme and preventing the synthesis of a polysaccharide fiber (*b*). The active sites of the lectin might be blocked by a similar analogue (*c*), or a blocking agent that mimics the glycocalyx material could be supplied to block the animal-cell glycoprotein receptors (*d*).**

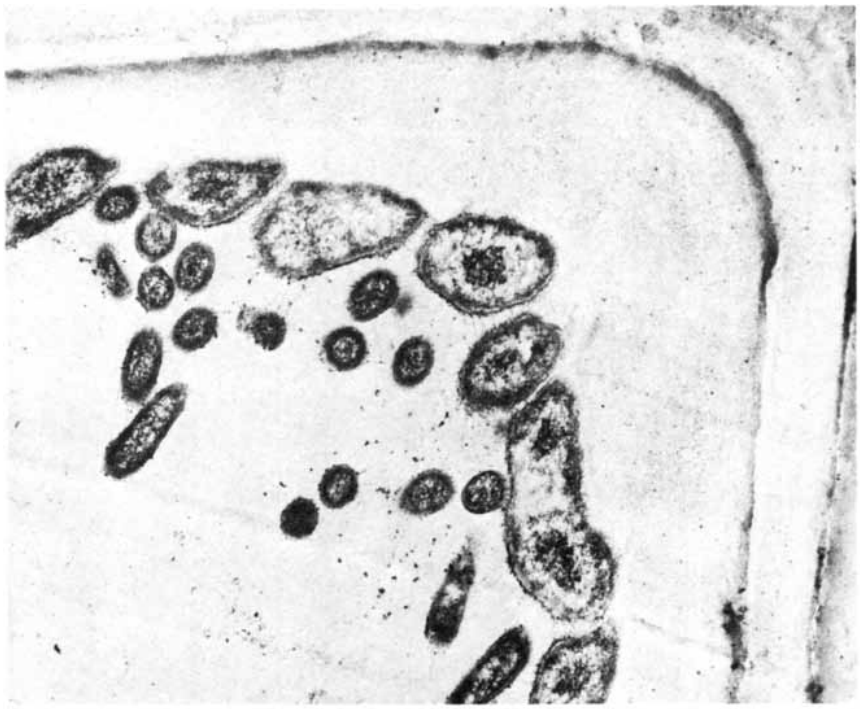
glycocalyx is of particular interest in pathogenic bacteria. We have found that single, uncoated *Pseudomonas aeruginosa* cells introduced into the lung of a rat are quickly phagocytized by the rat's white blood cells but that the defending cells cannot handle microcolonies of the same bacterium enclosed in a glycocalyx. Other experiments indicate that the glycocalyx prevents host antibodies from reacting with the surface of *P. aeruginosa* in the urinary tract.

In some diseases the manufacture of bacterial glycocalyxes, and thus the virulence of the bacteria, seems to be enhanced by excessive concentrations of certain nutrients; this may be the case in cystic fibrosis, in which there may be an excess of sodium and magnesium ions in the affected lung. Some of our preliminary results indicate that antibiotics may not be able to overcome the binding capacity of the glycocalyx in order to reach their bacterial target. Since infecting bacteria in many diseases are surely grouped in microcolonies protected by glycocalyxes, it is clear that antibiotics should be designed with the penetration of those thickets of polysaccharide in mind.

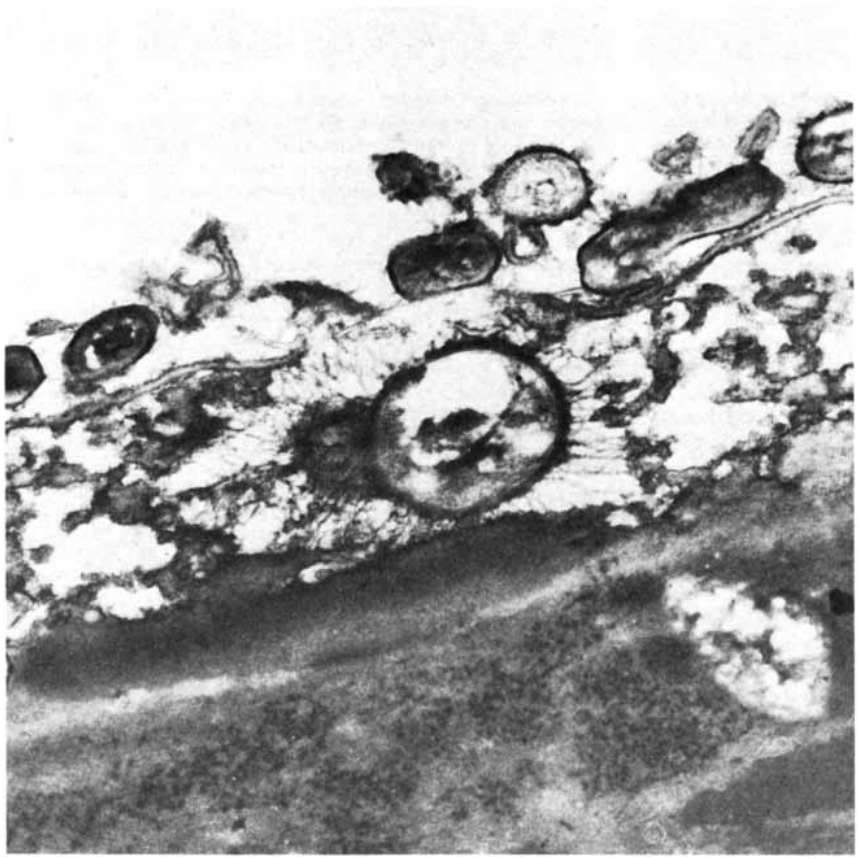
In addition to maintaining bacteria in an advantageous position and protecting them the glycocalyx may at times group bacteria in something approaching an organized community. In lake sediments and in other ecosystems two or more species of bacteria may act together in a "consortium" to carry out physiological processes. Ralph S. Wolfe of the University of Illinois at Urbana-Champaign has described consortiums in which one species of bacterium releases hydrogen from organic compounds and passes it on to another species, which uses it to reduce carbon dioxide to methane ($CH_4$). A physiological transfer of this kind requires that the members of the consortium be held close together, and that may be accomplished by chemical affinity between the glycocalyxes of the associated cells. Such a physiological consortium seems to be represented in some micrographs we have that show bacteria digesting the cellulose of a plant-cell wall while bacteria of another species are ranged along the cellulose-digesting bacteria's "free" sides.

These bits of evidence suggest a community structure, with cells of a particular species adhering in a favorable niche close to the source of a necessary nutrient. Such adherent populations could respond with unique plasticity to changes in nutrient conditions or in other environmental conditions: bacteria in the "wrong" niche would simply die, leaving space and nutrients for other bacteria more suited to the location. The entire mixed population of cells would have some of the characteristics—but none of the vulnerability—of a multicellular higher organism.

If adhesion has a central role in the success of pathogenic bacteria, then the prevention of adhesion should be an effective way to prevent or combat bacterial infection. It should be possible to develop a new class of antibiotics that interfere with glycocalyx formation or function in specific pathogens. There are at least three ways in which such inhibition might be achieved. One way would be to disrupt the synthesis of glycocalyx



**CELLULOSE-DIGESTING BACTERIA (*larger cells*) adhere to the wall of a plant cell in animal feed. An associated population of smaller bacteria adheres to the large ones, presumably through a specific interaction of the two bacterial glycocalyxes. The two bacterial species may be an example of a "consortium," a cooperative association in which a metabolic product of one bacterial population provides a nutrient that is required by the second population.**



**DEAD CELL at the surface of an animal tissue has been invaded by an adherent bacterium, which has broken through the cell membrane and has sunk into a cavity produced by its own digestive activity. A population of other bacteria adheres to the animal cell's membrane.**

fibers. The bacterial polymerase that links sugar molecules to form these fibers should be inhibited if it is presented with a compound that mimics its normal substrate and therefore occupies the enzyme's active site, but that cannot be processed to build the normal polysaccharide fiber. In the absence of such fibers there would be no glycocalyx, no adhesion and no resistance to white cells. One might also find a compound that would occupy and block the active site of a lectin mediating the adhesion of bacterial glycocalyx fibers to the fibers of host cells.

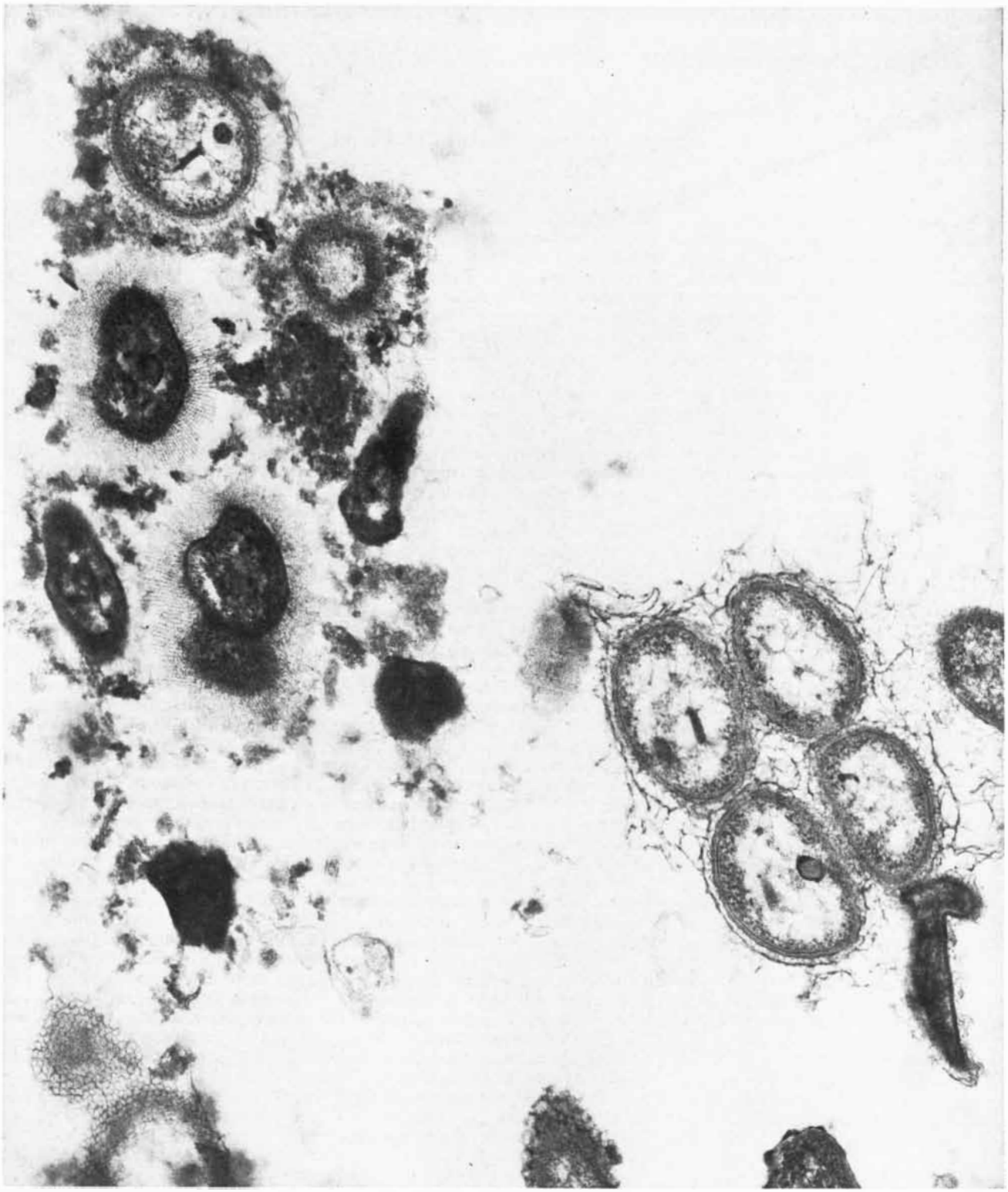
Finally, it should be possible to block the "receptor" sites on host cells, that is, the glycoprotein fibers to which bacterial fibers adhere directly. One attractive aspect of an antibiotic directed against the glycocalyx is that it need not enter the host cells or the bacterial cells, thereby avoiding two common problems in antibiotic therapy: toxicity to host cells and the induction of bacterial resistance based on changes in the permeability of the bacterial-cell membrane.

Perhaps the best initial approach would be the last one listed above. Since polysaccharides are rather simple chains of sugar molecules and since specific adhesion depends on the chemical affinity of bacterial polysaccharides for host-cell polysaccharides, adhesion should be blocked by bits of the glycocalyx material, by its sugar subunits or by chemical analogues of those subunits. Onderdonk has already found that treating mouse peritoneal tissue with glycocalyx material from *Bacteroides fragilis* prevents the attachment of *B. fragilis* cells to the mouse tissue; he attributes this effect to the blockage of the receptor sites on the mouse cells by the glycocalyx polysaccharide. Other workers have reported a similar effect in the crop of the chicken, which normally comes to be covered by a layer of lactobacilli. Treating the crop of germ-free chickens with the polysaccharide of the lactobacillus glycocalyx prevents the formation of the bacterial layer.

Presumably the cells of the human throat have receptors for the streptococci that cause "strep throat" and its more serious aftereffects, including rheumatic heart disease. It should be possible to block those receptors with simple sugar molecules that are analogues of the polysaccharide units of the streptococcal glycocalyx. The task of identifying those polysaccharides and developing analogues would be particularly worthwhile in the case of streptococcal infection and other diseases where repeated infections can have serious effects. The practice in such conditions is to avoid reinfection by long-term prophylactic treatment with a conventional antibiotic, and that can be dangerous. The antibiotic kills many of the nonpathogenic

adherent bacteria that are normally present and that occupy many of the receptor sites, thus opening the way to colonization by drug-resistant strains of the pathogenic bacteria, strains that tend to be selected by the very treatment designed to suppress the pathogen. An antiglycocalyx therapy would make it unnecessary to maintain the patient on a conventional antibiotic.

The development of new ways to control bacterial infections will require a much more detailed knowledge than we now have of the polysaccharide constituents of particular pathogens and host-tissue cells. That will take time. The basic perception that bacteria colonize their environments by adhering tenaciously and specifically is a first step toward an ability to manipulate and control bacteria more effectively.



**FIBERS OF GLYCOCALYX display a wide variety of patterns in electron micrographs of thin sections. The fibers of this mixed population from a cow's rumen form (*clockwise from lower right*) patterns designated random, honeycomb, radial and radial-concentric.**

# The Efficiency of Algorithms

*Some mathematical problems can be solved only by methods too slow for even the fastest computers. More efficient methods have not been found, but neither has it been proved that there are no better methods*

by Harry R. Lewis and Christos H. Papadimitriou

Suppose you were asked to plan an itinerary for a traveling salesman who must visit a number of cities. You are given a map on which the distances between the cities are marked and you are asked to find the shortest route that passes through all the cities and returns to the starting point. An approach to this problem that is certain to give the correct answer is to trace all the possible routes, measure their length and pick the shortest one. If the tour included more than a few cities, however, hundreds or thousands of routes would have to be checked. If there were 100 cities, then even the fastest computers would require weeks of calculation to find the shortest path.

In the search for a quicker solution you might try some less rigorous methods. One idea that seems reasonable is always to visit nearby cities before going on to those farther away. You would soon discover, however, that this procedure does not invariably yield the correct answer. Other shortcuts also fail. In fact, the best methods known for solving the problem are not much better than the obvious but laborious procedure of checking all the possible itineraries. Mathematicians now suspect that this problem and many others like it may forever remain beyond our ability to solve in any efficient way. That speculation itself, however, is unconfirmed; although no faster methods of solution have been found, neither has it been proved that faster methods do not exist.

In the problem of the traveling salesman's tour it is not the solution for a particular set of cities that is of the greatest importance but a general method for finding the solution for any cities. Such a method is called an algorithm; it is a precisely stated procedure or set of instructions that can be applied in the same way to all instances of a problem. If the problem is to be solved with the aid of a computer, an algorithm is indispensable, because only those procedures that can be stated in the explicit and unambiguous form of an algorithm can be presented to a computer. Instructions that are vague or that rely on intuition are unacceptable.

An example of an algorithm is the procedure taught in the schools for the subtraction of whole numbers. If each of the steps in this procedure is applied correctly one at a time, the algorithm will always yield the correct result. What is more, once the algorithm has been learned or stored in the memory of a computer or embodied in the circuitry of an electronic calculator, it can be applied to an infinite set of subtraction problems. With this one algorithm the difference between any two whole numbers can be determined.

In principle any problem for which an algorithm can be devised can be solved mechanically. It may therefore seem surprising that there are problems for which algorithms exist but for which we so far have no practical general solution. The algorithms for solving these problems always give a correct answer, but they often require an inordinate amount of time. The problem of the traveling salesman's tour is among these intractable tasks.

The efficiency of computer algorithms is a topic of obvious practical importance. It is also of interest in more formal areas of mathematics. There are some problems in mathematics and logic for which no algorithm can ever be written, and there are many others for which efficient, fast algorithms are already known. Between these two groups is a third class of problems that can always be solved in principle but for which there are only inefficient (and therefore largely unusable) algorithms. For some of these difficult problems mathematicians have been able to demonstrate that efficient algorithms can never be designed. For many of the most important problems, however, there is only the suspicion that good algorithms are impossible.

A given problem can have more than one algorithm for its solution. For example, children in Europe learn a procedure for subtraction slightly different from the one taught in the U.S. Both of the subtraction algorithms, however, give the same result in the same amount of time. That is not invariably the case with different algorithms for solving the same problem. One celebrated problem that can be solved by either a "fast" algorithm or a "slow" one is the problem of the Königsberg bridges.

In the 18th-century German city of Königsberg (which is now the Russian city of Kaliningrad) a park was built on the banks of the river Pregel and on two islands in the river. Within the park seven bridges connected the islands and the riverbanks. A popular puzzle of the time asked if it was possible to walk through the park by a route that crossed each of the bridges once and only once.

For the solution of the problem the size and shape of the islands and the length of the bridges are immaterial; the only essential information is the pattern of interconnections. This information can be presented compactly in the mathematical structure known as a graph, which is merely a set of points with lines drawn to join them. In the case of the Königsberg park each of the riverbanks and each of the islands is condensed to a single point and each of the bridges is represented by a line between two points. Thus the graph consists of four points and seven lines. If the lines are labeled, any path through the park can be specified by a simple listing of labels.

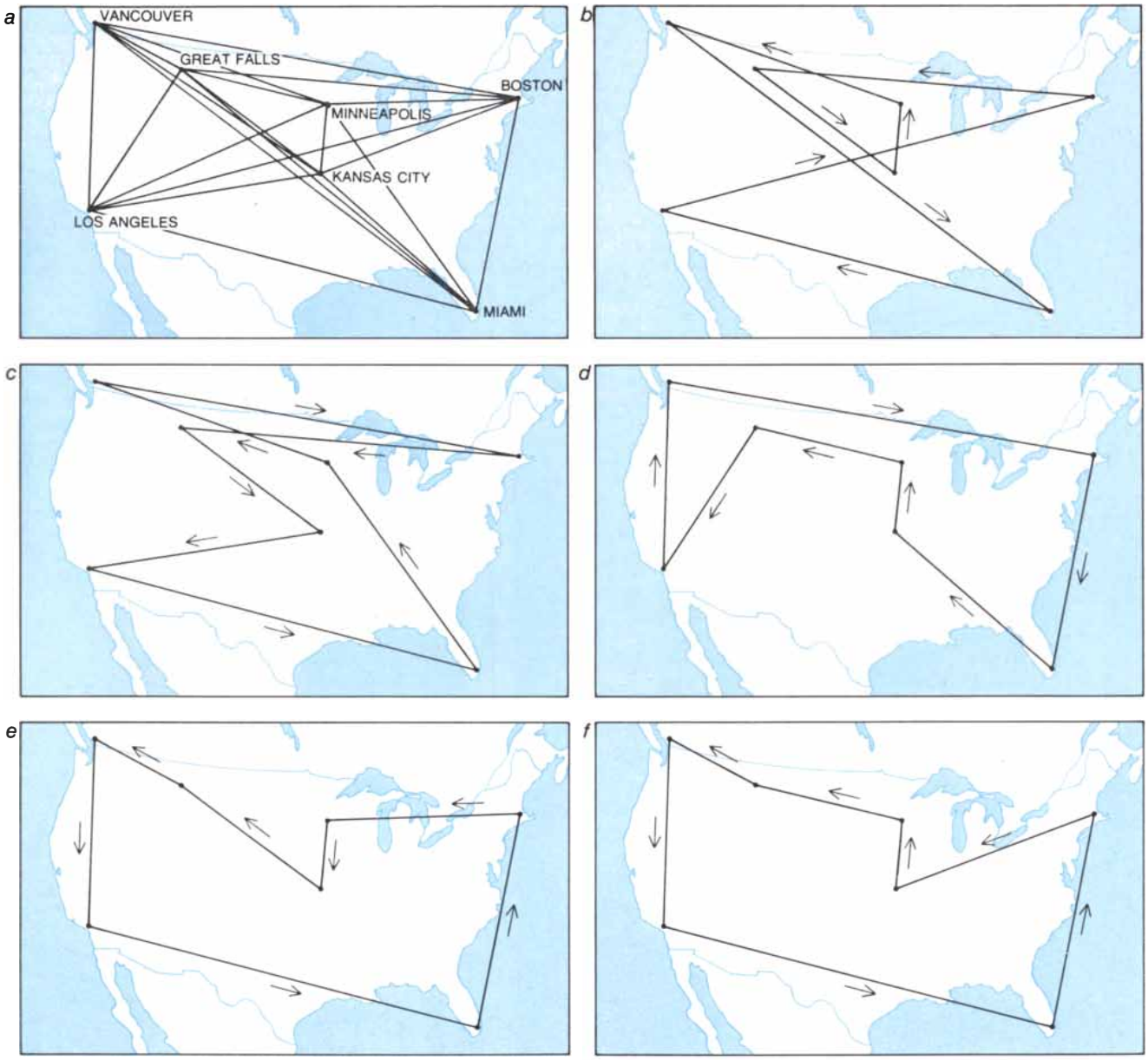
The obvious approach to the problem is to list all the paths that cross all the bridges and to eliminate from consideration those that cross any bridge more than once. This is the technique of exhaustive search, similar to the one employed in the problem of the traveling salesman. When the mathematician Leonhard Euler was presented with the problem of the Königsberg bridges, he recognized the limitations of the technique and found another method. In recognition of his contribution a path that traverses each line of a graph exactly once is now called an Eulerian path.

Euler wrote: "The particular problem of the seven bridges of Königsberg

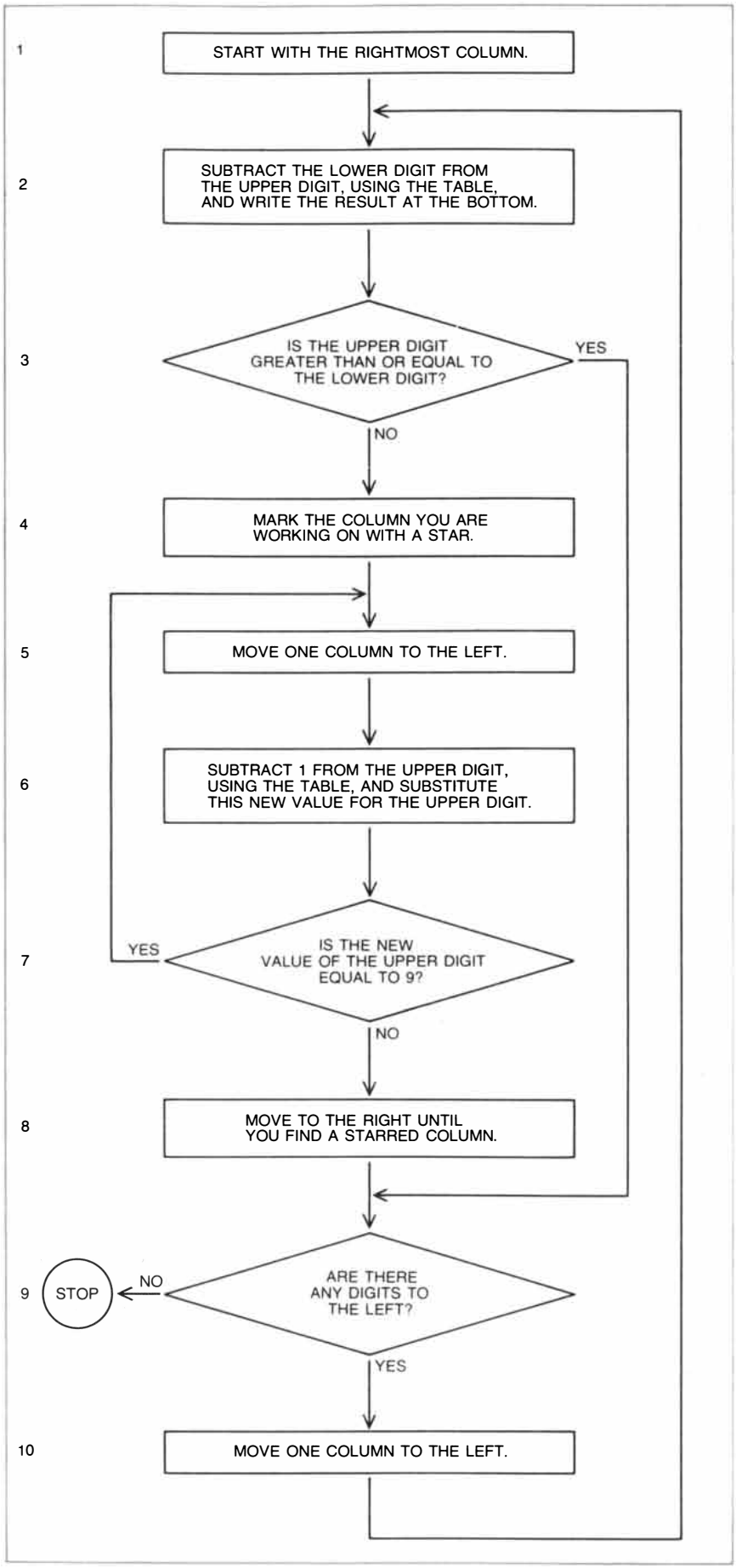
could be solved by carefully tabulating all possible paths, thereby ascertaining by inspection which of them, if any, met the requirement. This method of solution, however, is too tedious and too difficult because of the large number of possible combinations, and in other problems where many more bridges are involved it could not be used at all."

Euler's alternative method is much simpler. He showed that a tour of the kind sought must exist if the graph meets two conditions. First, it must be possible to go from any point in the graph to any other point by following the lines of the graph; in other words, the graph may not be disconnected. Second, every point of the graph, with two possible exceptions, must be at the junction of an even number of lines.

It is not hard to understand why a graph cannot have an Eulerian path unless it meets these conditions. All regions of the graph must be connected to one another if there is to be any path that traverses all the lines. Each point must have an even number of lines because half of them are required to reach the point and the other half to leave it. Two points with an odd number of lines can be allowed if they are chosen as the starting and finishing points of the path. Demonstrating that any graph




**TRAVELING SALESMAN'S TOUR is a commonplace problem that reveals a deep deficiency in the methods available to mathematics. The problem can be solved, but only by means so arduous and time-consuming that the solution is generally impractical. The problem is, given a map showing the airline routes connecting several cities (*a*), to find the itinerary for the shortest round-trip tour of the cities. One approach to the problem that works consistently is to plot all the possible tours, measure their length and choose the shortest one. Even with only seven cities, however, there are more than 350 tours, and the number increases rapidly (as the factorial of the number of cities). Several shortcut methods might be attempted. For the salesman always to fly to the city farthest from the one where he is clearly gives a wrong answer; the resulting tour (*b*) is among the longest tours possible. Even an arbitrary route, such as one visiting the cities in alphabetical order (*c*), gives a better result. Dividing the country into segments and visiting first the East, then the Middle West and then the West gives a still shorter tour (*d*). Finally, the salesman might always fly to the nearest city, a procedure that leads to two reasonable itineraries, the difference between them depending on whether he begins at Minneapolis (*e*) or Kansas City (*f*). Even those routes are not the shortest possible. (The optimum tour is shown in the illustration on page 107.) The kind of solution sought is an algorithm, or set of instructions, that will find the shortest tour for any group of cities. No known algorithm is significantly better than exhaustive search.**




that meets these conditions actually has an Eulerian path requires a somewhat more complicated argument, which we shall not present here, but Euler was able to give a rigorous proof.

It is an easy matter to express Euler's solution to this problem in an algorithm that could be executed by a computer. The first requirement, connectivity, can be established by marking some point of the graph, then similarly marking all points connected to it by lines, then all the points connected to the newly marked points, and so on. The graph is connected if at the end all points have been marked. The second requirement is just as easily tested: the machine is instructed to examine each point of the graph and count the number of lines that terminate at that point. If no more than two of the points have an odd number of lines, the graph has an Eulerian path. The park at Königsberg met the first condition but not the second, and so there was no Eulerian tour of the seven bridges.

Euler's method is unquestionably the more economical approach to the problem of the Königsberg bridges: it requires that each point and line of the graph be listed just once, whereas the exhaustive search is not completed until every path that crosses all the bridges has been listed. The number of such paths is much larger than the number of points and lines in the graph. In that sense Euler's method is the better algorithm, but how much better? How can

DIMINUEND

| SUBTRAHEND | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| 1 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| 2 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 3 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 4 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 |
| 5 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 |
| 6 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 |
| 7 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 |
| 8 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 |
| 9 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 |

```
  2 * 3 *
  3 0 4 7   DIMINUEND
- 1 4 3 8   SUBTRAHEND
  1 6 0 9   DIFFERENCE
```

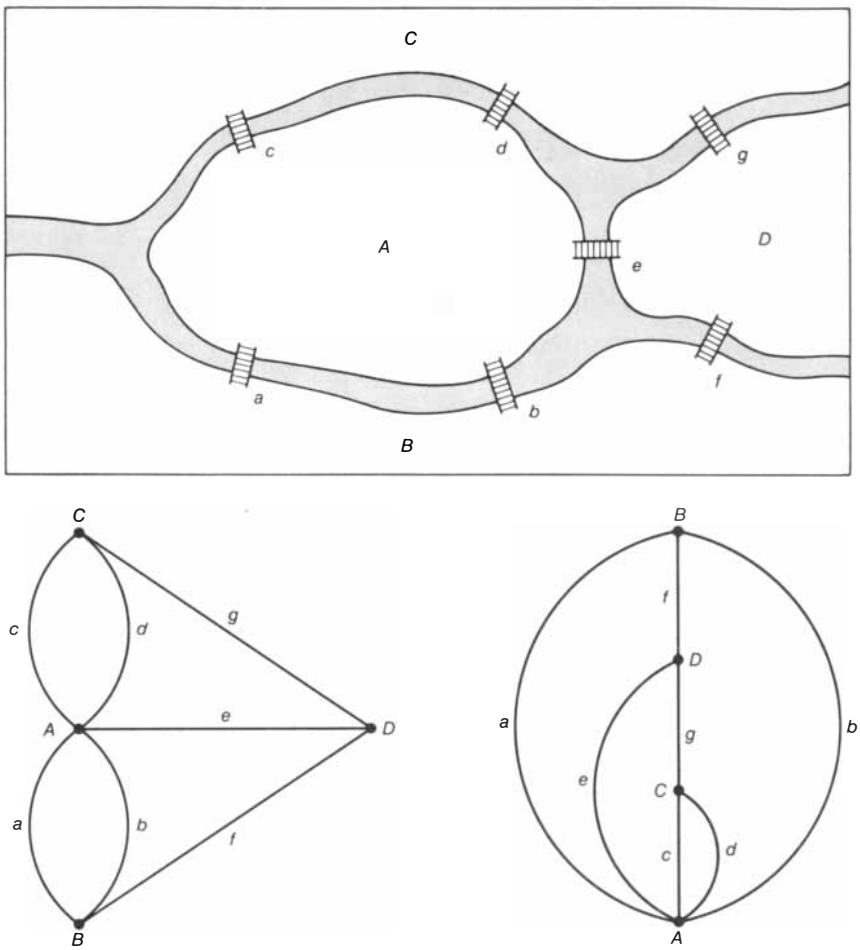
**ALGORITHM for the subtraction of whole numbers defines an explicit procedure that can be followed without any need for intuition and even without an understanding of the significance each step has to the operation as a whole. The algorithm, which is assumed to incorporate the table of differences shown here, can be applied to an infinite number of subtraction problems. Other algorithms are equally effective. A method of subtraction taught in European schools, for example, differs in the treatment of borrowing where it specifies that 1 should be added to the lower digit instead of being subtracted from the upper one.**

the difference be measured, and how can one tell if the difference is significant?

For a graph with only four points and seven lines both techniques are fast enough to be considered practical. Suppose, however, that more islands and bridges were added to the park, or in other words that more points and lines were added to the graph. If the problem is being solved by Euler's method, each new point merely adds one item to the list of points that must be checked. If the paths are to be examined by exhaustive search, on the other hand, then with each new point and line the size of the list is multiplied by some factor. A moderate increase in the size of the graph results in an explosive increase in the number of paths. Ultimately the list of paths must become prohibitively long.

In this comparison of the two solutions to Euler's problem there is the basis for a completely general method of evaluating the speed or practicality or efficiency of any algorithm. We imagine that the algorithm is supplied with larger and larger inputs, and we note the rate at which the execution time of the algorithm increases. In this way it is possible to make unequivocal judgments of algorithms. Exhaustive search not only is a slower method; in general it is too slow to be of any value. Euler's method remains practical for problems of essentially unlimited size.

As the size of the graphs being examined increases, the lists produced by the method of exhaustive search grow exponentially. Each time some fixed number of points and lines are added to the graph the size of the list doubles. Growth of this kind can be described by a mathematical function such as $2^n$, where $n$ is some measure of the size of the graph. Many other functions have similar or even higher rates of growth. Among them are $n^n$ and $n!$ (which is read as "$n$ factorial" and signifies $n$ multiplied by all the integers between 1 and $n$). For the purposes of this discussion all these functions can be regarded as having the same property of exponential growth.




abcdef abcdeg abcged abcgf abdcef abdceg abdgec abdgf
acdbfe acdbfg acdefb acdeg acgebf acged acgfbd acgfbe
adceg adgebf adgec adgfbc adgfbe aefbcd aefbcg aefbdc
aegdbf aegdc abegcd abegdc abef adcbfe adcbfg adcefb
aefbdg aegcbf aegcd

A: abcde = 5
B: abf = 3
C: cdg = 3
D: efg = 3

**EULER'S PROBLEM asks whether there is a path through a graph that traverses each line exactly once. In this context a graph is defined as any collection of points with lines connecting them. The problem was first stated in terms of a walking tour through a park in the 18th-century German city of Königsberg (*top*); the path sought was required to cross each of the seven bridges in the park exactly once. The park can be represented as a graph in at least two equivalent ways. One approach to the problem is to list all the paths through the graph that continue as far as they can without repeating a line. Even for a small graph, however, there are many such paths; the sample listing shown includes only those paths that begin with line *a*. A more efficient algorithm was discovered by Leonhard Euler. A graph has a path that traverses each line once, he showed, if every point of the graph (with two possible exceptions) is at the junction of an even number of lines. Such a path is now called an Eulerian path. Counting the lines meeting at each point shows that the graph of the Königsberg park has no Eulerian path.**

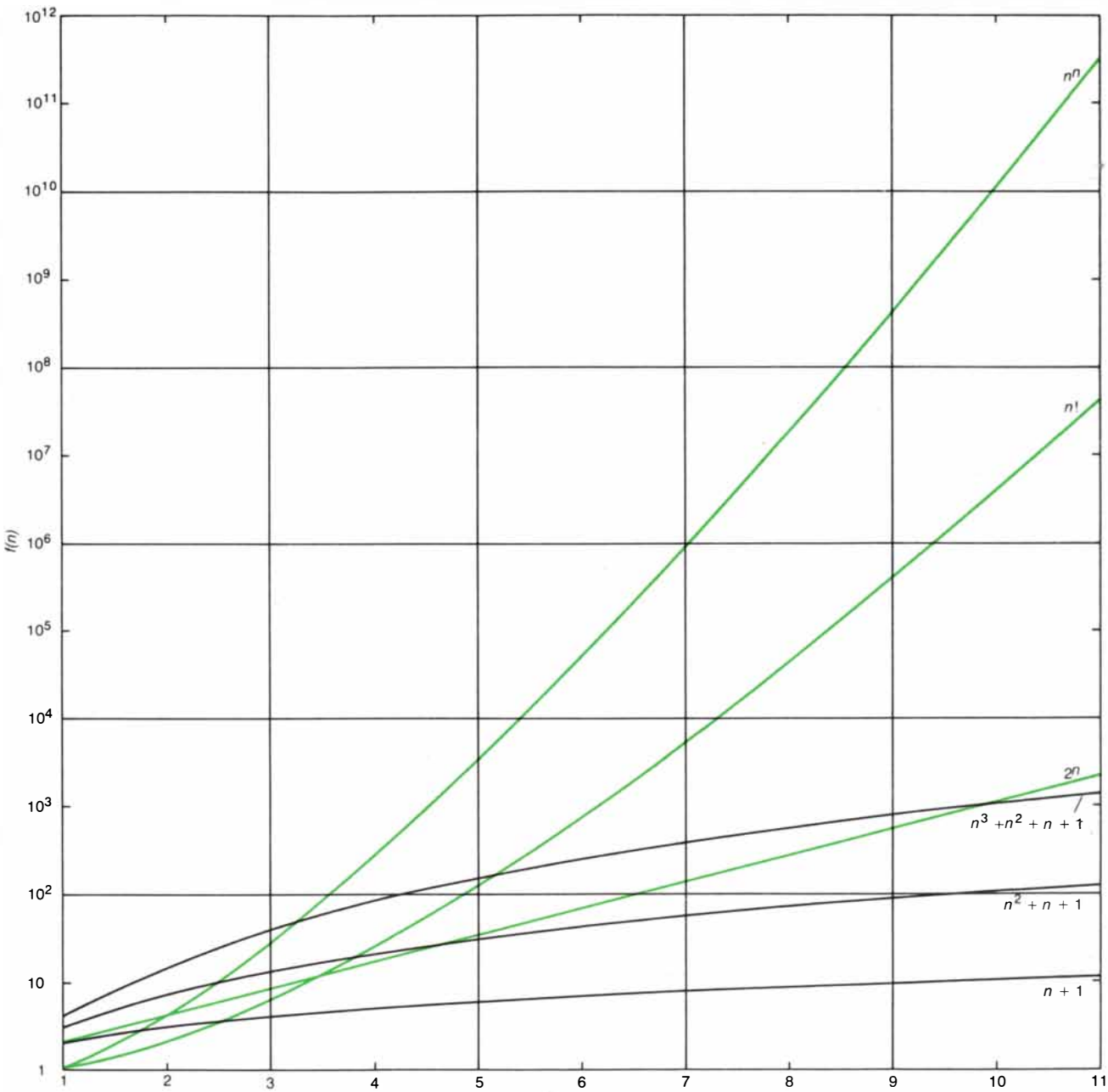
Mathematical functions of another kind are known as polynomials. The simplest members of this class are linear functions, such as $3n$, which designate a simple relation of proportionality. The time needed to solve the problem of the Königsberg bridges by Euler's method increases as a linear function of the size of the graph. Other polynomials are $n^2$, $n^3$ and so on, and the sums of such functions. What distinguishes polynomials from exponential functions is that $n$ never appears in an exponent.

For sufficiently large values of $n$ any exponential function will overtake and exceed any polynomial function. It was the certainty of this result that dismayed Thomas Malthus when he compared the exponential rate of population growth with the polynomial rate of increase in the food supply. For small values of $n$ a given polynomial function may well exceed a given exponential one, but there is always a value of $n$ beyond which the exponential function is the greater. The exact form of the polynomial makes little difference, except in changing the point at which the polynomial function is overtaken.

It is now generally agreed by computer scientists that algorithms whose execution time increases exponentially as a function of the size of the input are not of practical value. We shall call algorithms of this kind "exponential time" algorithms, or simply inefficient algorithms. The only algorithms that are considered fast or efficient enough for general application are "polynomial time" algorithms.

Of course, even among efficient algorithms some are faster than others, but for the purposes of this discussion it is important only to distinguish polynomial-time algorithms as a class from exponential-time algorithms. Moreover, this system of classification has the advantage that it makes the speed of an al-

| FUNCTION | TYPE | $n = 1$ | $n = 2$ | $n = 3$ | $n = 4$ | $n = 5$ | $n = 6$ | $n = 7$ | $n = 8$ | $n = 9$ | $n = 10$ | $n = 11$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $f(n) = n + 1$ | POLYNOMIAL | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| $f(n) = n^2 + n + 1$ | POLYNOMIAL | 3 | 7 | 13 | 21 | 31 | 43 | 57 | 73 | 91 | 111 | 133 |
| $f(n) = n^3 \quad n^2 + n + 1$ | POLYNOMIAL | 4 | 15 | 40 | 85 | 156 | 259 | 400 | 585 | 820 | 1,111 | 1,464 |
| $f(n) = 2^n$ | EXPONENTIAL | 2 | 4 | 8 | 16 | 32 | 64 | 128 | 256 | 512 | 1,024 | 2,048 |
| $f(n) = n!$ | EXPONENTIAL | 1 | 2 | 6 | 24 | 120 | 720 | 5,040 | 40,320 | 362,880 | 3,628,800 | 39,916,800 |
| $f(n) = n^n$ | EXPONENTIAL | 1 | 4 | 27 | 256 | 3,125 | 46,656 | 823,543 | 16,777,216 | 387,420,489 | 10,000,000,000 | 285,311,670,611 |




**RATE OF GROWTH in the execution time of an algorithm as the size of the problem increases determines the practicality of the algorithm. The rates of growth are described by mathematical functions that can be classified as either polynomial or exponential. The values of $n$ in the table at the top and in the graph at the bottom represent some measure of the size of a problem; the values of the functions correspond to execution time. For any exponential function and any polynomial one there is a value of $n$ beyond which the exponential function is always the greater. For this reason algorithms whose execution time increases as an exponential function of the problem size are considered inefficient and in general are of no practical value. Algorithms whose execution time increases as a polynomial function, in comparison, are efficient. In the problem of the Königsberg bridges the execution time for Euler's method increases as a polynomial function such as $n + 1$; the method of exhaustive tabulation requires an amount of time given by an exponential function such as $2^n$.**

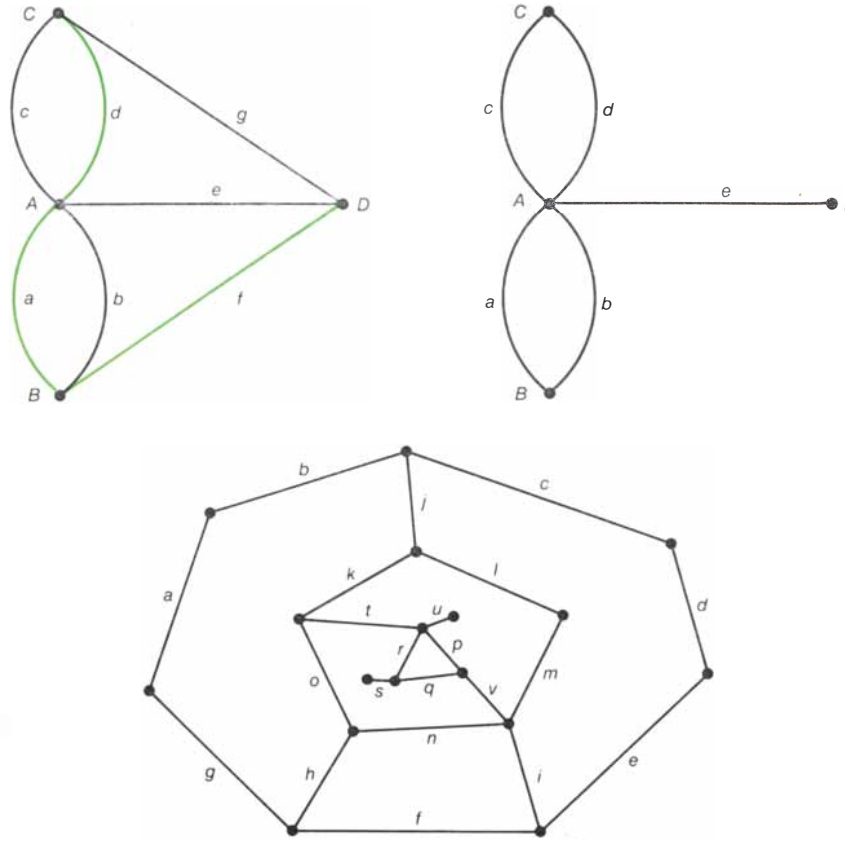
gorithm a property of the algorithm itself and independent of the machine on which it is executed. For sufficiently large problems a polynomial-time algorithm executed on even the slowest machine will find an answer sooner than an exponential-time algorithm on the fastest computer.

The mathematical properties of algorithms were studied in the 1930's by the British mathematician A. M. Turing, the inventor of the imaginary computer now called a Turing machine. The Turing machine was conceived to be an automaton equipped with an infinite supply of paper tape marked off in square regions. The machine was capable of just four actions: it could move the tape one square, it could place a mark in a square, it could erase a mark already present and at the end of a calculation it could halt. These operations were to be performed according to a sequence of instructions built into the internal mechanism. Of course, Turing never built such a machine; it was merely a conceptual device for automatically solving problems in mathematics and logic. Indeed, Turing was interested not in actually solving problems with the machine but rather in investigating what kinds of problems it could solve and what kinds it could not.

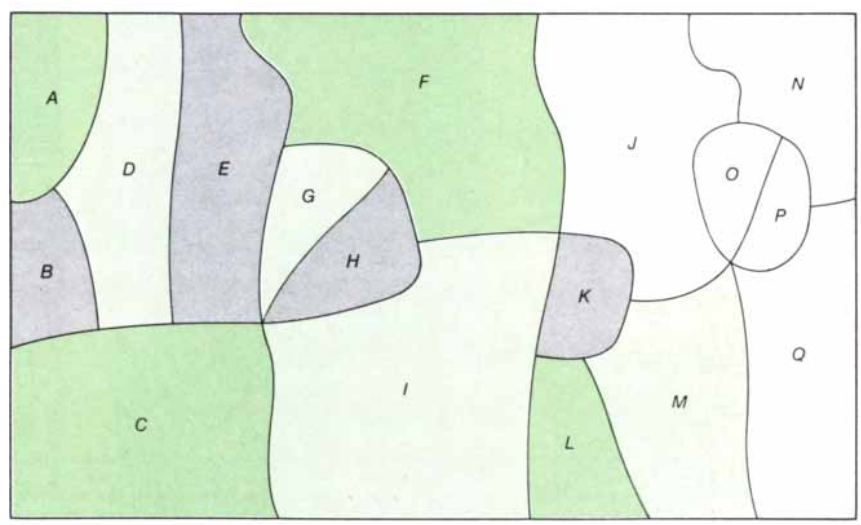
Turing discovered that even a machine as simple as this one could solve any problem for which an algorithm could be devised. The computation might be laborious and indirect, but given enough time and paper tape the machine would eventually find the solution and halt. Reduced to its essentials the Turing machine is a language for stating algorithms, as powerful in principle as the more sophisticated languages now employed for communicating with computers.

In addition to conceiving these machines Turing demonstrated their limitations. In 1936 he showed that there are problems that cannot be solved by Turing machines, and it follows that these problems cannot be solved by any automatic computer. They are problems for which algorithms cannot be written, even in principle. The example first studied by Turing is the problem of predicting whether a particular Turing machine, once it is set in motion, will ever finish its calculation and halt. Through an analysis of this problem he was able to show that there can be no general procedure for telling whether mathematical propositions are true or false. Since then a variety of other problems with the same properties have been proposed.

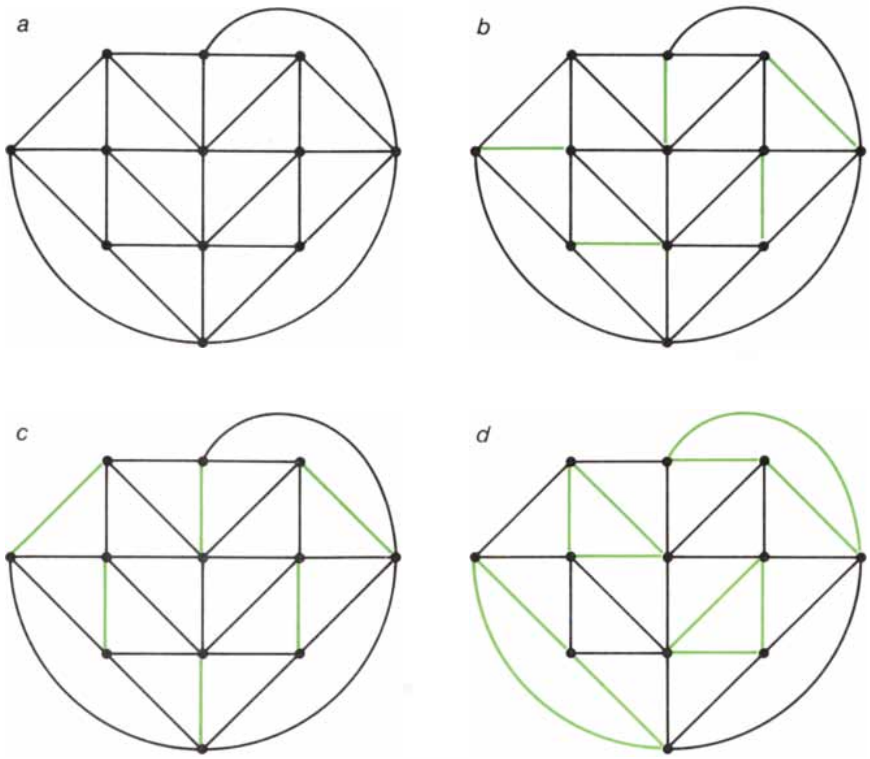
One result of Turing's work was to divide all imaginable problems in mathematics into two classes. Those problems for which algorithms can never be written are in a formal sense permanently unsolvable. Some instances of these problems may be solved by rare percep-



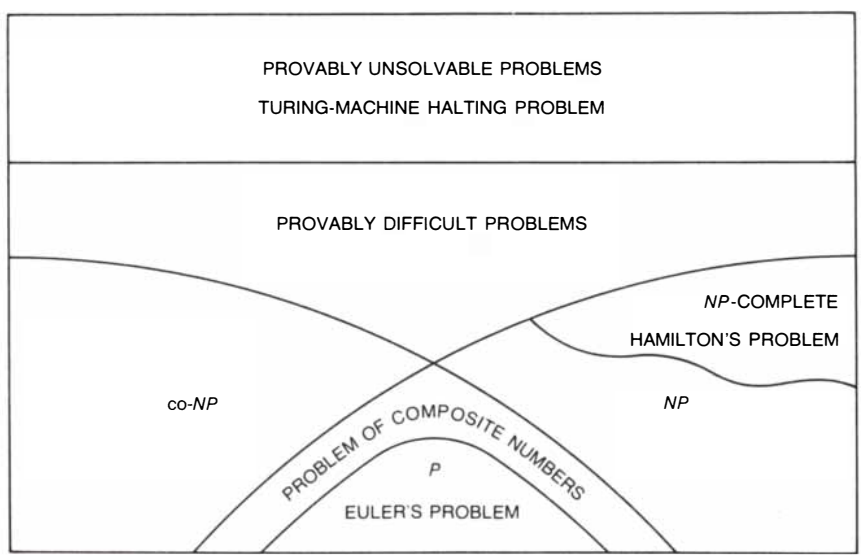

**HAMILTON'S PROBLEM, formulated by the Irish mathematician William Rowan Hamilton, resembles Euler's problem superficially but asks whether there is a path through a graph that touches each point (instead of each line) exactly once. The graph derived from the park at Königsberg (*top left*) has a Hamiltonian path (*color*) although it has no Eulerian path. By removing two lines a graph is formed (*top right*) that has an Eulerian path (beginning at point *D* and proceeding, for example, along lines *e, a, b, c* and *d*) but not a Hamiltonian path. The third graph (*bottom*) has neither property. No efficient algorithm for solving Hamilton's problem is known; the available methods are fundamentally no better than exhaustive search.**




**THREE-COLOR-MAP PROBLEM asks whether three colors can be applied to the regions of a map so that no two regions sharing a border have the same color. The problem can be solved for any particular map by examining all the possible colorings, but such a procedure is extremely tedious; for the map shown there are $3^{16}$, or about 43 million, possible colorings. No efficient algorithm for solving the problem is known, but where a three-color solution does exist it can be found in principle by guessing (a nonalgorithmic method). If there is no solution, guessing is of no value. The trial solution shown fails on reaching country *J*, which communicates with all three colors and therefore cannot be assigned any of them. The failure does not prove, however, that no three-color solution exists; indeed, a solution can be found fairly quickly by exchanging the colors of regions *K* and *L* and continuing the series of guesses in the rest of the graph.**




**ASSIGNMENT OF ROOMMATES in a dormitory is a problem that can be solved efficiently if two students are to share each room but not if there are to be three in a room. The problem can be stated in the form of a graph in which each student is represented as a point, and a line is drawn between two points if those students would be compatible roommates (*a*). Pairs of students can be assigned to rooms by a technique called augmentation: an incomplete assignment is made (*b*), then a complete one is found (if it exists) by a sequence of rearrangements (*c*). There is no equivalent efficient procedure for finding groups of three compatible students. Once an acceptable room assignment has been found, however, it can readily be exhibited (*d*).**




**CLASSIFICATION OF PROBLEMS derives from present conjectures about the existence and nonexistence of efficient algorithms. The "provably unsolvable" problems are those for which there are no algorithms of any kind; the "provably difficult" ones have only exponential-time algorithms. Problems with known polynomial-time algorithms are assigned to the class *P* (for polynomial). The status of the remaining problems is less certain: only exponential-time algorithms are known for them, but it has not been proved that efficient algorithms do not exist. These problems are assigned to the classes *NP* and co-*NP*. The letters *NP* stand for "nondeterministic polynomial" and signify that the problems can be solved quickly by guessing. Co-*NP* includes the yes-or-no problems whose complementary no-or-yes versions are in *NP*. *NP*-complete is a subset of the class *NP* made up of problems with a special property: if any one of them could be solved by an efficient algorithm, then all other problems in the class *NP* could also be solved efficiently. The discovery of such an algorithm would constitute a proof that *P* and *NP* are identical, but most mathematicians believe such an algorithm does not exist.**

tion or by luck, but a general method for their solution will never be found.

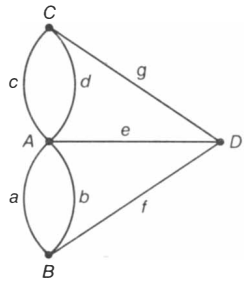
All other problems in mathematics and logic can be solved by algorithms. As we have seen, however, some algorithms are more useful than others. The class of solvable problems can therefore be divided into two subgroups: those for which there are efficient, polynomial-time algorithms and those for which there are only exponential-time algorithms. Euler's problem is a member of the class with polynomial-time solutions, since Euler's method is itself a polynomial-time algorithm. Problems that can be proved to have only exponential-time algorithms are also known, although they are rather obscure.

Although these two groups of problems are quite distinct, it is not always a straightforward task to assign a problem to one group or the other. Indeed, a very interesting class of problems seems to fall somewhere between them. For these problems no efficient algorithms are known and the best available solutions require exponentially increasing time, yet no one has been able to prove that the problems do not have polynomial-time solutions.

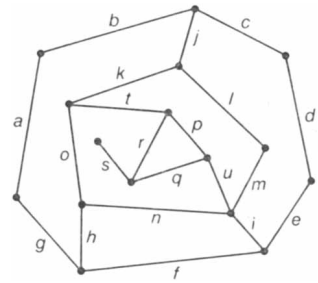
One such problem was considered in the 19th century by the Irish mathematician William Rowan Hamilton. Superficially Hamilton's problem is much like Euler's. The problem is to decide whether a given graph has a path that takes in each point exactly once (whereas Euler looked for a path that traversed each line once). Actually the tasks are quite different, and Euler's method cannot be applied to Hamilton's problem. The graph derived from the plan of the park at Königsberg has a Hamiltonian path, although, as we have seen, it has no Eulerian path. On the other hand, removing two lines results in a graph that has an Eulerian path but not a Hamiltonian one. Many other graphs have neither kind of path.

Hamilton's problem can be solved by exhaustive search; indeed, the procedure is not substantially different from that employed in listing all the possible paths that might have the Eulerian property. For Hamilton's problem, however, no efficient algorithm comparable to Euler's method has been found. The problem has been pondered by many of the best mathematicians of the past century, but the most efficient methods available today are fundamentally no better than exhaustive tabulation. On the other hand, all attempts to prove that there is no better method have also failed, and it must be considered a possibility that an efficient algorithm will be discovered tomorrow.

Problems that are known to have polynomial-time solutions, such as Euler's problem, are said to be members of the class *P* (for polynomial). Hamilton's problem is a member of another class,

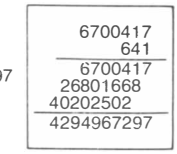








4,294,967,297

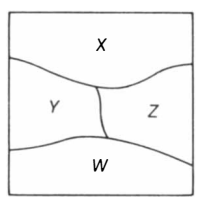


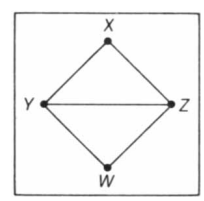

**PROBLEMS IN THE CLASS *NP* ask a yes-or-no question that often can be answered only through a time-consuming, inefficient procedure, but when the answer is known to be yes, that fact can be demonstrated by a short proof. For the present there is no efficient way to determine whether a graph has a Hamiltonian path, but if it does have one, a brief "certificate" can be issued to prove it. The certificate simply lists the lines of the path in the order they are traversed. Another problem in the class *NP* asks whether a number is composite, that is, whether it can be written as the product of two other numbers. No efficient way of answering this question is known; indeed, in one case it took almost 100 years to show that a number is composite. In 1640 Pierre de Fermat proposed that 4,294,967,297, which is equal to $2^{32} + 1$, is a prime number, and he was not proved wrong until Euler discovered the factors of the number in 1732. Once a number is known to be composite, however, that fact can be demonstrated by exhibiting a multiplication that gives the number as an answer.**

designated by the letters *NP*, signifying "nondeterministic polynomial." The class *NP* encompasses all the problems in *P*, or in other words *P* is a subset of *NP*. In addition *NP* includes problems whose status is less certain. They are all solvable problems in principle; they have algorithms, but for now only exponential-time algorithms are known. They may also have polynomial-time algorithms (in which case *NP* and *P* are identical), or they may prove to be permanently intractable, with only inefficient solutions.

The problems considered here, and all problems classified in this way, can be described as an infinite set of similar questions each of which can be answered yes or no. For problems that are formally unsolvable, such as the problem of predicting whether a Turing machine will halt, these questions simply cannot be answered by any algorithmic procedure. For problems of the class *P* the questions can invariably be answered, whether the answer turns out to be yes or no, by an efficient procedure. In order for a problem to qualify for the class *NP* there need not be an efficient means of answering the yes-or-no questions. What is required is that whenever the answer is yes there be a short and convincing argument proving it.

Hamilton's problem, for example, meets this condition. It is not possible to tell by any efficient means known today whether a graph has a Hamiltonian path, but if it does, then the path itself can be exhibited. Hence for every Hamiltonian graph it is possible to issue a "certificate" that proves its membership in this special class of graphs. Such a




∧ AND

∨ OR

¬ NOT

| | |
|---|---|
| $[(X \text{ is red}) \vee (X \text{ is blue}) \vee (X \text{ is green})] \wedge [(Y \text{ is red}) \vee (Y \text{ is blue}) \vee (Y \text{ is green})] \wedge [(W \text{ is red}) \vee (W \text{ is blue}) \vee (W \text{ is green})] \wedge [(Z \text{ is red}) \vee (Z \text{ is blue}) \vee (Z \text{ is green})] \wedge$ | Every country is colored some color. |
| $[(\neg X \text{ is red}) \vee (\neg X \text{ is blue})] \wedge [(\neg X \text{ is red}) \vee (\neg X \text{ is green})] \wedge [(\neg X \text{ is blue}) \vee (\neg X \text{ is green})] \wedge$<br>$[(\neg Y \text{ is red}) \vee (\neg Y \text{ is blue})] \wedge [(\neg Y \text{ is red}) \vee (\neg Y \text{ is green})] \wedge [(\neg Y \text{ is blue}) \vee (\neg Y \text{ is green})] \wedge$<br>$[(\neg W \text{ is red}) \vee (\neg W \text{ is blue})] \wedge [(\neg W \text{ is red}) \vee (\neg W \text{ is green})] \wedge [(\neg W \text{ is blue}) \vee (\neg W \text{ is green})] \wedge$<br>$[(\neg Z \text{ is red}) \vee (\neg Z \text{ is blue})] \wedge [(\neg Z \text{ is red}) \vee (\neg Z \text{ is green})] \wedge [(\neg Z \text{ is blue}) \vee (\neg Z \text{ is green})] \wedge$ | No country is colored two colors. |
| $[(\neg X \text{ is red}) \vee (\neg Y \text{ is red})] \wedge [(\neg X \text{ is blue}) \vee (\neg Y \text{ is blue})] \wedge [(\neg X \text{ is green}) \vee (\neg Y \text{ is green})] \wedge$ | *X* and *Y* are not the same color. |
| $[(\neg X \text{ is red}) \vee (\neg Z \text{ is red})] \wedge [(\neg X \text{ is blue}) \vee (\neg Z \text{ is blue})] \wedge [(\neg X \text{ is green}) \vee (\neg Z \text{ is green})] \wedge$ | *X* and *Z* are not the same color. |
| $[(\neg Y \text{ is red}) \vee (\neg Z \text{ is red})] \wedge [(\neg Y \text{ is blue}) \vee (\neg Z \text{ is blue})] \wedge [(\neg Y \text{ is green}) \vee (\neg Z \text{ is green})] \wedge$ | *Y* and *Z* are not the same color. |
| $[(\neg Y \text{ is red}) \vee (\neg W \text{ is red})] \wedge [(\neg Y \text{ is blue}) \vee (\neg W \text{ is blue})] \wedge [(\neg Y \text{ is green}) \vee (\neg W \text{ is green})] \wedge$ | *Y* and *W* are not the same color. |
| $[(\neg W \text{ is red}) \vee (\neg Z \text{ is red})] \wedge [(\neg W \text{ is blue}) \vee (\neg Z \text{ is blue})] \wedge [(\neg W \text{ is green}) \vee (\neg Z \text{ is green})]$ | *W* and *Z* are not the same color. |

**PROPOSITIONAL CALCULUS serves as a universal language for problems in the class *NP*. The problem considered here is that of applying three colors to a map or its equivalent graph. A sentence in the propositional calculus is made up of statements, such as "*X* is red," joined by the logical connectives "and," "or" and "not." If two statements are joined by a logical "and," the sentence is true only if both statements are true; a logical "or" requires that at least one of the statements be true, and "not" signifies that a statement is false. The sentence representing the map-coloring problem has three parts: the first two establish that every country has exactly one color and the third part lists the countries that cannot have the same color. The map-coloring problem is thereby reduced to the problem of determining whether this sentence can be satisfied, that is, whether it is possible to assume some statements to be true and others false in such a way that there are no contradictions. Since all problems in *NP* could be expressed as sentences in the propositional calculus, an efficient general method for solving the satisfiability problem could be applied to all those problems. No such method is known. The sentence given here can be satisfied by assuming, for example, that only the statements "*X* is red," "*Y* is blue," "*Z* is green" and "*W* is red" are true.**

certificate would name the lines in the graph in the order the path traverses them. Finding the path might require weeks of tabulation, but once it has been found it can easily be exhibited. Another problem that belongs to the class *NP* is the question of whether a whole number is composite, that is, whether it can be written as the product of two other numbers. Again there is no known efficient procedure for answering the question, but if the number is indeed composite, there is a succinct proof of that fact, namely a correctly worked-out multiplication with the number on the bottom line.

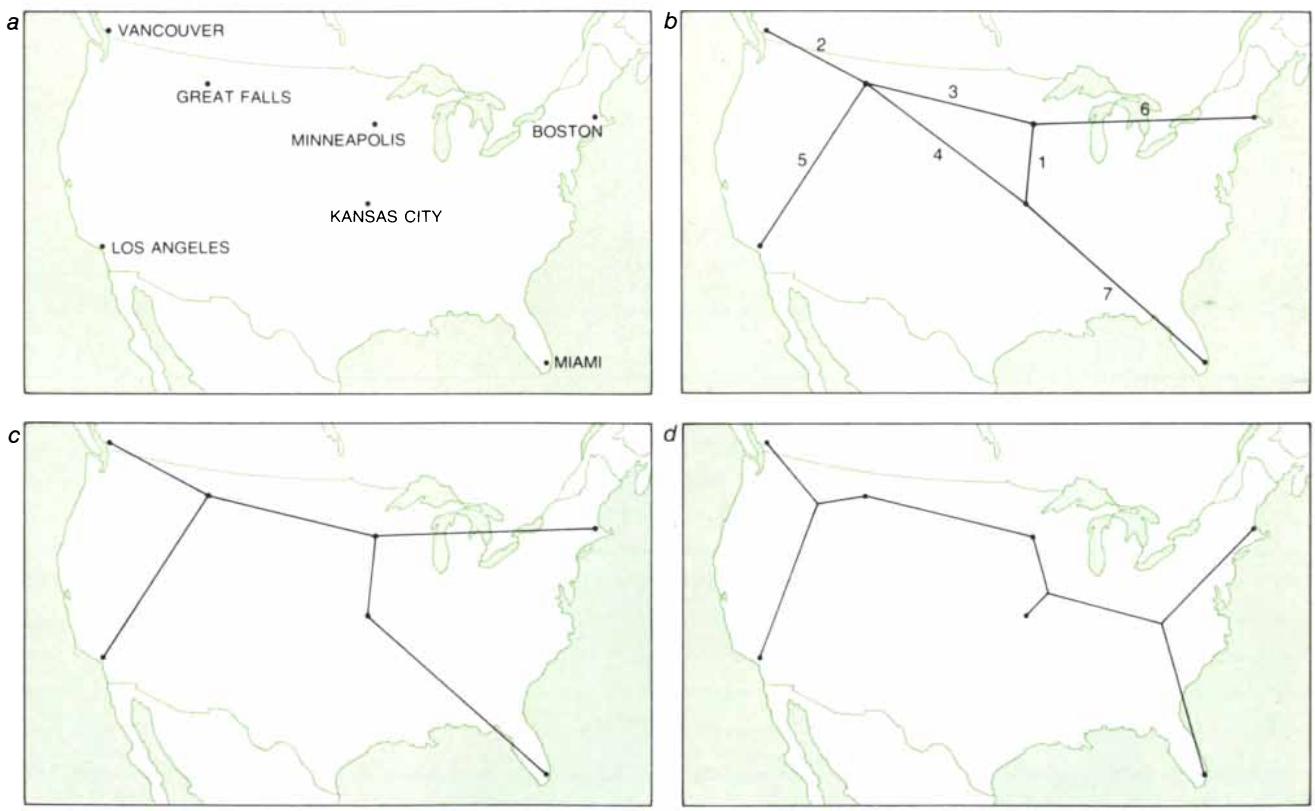
Care must be taken when asking the yes-or-no question of a problem in the class *NP*, since the complementary no-or-yes problem might not be in the same class. For example, the complement of Hamilton's problem, in which one is asked to show that a graph does not have a path passing once through each point, may well not be in the class *NP*. For now the only way to demonstrate the absence of such a path is to list all possible paths, and such a proof is too lengthy to qualify as a certificate of membership in *NP*. On the other hand, the complement of the composite-number problem, which asks if a number is prime, turns out to be in the class *NP*. The reason, which is far from obvious, is that relatively short proofs demonstrating that a number has no factors other than 1 and itself were discovered in 1975 by Vaughan Pratt of the Massachusetts Institute of Technology. Still, it is not known whether the composite-number problem and its complement are in the class *P*.

It is easy to show that every problem in the class *P* is also in the class *NP*. If a problem is in *P*, then by definition there is an efficient algorithm for it. To produce a short and convincing proof that the answer to some instance of the problem is yes, all we need to do is follow the algorithm; a record of its operation constitutes the required certificate.

Another way of defining *NP* is as the class of yes-or-no problems that can be solved by guessing certificates. If one is given an instance of a problem in *NP* for which the answer happens to be yes, then with luck one may discover the required certificate fairly quickly by making a sequence of guesses; if the answer is no, guessing cannot possibly yield an answer any faster than an exhaustive search could. For example, in solving Hamilton's problem one might find a correct path (if there is one) on the first try by tracing small portions of the path and guessing at each stage how to proceed. Such a procedure, it should be emphasized, is not an algorithm. It could be made into an algorithm only by crossing off each trial path as it is tested and checking all possible paths, but that is equivalent to the method of exhaustive search.

A mathematical procedure defined in terms of lucky guesses may seem bizarre, but it is a quite legitimate approach to defining the problems in the class *NP*. In principle the procedure could even be mechanized by building a device called a nondeterministic Turing machine. This device can do all that an ordinary Turing machine can do; in addition, at some points in its operation it may have more than one choice of what to do next. Such a machine would be considered to answer yes to a question if there were some sequence of choices that could lead it to a yes conclusion. *NP*, the class of nondeterministic polynomial-time problems, consists of precisely those problems whose yes instances can be identified by machines making comparatively short guessing computations.

The inclusion of guessing in the definition of these problems suggests strongly




**SPANNING-TREE PROBLEM calls for the shortest network of lines connecting a set of points, or equivalently the shortest railroad system connecting a set of cities (*a*). If the lines are allowed to meet only at cities, the problem can be solved by an efficient procedure called the greedy algorithm. First the closest pair of cities are joined, then the next-closest and so on (*b*). Lines joining cities that are already connected indirectly (such as line *4* here) are omitted. The result is the optimum spanning tree (*c*). An even shorter network is possible if lines are allowed to meet at isolated junction points. At each such point three lines should meet at angles of 120 degrees, but there is no efficient algorithm for determining where junction points should be introduced. The optimum network (*d*) was found by exhaustive search.**

to many mathematicians that *P* and *NP* are not the same set and hence that efficient algorithms can never be found for the intractable problems in the class *NP*. If every problem in *NP* were actually in *P*, then all the guesswork and luck could be replaced by some systematic procedure without great sacrifice in time. It is hard to believe the ability to guess and be lucky could win so little.

The class *NP* includes a variety of commonly encountered problems that seem to defy efficient solution. We have already mentioned Hamilton's problem and the problem of composite numbers. Another example is known as the matching problem. It can be considered in terms of the task faced by the colleges every September, when a new class of freshmen must be assigned to shared dormitory rooms.

For the sake of simplicity let us assume that all the information gathered about the students' smoking habits, bedtime hours, taste in music and so forth results in a single yes-or-no decision as to the compatibility of each possible pair of students. The entire class can then be represented as a graph in which the points correspond to students and a line is drawn connecting every two students who can be placed in the same room. If each room holds just two students, the assignment can be made efficiently by a clever polynomial-time algorithm discovered by Jack Edmonds of the University of Waterloo. If each room is to be shared by three students, however, there is no known efficient algorithm. The problem is in the class *NP*, since all yes instances have succinct certificates: an acceptable room assignment, once it is discovered, can easily be exhibited. Of course, a solution could be found by exhaustive search, albeit inefficiently. With luck a suitable assignment, if there is one, can be guessed quickly.

Map coloring is a problem in the class *NP* that concerns mathematicians more than it does cartographers. The question is whether the countries on a given map can be colored with a given number of colors so that no two countries that share a border have the same color. It is easy to find out if a map can be colored with two colors: it can be if there are no places on the map where an odd number of countries meet at one point. It is even easier to tell if a map can be colored with four colors; indeed, there is no need even to look at the map, since Kenneth Appel and Wolfgang Haken of the University of Illinois proved in 1975 that four colors suffice for any map. Surprisingly, however, no efficient algorithm is known for determining whether three colors are enough for a given map. The problem is in the class *NP*, since a correctly colored map can serve to certify a yes answer.

Map coloring can be regarded as a

special case of another problem called graph coloring. Any map can be converted into a graph by reducing each country to a point and drawing a line between two points if the corresponding countries share a border. Coloring the graph is then equivalent to coloring the map, subject to the rule that two points connected by a line cannot have the same color. Graph coloring, however, is a more general problem, with applications outside graph theory. For example, a graph can represent the scheduling of work in a factory. Each point of the graph stands for some job to be done, and two points are connected by a line if the jobs cannot be done concurrently, perhaps because they require the same piece of machinery. A coloring of the graph with three colors would then supply a schedule dividing the work of the factory into three shifts. Like map coloring, the graph-coloring problem is in the class *NP*.

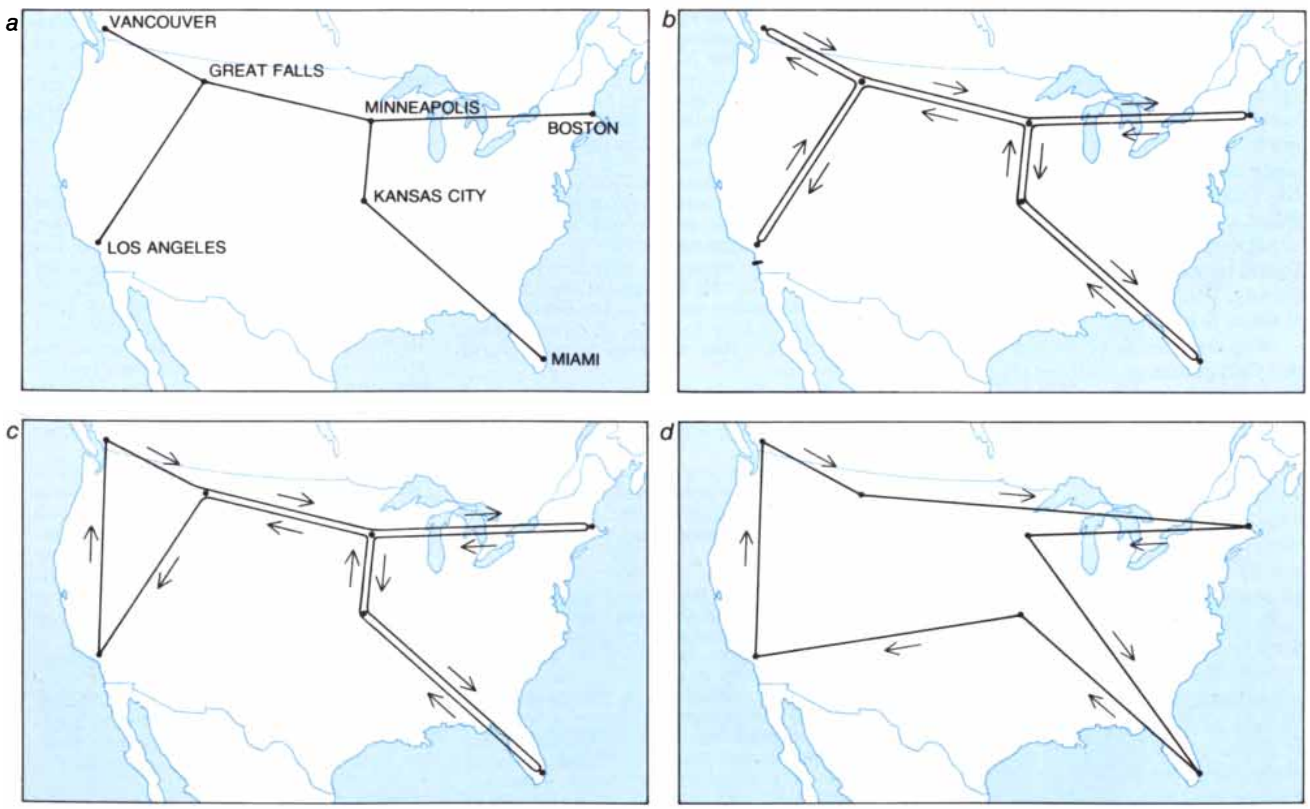
It often happens that if one problem can be solved efficiently, so can many others. For example, if an efficient algorithm could be found for the problem of graph coloring, it could be applied with only minor modifications to the problems of map coloring and factory scheduling. Map coloring and factory scheduling are therefore said to be efficiently reducible to graph coloring. In the past several years it has become apparent that some of the problems in the class *NP* have a remarkable property: all the problems in *NP* are efficiently reducible to them. These elite problems within the class *NP* are called *NP*-complete. If any one of them has an efficient algorithm, then every problem in *NP* can be solved efficiently.

The first proof that a problem is *NP*-complete was presented in 1971 by Stephen A. Cook of the University of Toronto. His reasoning follows a path essentially parallel to the path of Turing's earlier work on mathematical machines and their relation to problems of formal logic. Cook stated his proof in terms of the propositional calculus, the formal language in which separate logical statements, which individually may be either true or false, are joined together by the lexical elements "and," "or" and "not." In general a sentence in the propositional calculus can be shown to be either true or false depending on which of its component statements are assumed to be true or false. Certain sentences, however, cannot be true under any interpretation because they are self-contradictory. Sentences that cannot be made true are said to be unsatisfiable.

Cook employed the propositional calculus to describe the operation of the nondeterministic Turing machines, the mechanized guessing devices essential to the definition of the class *NP*. He showed that the calculations of any such machine can be described succinctly by sentences of the propositional calculus. When the machine is given a yes instance of a problem in *NP*, its operation is described by a satisfiable sentence, whereas the operation of a machine given a no instance is described by a sentence that cannot be satisfied.

It follows from Cook's proof that if one could efficiently determine whether a sentence in the propositional calculus can be satisfied, one could also determine efficiently in advance whether the problem presented to a nondeterministic Turing machine will be answered yes or no. Since the problems in the class *NP* are by definition all those that can be solved by nondeterministic Turing machines, one would then have an efficient method for solving all those problems. The catch, of course, is that there is no known efficient method of determining whether a sentence in the propositional calculus can be satisfied.

Cook's argument states in essence that the propositional calculus is a universal language for describing problems in the




**SHORTEST-TOUR PROBLEM cannot be solved exactly by any known efficient algorithm. Mathematicians have therefore devised solutions that are good even if they are not optimal. One efficient procedure draws a tour guaranteed to be no more than twice the shortest length. The method begins with the optimum spanning tree (*a*), which can be found efficiently with the greedy algorithm. The spanning tree can be converted into a tour simply by traversing each line once in each direction (*b*). This tour is clearly just twice as long as the spanning tree itself; the spanning tree in turn must be shorter than any tour of the cities, since a tour could be made into a tree (albeit a tree with no branches) by omitting one segment. The tour produced by this method can generally be improved by taking shortcuts (*c*, *d*).**

class *NP*. Every instance of such a problem corresponds to a sentence in that language, and if the sentence is satisfiable, the instance has a yes answer. Many other problems have since been shown to be *NP*-complete because the satisfiability problem can efficiently be reduced to them. Hamilton's problem, the problem of matching groups of three roommates and the problem of coloring graphs with three colors are all *NP*-complete. The first to point out the broad applicability of this theory was Richard M. Karp of the University of California at Berkeley. Similar investigations were independently conducted by the Russian mathematician P. A. Levin. Since *NP*-complete problems capture the difficulty of all other problems in *NP*, it is widely thought today that all *NP*-complete problems are computationally intractable. A proof that a problem is *NP*-complete is usually considered a strong argument for abandoning further efforts to devise an efficient algorithm for its solution.

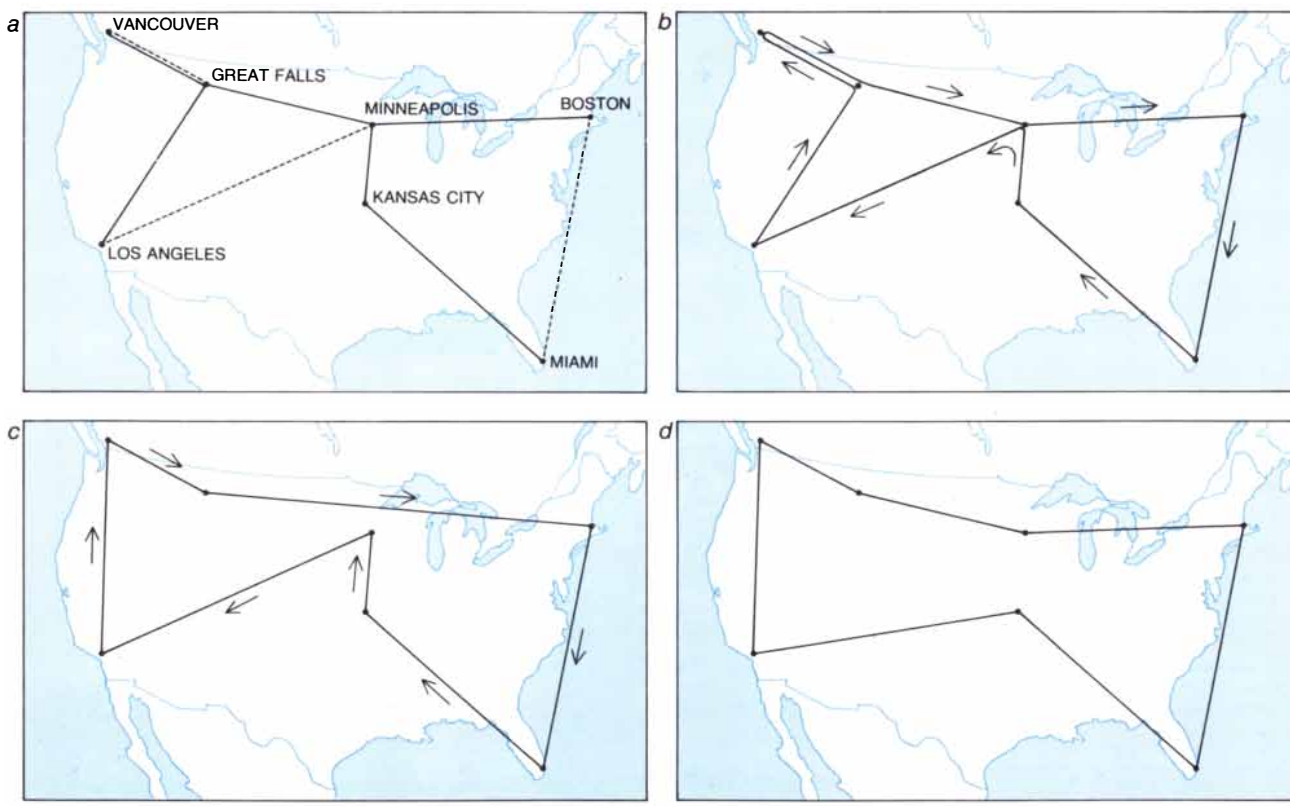
Even the assumption that all *NP*-complete problems are intractable would not settle all questions about the class *NP*. In addition to the mystery of the *NP*-complete problems there is an even more obscure area: problems in *NP* for which no efficient algorithms are known but which have not been proved to be *NP*-complete either. The problem of composite numbers is one of these.

Not all problems that can be solved by a computer are of the yes-or-no type. Another common type is the optimization problem. For example, suppose one is given the positions of some cities on a map and asked to find the shortest possible railroad network connecting them. In one version of this problem one is allowed to lay down a straight section of track between any two cities, but one is not allowed to install isolated junction points; tracks can be joined only at cities. One property of the solution to this problem is immediately apparent: the optimum network can never include a closed circuit, because if it did, the network could be made shorter simply by omitting one link in the circuit. Thus the best network always branches like a tree, and the problem itself is called the spanning-tree problem.

The spanning-tree problem can be solved correctly and quite efficiently by a method called the greedy algorithm, devised by Joseph B. Kruskal of Bell Laboratories. The procedure is simply to connect the closest pair of cities, then the next-closest and so on without adding any superfluous lines (lines joining cities that are already linked indirectly). It is far from obvious that this method always yields the shortest network, but it does, and it has the pleasant property of requiring no foresight and no reconsideration of earlier decisions.

The greedy algorithm can be relied on to find the shortest network between cities under the rules specified, but in general that network will not be the shortest possible one. Further savings can be achieved by establishing junction points between cities. The properties of networks with such junctions were studied by the Swiss mathematician Jakob Steiner. It can be shown that any shortest network must be arranged so that each junction point is made up of three lines that meet at angles of 120 degrees. This rule provides some guidance in evaluating networks, but there are many possible networks with Steiner junction points. No algorithm has been discovered that finds the best network quickly.

The problem of the traveling salesman's tour is closely related. Again one is given a set of cities, but now one is asked to find the shortest round-trip tour. As a first guess the greedy algorithm suggests that perhaps the salesman should always go to the nearest city he has not yet visited, but this procedure



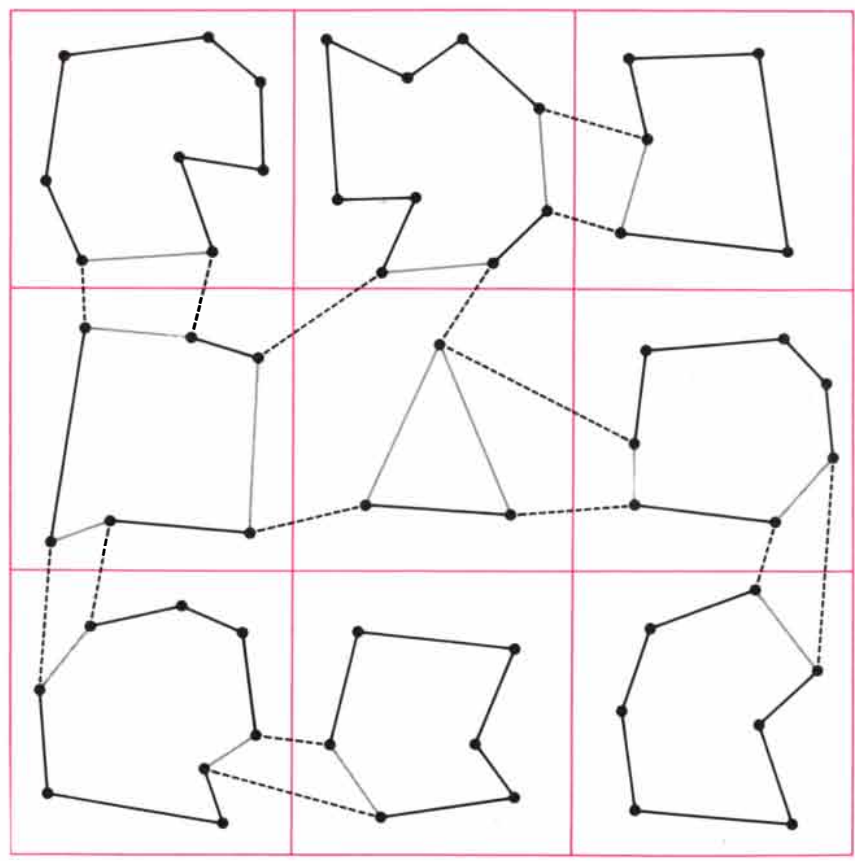

**IMPROVED ALGORITHM for the traveling salesman problem yields a tour that is certain to be no more than 50 percent longer than the optimum. The procedure was devised by Nicos Christofides of the Imperial College of Science and Technology. The first step is again to generate the shortest spanning tree. All the cities that are linked to an odd number of cities are then singled out; in this example all the cities except Kansas City have an odd number of connections. These cities are next linked in pairs by a procedure similar to the one employed in matching pairs of students to yield a tour (*a*) that can be improved by making shortcuts (*b, c*). The result is only a little longer than the optimum tour (*d*), found by exhaustive search. Note that in the shortest tour the line between the two closest cities is omitted.**

does not work. Indeed, the problem is notorious for having resisted all attempts to find an efficient solution.

Optimization problems are not fundamentally different from those that ask yes-or-no questions; in fact, every optimization problem can be rewritten in a yes-or-no form. In the traveling salesman problem, for example, we might be given a length along with a set of cities and asked to state whether a tour can be constructed that does not exceed the specified length. This yes-or-no problem cannot be harder than the associated optimization problem, because if the optimum tour could be found by some efficient method, it would be a trivial task to determine whether it exceeds a given number. Hence if the yes-or-no version is computationally intractable, one cannot hope to solve the optimization problem itself efficiently. For this reason certain optimization problems, such as that of the traveling salesman's tour and that of placing Steiner junction points, are said to be *NP*-complete.

Optimization problems are encountered often in engineering, economics, operations research and other fields. The discovery that at least some of these problems are *NP*-complete is therefore of considerable practical interest. Since the *NP*-complete problems probably have no efficient algorithms, there would seem to be little point in expending further effort in seeking optimum solutions. An alternative that has recently been adopted is to seek approximate solutions that are good even if they are not precisely optimal.

One technique that has been applied to the traveling salesman problem offers a solution that may not be optimum but is guaranteed to be no worse than twice the optimum path. The procedure starts with the shortest spanning tree, which can be generated efficiently by the greedy algorithm. This network can be converted into a tour of the cities by traversing each line twice and returning to the origin. It is known that the optimum spanning tree must be shorter than any possible tour of the cities, since a tour can be converted into a spanning tree (albeit one without any branches) by simply omitting one segment. Thus twice the length of the optimum spanning tree cannot be longer than twice the optimum tour. The meth-



**APPROXIMATE SOLUTION to the traveling salesman problem is not guaranteed to find a tour within any specified range of the optimum tour, but when it is applied to many instances of the problem, it is not far wrong very often. The algorithm, which was devised by Richard M. Karp of the University of California at Berkeley, divides a map into many small regions, each one containing only a few cities. Within each region the optimum tour is found by exhaustive search, a procedure that is practical since the number of points is small. Each of the small tours is then regarded as a single entity to be linked to the others, a task that can be performed efficiently by an algorithm similar to the greedy algorithm for finding an optimum spanning tree.**

od is a polynomial-time algorithm. Recently Nicos Christofides of the Imperial College of Science and Technology in London has found a way to improve the algorithm so that it yields a tour guaranteed to be no more than half again as long as the optimum.

A more profound compromise gives up not only the requirement that a solution be optimal but also the insistence that a less than optimum solution be guaranteed to fall within a specified range. Instead the assurance is given that the solution will not often deviate far from the optimum. An underlying assumption of such techniques is that the maps encountered in practice are not concocted to confound basically plausible techniques; such maps are encountered frequently only when they are constructed by computer scientists to reveal the flaws in methods proposed by their colleagues. Indeed, if the salesman's-tour algorithms discussed above are applied to "natural" maps, they deliver far more than they promise. The resulting tours are not 100 percent or 50 percent longer than the optimum but closer to 5 percent.

A reasonable assumption about the properties of many maps is that cities are randomly placed. A theorem describing the statistical properties of optimum tours through such randomly distributed points was proved in 1958 by Jillian Beardwood, J. H. Halton and John M. Hammersley of the University of Oxford. Relying on that theorem, Karp has shown that a simple method of constructing tours almost always yields near-optimum results when it is applied to maps with many cities.

Karp begins by dividing the map into many small regions. Within each of these regions the cities are sufficiently few to find the optimum tour by exhaustive search, even though that method involves an exponential-time algorithm. The tours of the small areas are then linked by a variant of the greedy algorithm. Perhaps significantly, the method is not very different from the method usually adopted by people solving the problem manually.

Efficient but approximate solutions can be found for many *NP*-complete optimization problems. From the standpoint of mathematics, however, the important question is whether *NP* is identical with *P*. The repeated failure of attempts to find an efficient algorithm for the *NP*-complete problems has created considerable confidence that *NP* and *P* are not the same. There is now suspicion that they are not identical, but the proof of their distinctness may be beyond present mathematical capabilities. The question may join that select group of mathematical enigmas that remain unresolved for decades, and the solution may have to await the development of new methods in mathematics.

**CHARIOTEER OF THE BLUES, whip held high in his right hand and reins gathered in his left hand, is portrayed in mosaic. He is one of four charioteers shown in the starting gates of a circus; the mosaic decorated the threshold of a dining room unearthed recently at a house site in Roman Carthage, where an international rescue-archaeology campaign began under UNESCO auspices in the mid-1970's.**

# Roman Carthage

*An international campaign of rescue archaeology at the ruined city is uncovering significant information about its successive roles as a Roman colony, Vandal prize and outpost of the Byzantine Empire*

by John H. Humphrey and John Griffiths Pedley

In the last quarter of the first millennium B.C. Carthage, the greatest Phoenician colony in Africa, and Rome, then a young expansionist republic in central Italy, fought bitterly for control of the western Mediterranean. Between 264 and 146 B.C. the two powers were at war with each other on three occasions. Although some 50 years of uneasy peace separated the end of the second Punic War from the start of the third, the Roman senate could not forget the rivalry. The senators heard Cato the Elder demand on every possible occasion that Carthage be destroyed.

Cato did not live to see his demand met; he died at age 85 in 149 B.C. just as Rome mounted its final campaign against the Carthaginians. After a grim three-year siege Scipio Aemilianus, the adopted grandson of the victorious Roman commander in the second Punic War, captured Carthage in 146 B.C. His first orders after the city's surviving defenders were dispersed would have won Cato's approval. Scipio's troops leveled every building in the city; the site was then placed under a curse and ceremonially sown with salt.

The seeming finality of Scipio's gesture was short-lived. Less than 25 years later the Romans reversed their policy. Reborn as a Roman colony, Carthage was soon once again a great city, destined to survive for another 800 years. The remarkable history of these Punic and post-Punic centuries is being read in the city's ruins today as the result of an unprecedented international project in rescue archaeology. Launched by the director general of the United Nations Educational, Scientific and Cultural Organization in 1972, the project involves workers not only from the host nation, Tunisia, but also from nine other nations of Europe and America.

That there is more than enough work for all these participants will be made clear by a brief outline of the history of Roman Carthage. The capital of Roman Africa by the end of the first century B.C., the new city of Carthage eventually became the third-richest metropolis in the Roman Empire, rivaling Rome itself in magnificence. The Emperor Augustus renamed the city Colonia Julia Carthago, and with the advent of Christianity it became a center of religious ferment. By A.D. 439, the year Carthage surrendered to the advancing Vandals, the city had at least 22 great churches, its own bishop, more than 500 other clerics and a population estimated at 250,000 or more.

Imperial troops eventually expelled the Vandals, but they were troops of the Eastern Roman Empire; the liberated city was renamed Colonia Justiniana Carthago in honor of the ruler at Byzantium, Justinian the Great. Carthage remained a Byzantine city from A.D. 533 to 697. It might have stood to this day if its people had not risen in 697 against a small Arab garrison left there by the Moslem governor of Egypt, Hassan ibn en Noman. As the French classical scholar E. C. F. Babelon relates the story, Hassan, whose forces were busy adding Algeria and Morocco to the expanding realm of Islam, was infuriated by what he considered a treacherous attack. He repeated the Roman razing of nearly 1,000 years earlier in A.D. 698, and the city of Carthage was once again left in ruins.

Today, some 1,300 years later still, the twice-obliterated city has become a modern building site, its empty acres disappearing under the suburban sprawl of adjacent Tunis. In 1881 Tunis was occupied by the French, and in the years after World War II what had been a trickle of urban development became a flood. By the early 1970's deserted Carthage was the most desirable real estate in all Tunisia. The new palace of the president had been built over the northern ruins of Carthage, and large villas appeared on the heights overlooking the coast where the ancient Byrsa, or citadel, of the Punic city had stood. The adjacent heights, known in Roman times as the Hill of Juno and where the ruins of a Roman theater are still visible, were also invaded. Still other villas were built along the coast northward from Tunis up to and beyond the former Roman city wall, creating the present-day suburbs of Salammbo, Dermech and Sayda.

When the program of rescue archaeology was begun in the mid-1970's, the few undeveloped sections of the ruined city were apportioned among the various participating archaeologists. The Tunisian National Institute of Archaeology and Art, under the direction of Azedine Beschaouch, and Abdelinajid Eunatti, the conservator for the site of Carthage, have continued earlier work at Roman villas near the Roman theater. The principal undertaking of the host country, however, has been a massive cataloguing program, covering not only current discoveries but also the scattered findings of the past 150 years. A British group, directed by Henry Hurst, was assigned the island in the middle of one of the city's two artificial harbors: a circular harbor traditionally considered to have been military. The British were also given a second site on the north side of that harbor and a third area a short distance inland on the line of the Roman city wall.

From the circular harbor northward to city limits only one small unoccupied area remained on the shore of the Gulf of Tunis (except for the ruins of the Antonine baths, a part of Roman Carthage that was turned into an archaeological park some years ago); it was assigned to German excavators under the leadership of F. Rakob. Several other groups concentrated on parts of the city lying farther inland: the French, under S. Lancel, chose to continue earlier work by a succession of French investigators at the Punic citadel; one Canadian group, under P. Sénay, undertook to extend earlier French work at a circular monument on the Hill of Juno and another, under C. Wells and E. Wightman, traced the city wall nearby. Still farther inland an Italian group under A. Caran-

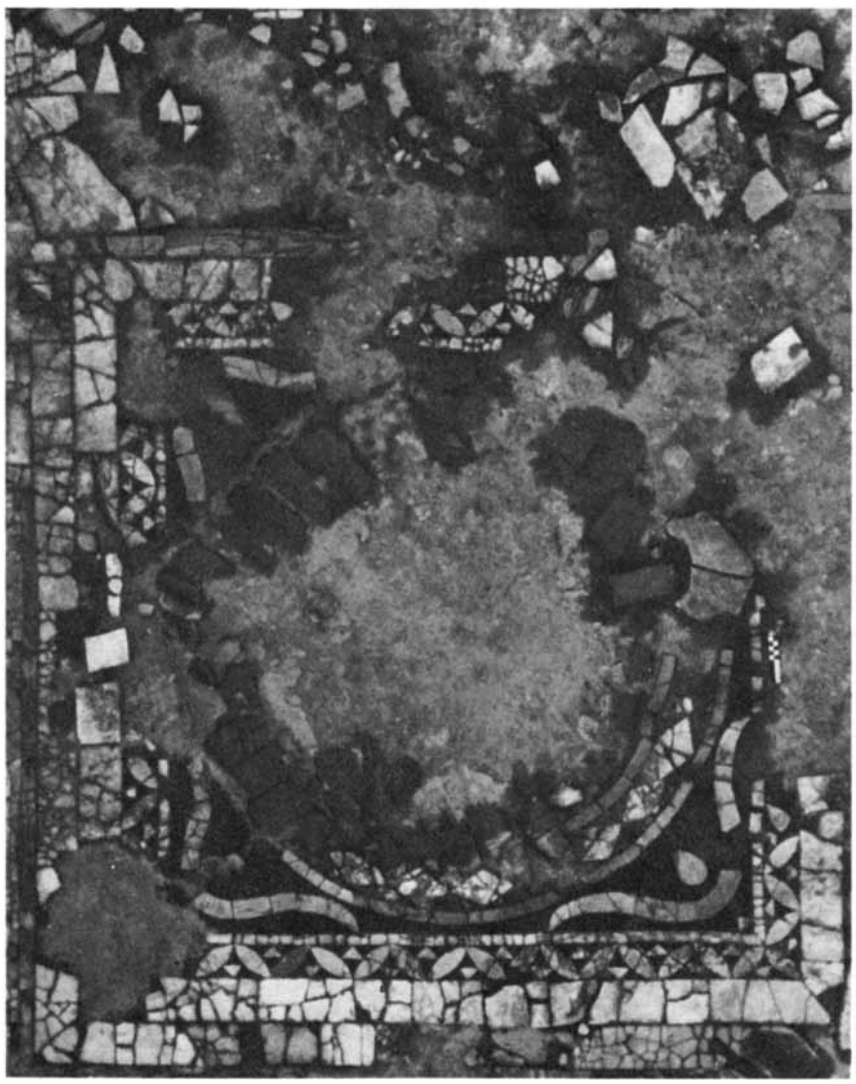
dini set out to survey the street grid and city wall at the northwest corner of the city; archaeologists from the Institute of the History of Material Culture in Warsaw surveyed the southwest sector, particularly in the vicinity of the Roman circus, measuring variations in the electrical resistivity of the soil. The Danish group, under S. Dietz, chose a coastal site north of the Antonine baths, outside the city limits, where earlier Danish excavations had taken place. A Bulgarian group worked on another site outside the city limits: Damous el Karita, a famous early Christian church.

Two American groups were engaged in the rescue project. One group, a joint enterprise of Harvard University and the University of Chicago, was assigned an area on the western edge of the second harbor at Carthage, a rectangular harbor traditionally considered to have been mercantile, located just south of the circular one. The area extends westward into the adjacent Punic tophet, or burial ground, and the sanctuary of the Carthaginian goddess Tanit. Our own group, from the University of Michigan, was assigned a rectangular field 100 by 400 meters. Modern urban development had been forestalled in this area parallel to the rail line from Tunis to La Marsa, which lies west of the circular harbor. We were in a very real sense continuing a Michigan tradition: the first American archaeologist to work in Carthage, in 1925, was a classics scholar from the university, Francis W. Kelsey; his name was given to the university museum we represent as field director and principal investigator respectively. The work of both American groups is sponsored by the American Schools of Oriental Research and is funded largely through the Foreign Currency Program administered by the Smithsonian Institution.

The father, or perhaps more appropriately the grandfather, of modern Carthaginian studies was a 19th-century scholar, C. T. Falbe, who served as the Danish consul at Tunis and published the first detailed map of the ruined city in 1833. A number of other antiquarians, including the Anglican chaplain at Tunis, Nathan Davis, worked at the site in the years that followed, and soon after the French occupation French scholars drew up a comprehensive but wildly overambitious plan for excavating all the city's most important features.

The efforts of these and later investigators, however, served mainly to fill the museums of Europe with works of art such as mosaics and statuary whose exact location and archaeological context were generally undocumented. Indeed, until the current campaign began Carthage's checkered history of excavation and documentation gave it the reputation of being perhaps the hardest Classical Mediterranean site to study. Even the most resolute student of antiquity, attempting to interpret the work that had been done there, was frustrated by an absence of stratigraphic records, of pottery analyses, of numismatic studies and even of such a basic tool as an accurate large-scale map of the city.

That is now changing. Each of the groups working at Carthage has its own particular interests and specialties that for the first time are contributing to an overall view of the city's long history. For example, our group decided to concentrate on the later history of the city: from about A.D. 400, some 40 years before the Vandal takeover, to A.D. 698 and later, a period that includes the subsequent Byzantine revival of the city as Colonia Justiniana and its eventual seizure by the Moslems.



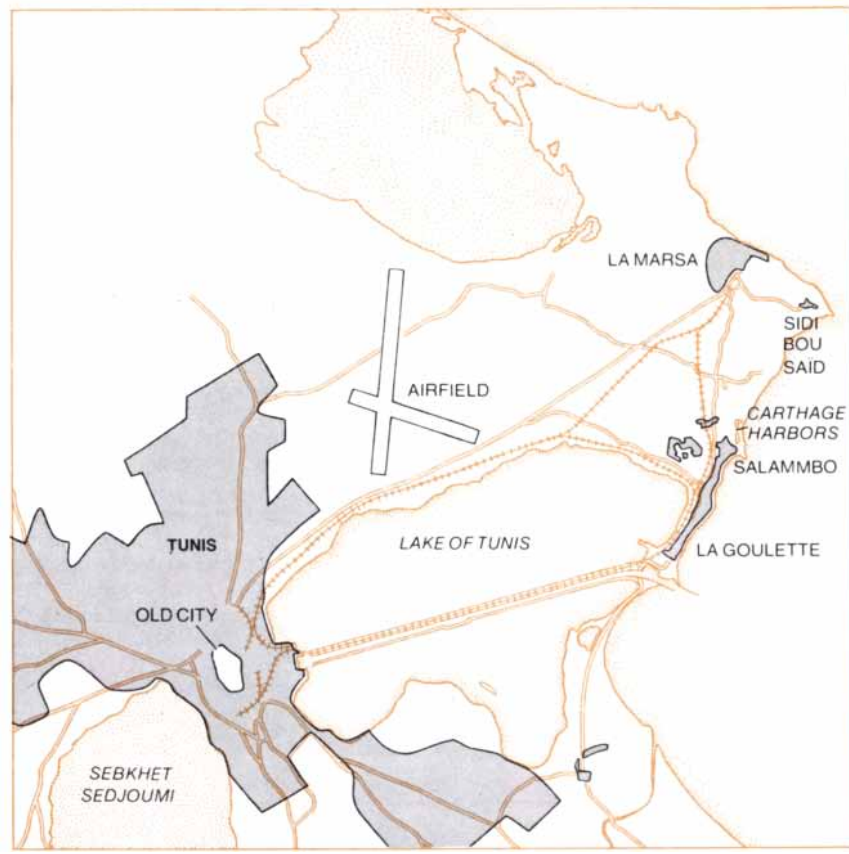
**MARBLE PAVEMENT covered the floor of the dining room after extensive rebuilding of the house that followed the Byzantine defeat of the Vandals in A.D. 533. The pavement, of a kind known as *opus sectile,* was repaired twice by the occupants of the house over the next century.**

The long field that was assigned to us was ideal for the purpose. A preliminary survey of the site by the Tunisian National Institute of Archaeology and Art in 1971 showed that the field included the ruins of a late Christian church and baptistry and those of a number of wealthy private houses of the late Roman (that is, Colonia Julia) period and of the subsequent Vandal and Byzantine periods. In a broadening of the traditional focus of Classical archaeology, long known for its emphasis on such artifacts of history as inscriptions, coins and the products of artists and artisans, we number among our group specialists interested in evidence relating to the economy of the city and the people's way of life. For example, the Carthaginians' diet is reflected in the remains of plants and animals, and their exploitation of the physical environment is indicated by their choices of building stone and of clay for pottery.

A chronological framework is a prerequisite for the study of any Classical site: it enables the investigators to relate

the various excavated levels and the objects found in each level to known periods of history. The greater the chronological precision is, the more detail that can be applied in the interpretation of different kinds of data. At Carthage, as at most Classical sites, the two kinds of artifacts that provide the basis for a firm chronological framework are coins and pottery. As an example, because of the relatively narrow intervals of time that are involved, coins usually provide a means of dating strata that is much more precise than carbon-14 determinations can be. Coins, however, are not necessarily found in abundance, whereas broken pottery is likely to be a major component of the rubbish at any Classical site.

In our oblong plot this once again proved to be the case. The coins we excavated in 1975 and 1976 were quickly identified for us by a University of Michigan colleague, T. V. Buttrey, and provided us with an overall chronological framework. It was, however, the potsherds found in stratigraphic association with the coins and the additional thousands of sherds from various intermediate strata at the site that enabled us to construct a series of quite precise chronological subdivisions based on pottery type and sequence.




**RUINS OF CARTHAGE lie to the north of modern Tunis, the capital city of Tunisia. In recent decades the northward spread of the suburbs of Tunis has almost engulfed the old city.**

The work was not without its difficulties. Until the rescue operation began at Carthage not even one publication had appeared describing the relative stratigraphy of the pottery unearthed in any part of the city. For example, what is known as African red-slip ware, a fine pottery produced in and around Carthage and elsewhere in northern Tunisia, was the main tableware in use not only in Carthage but also throughout the Roman Empire for half a millennium beginning in about A.D. 100. The history of this famous ware was painstakingly reconstructed by John W. Hayes of the Royal Ontario Museum in his book *Late Roman Pottery*. His definitive work, however, was based exclusively on the results of excavations elsewhere than in Tunisia, and the pottery classification depended largely on stylistic rather than chronological considerations. As for coins, until the rescue work began no catalogue of the coins excavated from the soil of Carthage had ever been published.

It is fortunate that Hayes and his colleague J. A. Riley of the University of Manchester have had charge of pottery analysis at our site. Because the fine red-slip tableware is easily recognized it has provided a general framework to which the other fine and coarse wares have been related. The two pottery specialists might readily have found themselves swamped with potsherds if they had not followed a program of selective preservation and disposal. The earth from each locus (that is, a distinguishable unit of soil) at the site was sifted after removal, and all the potsherds were segregated in buckets for washing (and for treatment with acid if they were heavily encrusted with earth). The sherds from each bucket were then sorted according to preestablished categories. Some were classified in terms of function. For example, pieces of lamps, jugs, basins, bowls, amphoras and cooking ware were separated. Other categories were based on the style of pottery, for example buff-finish wares, painted wares and red-slip wares.

The next step was to divide the sherds from each category into four groups: rim fragments, base fragments, fragments of the handle and fragments of the body of the vessel. The first quantitative measurements were made at this stage; the fragments in each of the four groups were counted and weighed and the data were recorded. If in the processing up to this point there was no evidence that the site had been contaminated, Hayes and Riley thought it was safe to combine all the buckets of pottery that had come from any single locus.

The next decision to be made was whether or not the locus was important. If the locus formed part of a "sealed" stratum (that is, an undisturbed level completely buried under an overlying stratum), or if the locus seemed important for some other reason, such as the presence of an unusual ware, Hayes and Riley kept and marked all the pottery recovered from it and assigned catalogue numbers to selected pieces, which were also measured and either photographed or drawn. If the locus was neither sealed nor otherwise important, they would discard most of the body sherds.

Because the body sherds usually far outnumbered rims, bases and handles, the actual quantity of pottery that was finally stored for later study was only a fraction of the quantity excavated. For example, on the basis of associated coins we were able to distinguish 81 varieties of pottery in strata belonging to the third quarter of the sixth century, an interval early in the Byzantine period at Carthage. Of a total of 7,343 sherds excavated only 881 were fragments of rims, bases and handles, and so more than 6,000 sherds could be discarded. Incidentally, the latest coin associated with this group of sherds was minted sometime between A.D. 548 and 565.

In 1976 Hayes and Riley published data relating to 17 separate groups or deposits of pottery from Carthage; the pottery ranged in date from early in the first century to the second half of the sixth. Last year they went on to publish

several additional pottery groups, most of them dating to the seventh century.

Among their findings is a decline in importation of fine wares from Italy early in the history of Colonia Julia. During the first century imported Arretine fine ware made up 27 percent of such wares at our site and imported amphoras made up another 2.5 percent. The proportion of imported wares fell to a fraction of that percentage early in the second century.

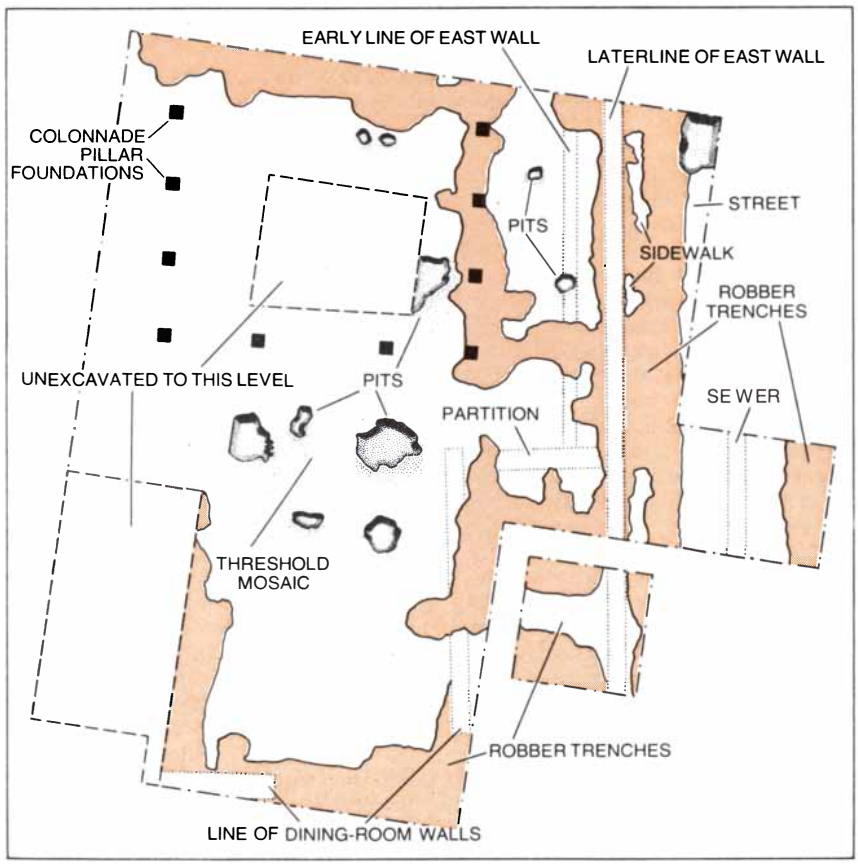
The sifting process yielded much more than potsherds alone. It enabled us to recover a far larger number of coins than might have come to light through more traditional methods of excavation. For example, we recovered more than 600 coins in our 1975 season and more than 1,200 in 1976. Unfortunately their quality did not match their quantity. The coins from the later centuries at Carthage, particularly those from the Vandal period, are very small and poorly minted and also are severely corroded by burial in the acid soil. Still another advantage of sifting, as will be seen below, was the recovery of an abundance of animal bones and plant materials.

Once a secure chronological framework is established it becomes possible to interpret the successive phases of occupation at a site as they relate to the surviving annals of Classical history. For example, we are now able to present at least in broad outline the history of the area we have excavated so far. Our work began in 1975 with the clearing of an area some 15 by 25 meters in the northwest part of our 100-by-400-meter field. The excavation exposed much of the floor plan of a Roman house fronting on one of the regular north-south streets of Colonia Julia. The Roman rebuilders of Carthage had constructed a terrace in this part of the city by piling a massive layer of fill over the underlying rubble of Punic Carthage. We have yet to excavate the Punic levels. Here, as evidently elsewhere in the city, the building of terraces over the Punic ruins took place before the laying out of the main street grid and its lesser subdivisions: the *insulae,* or city blocks. The main street our house fronts on is designated Kardo IX East on the street plan of Roman Carthage.

The date of the terrace construction falls early in the first century A.D. Thus it appears that this area, west of the circular harbor, was not among the first parts of the city to be rebuilt when Julius Caesar and his successor Augustus established Colonia Julia. In any event the city planners evidently subdivided the new city blocks into 16 equal lots, each lot about 17.5 meters square. The earliest of the Roman houses on our site, built no later than the middle of the first century, occupied one such lot or at most two.

We have not excavated all the earlier Roman levels at the house site. What we have exposed proves that a large cistern was built under the courtyard of the house early in the second century. In about the middle of that century pits were evidently dug at the edge of the street immediately outside the house; they were soon refilled and covered by a plastered sidewalk. If this outside activity marks some break in the continuity of occupation, it may have been associated with a great fire that swept Carthage in the reign of the emperor Antoninus Pius (A.D. 138–61) and destroyed much of this section of the city.

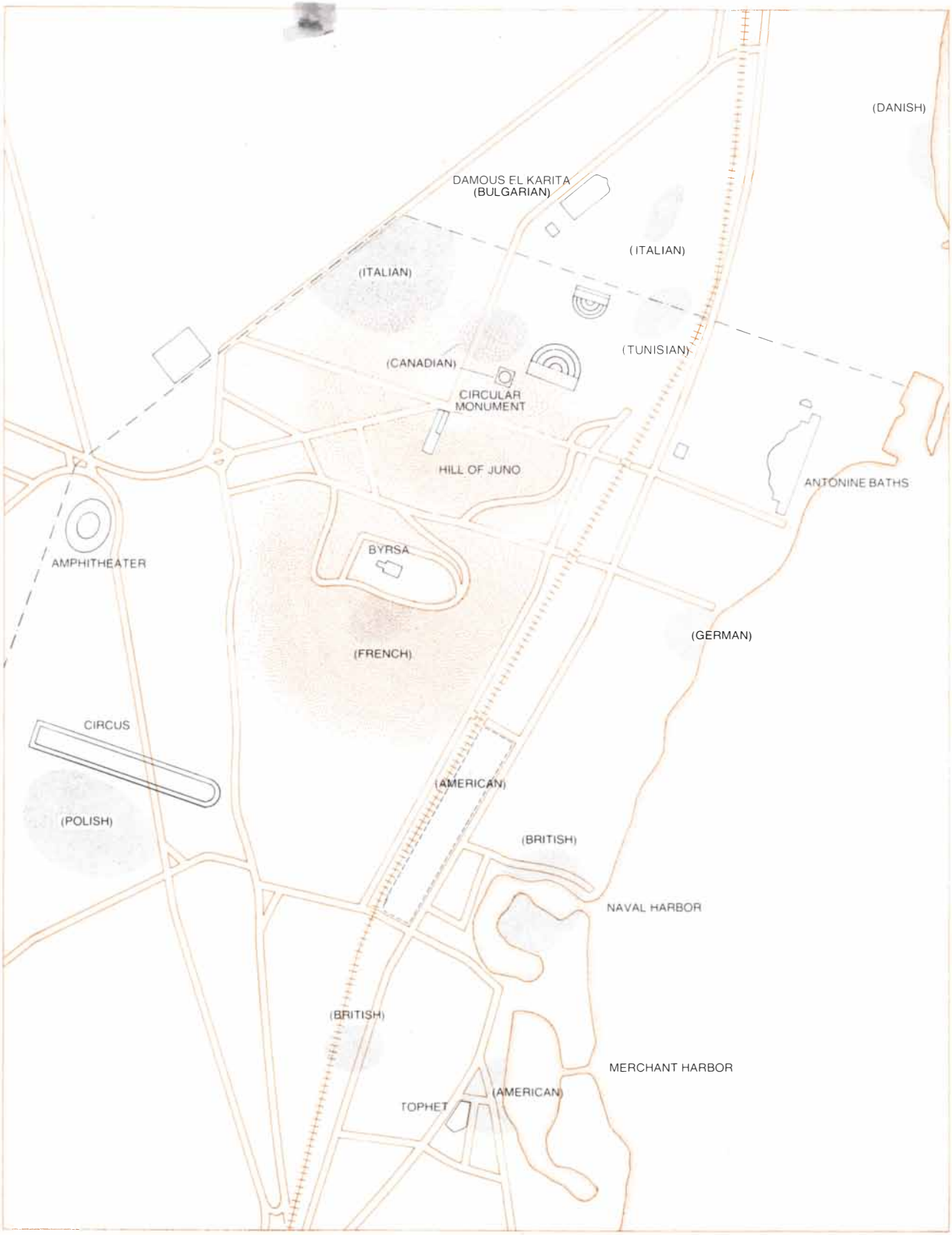
Our work has concentrated on the house remains that are attributable to the late Roman, Vandal and Byzantine periods at Carthage. With regard to the late Roman period it is apparent that the house was completely reconstructed just before A.D. 400. It may also have been enlarged at that time, although it still appears to have occupied only half of the width of the block. The main room of the house, roughly seven by nine meters, was a dining room; the dining-room wall on the east, or street, side of the house is five meters west of the outside wall. To the north the dining room opened onto a colonnaded courtyard, and the room itself was paved with a figured mosaic. The occupants of the house after the Vandal interlude at Carthage replaced the original Roman mosaic, but some of the mosaic's border, with an acanthus-scroll design, has survived.




**ROMAN RESIDENCE, later rebuilt according to the same ground plan, is being excavated by the University of Michigan group at Carthage. Located on one of the many terraces built by the Romans to cover the ruins of Punic Carthage, the first house was built on this site no later than the middle of the first century A.D. Just before A.D. 400 it was reconstructed as is shown here; the mosaic representing the Greek charioteers was at the threshold leading from dining room to courtyard. The extensive trenches (*color*) were dug by later masonry robbers.**

It happens that in the centuries after 100 B.C. Roman Carthage was the home of one of the most important schools of mosaic construction in the empire. The Tunisian reconnaissance of our field uncovered one example that was quite well preserved, and in 1975 we were able to show that the mosaic had stood at the threshold between the dining room and the courtyard of our house. The mosaic depicts four Greek charioteers standing in their chariots at the starting gates of a circus [*see illustration on page 110*]. A small pool occupied the southeast corner of the courtyard colonnade; both the courtyard and the colonnade were paved with mosaics, but only fragments of these paved areas have escaped destruction.

Where these and other mosaic fragments survive they tell us something




**RESCUE CAMPAIGN** is under way at 13 locations in Carthage. These are, from north to south, a site north of the Antonine Baths (Danish), a Christian church (Bulgarian), a street and city-wall survey (Italian), a circular monument and wall survey (Canadian), Roman villas (Tunisian), the Punic citadel (French), a site south of the baths (German), the University of Michigan area, a survey near the circus (Polish), two military-harbor sites and a city-wall site (British) and a joint Harvard University–University of Chicago concession.

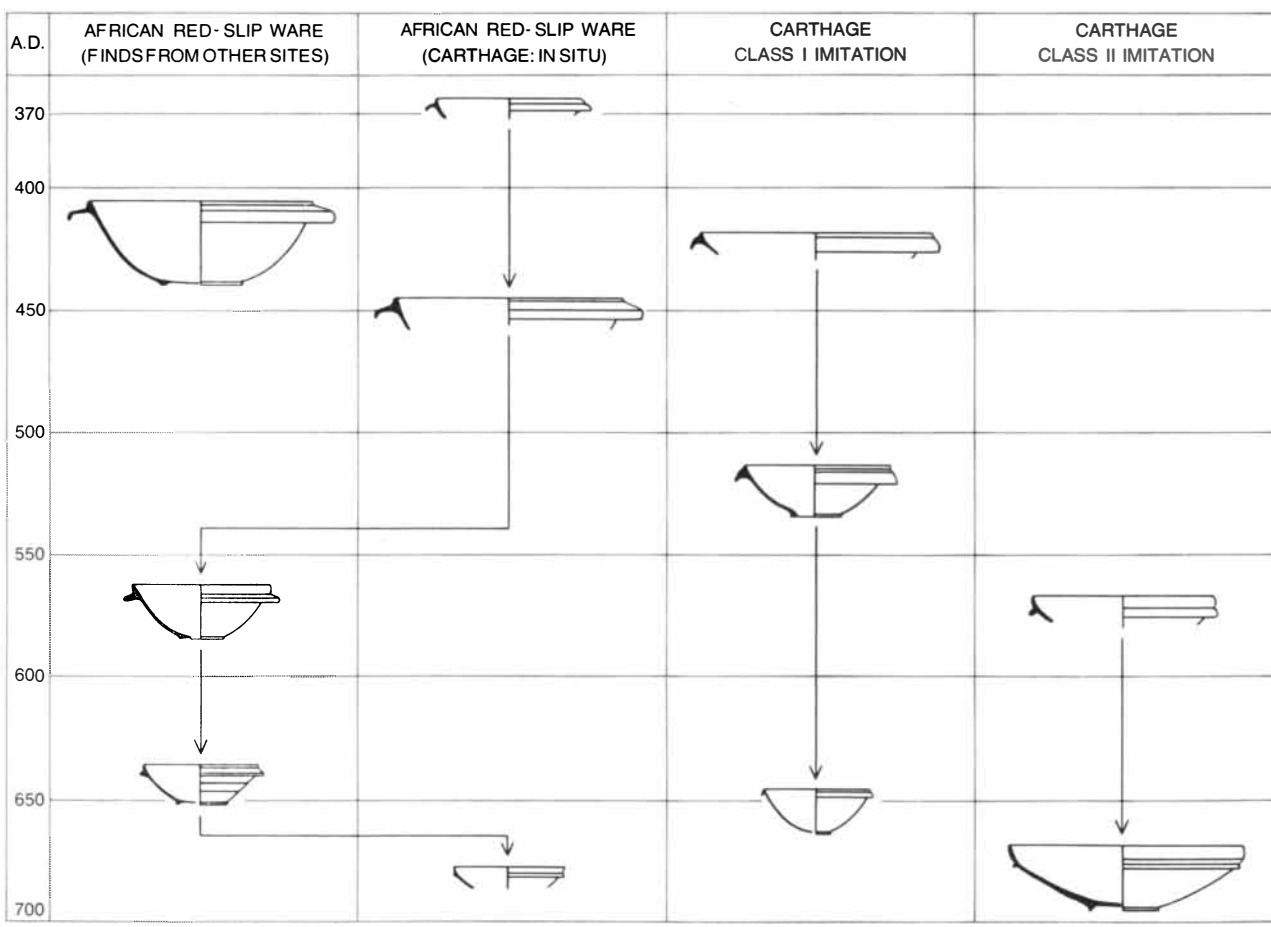
about the occupants of the house during the Vandals' rule. Katherine Dunbabin of McMaster University has analyzed all our mosaics. She finds that as the Roman mosaics wore through or otherwise became outdated the householders continued to put down new mosaics as replacements and even extended the paved areas. For example, remodeling during the Vandal period extended the house eastward until it included the area of the street that had formerly been a plastered sidewalk. The remodeling enlarged the rooms along the east side of the house, and at the same time the floors of the enlarged rooms were covered with new mosaics. Except for these repairs and alterations the century or so of Vandal rule at Carthage appears to have had little effect on what we now call the House of the Greek Charioteers.

Soon after the Byzantine general Belisarius expelled the Vandals from Carthage our house was extensively rebuilt. The work followed the existing floor plan and wall lines, but the dining-room floor was repaved with pieces of marble arranged in an elaborate pattern. This kind of pavement is known as *opus sectile* (the Latin phrase for cut work), and the example in our house is the first Byzantine *opus sectile* floor found in a private house in Carthage that can be confidently dated. The marble pieces have been studied by our geologist, Reuben G. Bullard. He reports that many of them were not freshly cut for the purpose but were salvaged from similar pavements elsewhere and reused here.

The Byzantine work on the house included the redesign of the area between the dining room and the street as an open court paved with flagstones. The remaining rooms fronting on the street received new mosaics, but afterward the house remained unaltered until the third quarter of the sixth century. At that time the courtyard in front of the dining room, which then had a plastered floor, was also covered with flagstones. The house appears to have been occupied until the Arab conquest of A.D. 698, but no new mosaics were laid. The occupants did make an attempt to preserve the *opus sectile* pavement in the dining room. Two distinct phases of repair can be detected. The first repair conformed as closely as possible to the original design; the second abandoned conformity and the repair disrupted the pattern.

When the railroad from Tunis was built adjacent to our site in 1909, the laborers who raised the right-of-way embankment scraped away the top levels of our house for fill. As a result little of the later Byzantine and post-Byzantine record of occupation remains. Fortunately traces of this period of Carthaginian history exist elsewhere on our plot, just to the south of the church and baptistry ruins located earlier by the Tunisian National Institute. We spent our entire 1976 season working in this part of the plot, and we uncovered evidence of five successive phases of occupation that extend from about A.D. 650 to about the 11th century.

During the earliest of the five phases the area had been part of an active ecclesiastical complex that included the church adjacent to our house. Our chronology indicates that the church and its associated structures had reached their final phase sometime before A.D. 647–59. The succeeding phase, which is dated slightly later than A.D. 659–68, marks




**CHANGING POTTERY STYLE over some 300 years is exemplified by a kind of flanged bowl that first became popular in late Roman times. A fine red-slip ware, it was copied throughout the empire; examples made in Africa (*first and second columns*) are found at Carthage and elsewhere. The bowls were imitated by potters in and near Carthage (*third and fourth columns*), who omitted the slip.**

the apparent collapse of the church complex. This collapse is indicated by the beginning of a process particularly familiar to Classical archaeologists: foundation robbing, that is, the theft of building stone from an abandoned site for use elsewhere.

The third phase, dated slightly later than A.D. 668–73, shows evidence of at least three distinct efforts to renew the flooring of the abandoned site; all were of poor quality. It is significant that the second and third phases at the site, which collectively imply less than prosperous and stable conditions in Carthage, both precede by many years the Moslem destruction of the city in A.D. 698. It would seem that the running conflict between Islam and the Byzantine Empire, starting in the Near East half a century earlier, had by this time also unsettled the North African provinces.

The fourth phase at our former church site, dated later than A.D. 687–95 and perhaps as late as the 10th century, is notable for a second major incursion by foundation robbers. The fifth and final phase may with some confidence be placed in the Islamic period: the 10th or even the 11th century. One of the floors of this phase overlaid and sealed an Islamic coin, and it seems safe to presume that the floor is later in date than the coin.

Classical archaeology normally pays little attention to a site in its declining phase, after the monumental buildings have fallen into disuse. Later occupations are likely to be marked by construction less substantial than masonry, for example partitions of mud brick or wattle and daub. Recent archaeological work in England has revealed something of what happened there after the Roman occupation of Britain had ended, but similar studies in other parts of the Roman Empire are rare. Therefore our findings in this corner of post-Punic Carthage throw a valuable light on the last days of the city.

In summary, even after the roofs had collapsed and the walls had been robbed of their stones, the shell of the church complex sheltered a series of domestic habitations where we can still recognize individual housing units. The conditions were mean, the partitions were superficial and the floors were paved not with mosaic or fine mortar but with trampled mud brick. Yet the quantity of coins and potsherds found even in the uppermost levels indicates that the last days of Colonia Justiniana were characterized by intense activity. The advance of Islam had in effect left Carthage besieged. The Christians from the surrounding countryside who flocked to the city evidently welcomed refuge however rudimentary it was.

The animal remains (which we have recovered by dry sifting) and the plant


**QUARRIES FOR CARTHAGE included one quarry on the city outskirts and three near enough to the coast to allow water shipment. Other main quarry sites were located between 12 kilometers (gypsum at Jebel Ayari) and 160 kilometers away (Numidian marble at Simitthu).**

remains (recovered from the sifted soil by flotation) provide an additional dimension to our picture of life in post-Punic Carthage. Bones and shells have not normally been regarded with any great enthusiasm by the Classical archaeologist, and we have been fortunate in having David Reese, a graduate student at the University of Cambridge, analyze our finds of these materials. He has discovered that pigs and sheep or goats were the city's most abundant meat animals. Cattle bones are less common; evidently these larger animals were more often harnessed to the plow and the cart than slaughtered for meat. At Carthage this preference remains the same from the early Roman period to the late Byzantine, and the evidence contrasts sharply with Latin literary sources that would have us believe beef was the most popular meat. That the camel was present here long before the Arab influx is evident from the discovery of camel bones in a pit filled with trash soon after A.D. 100.

Fishbones representative of various species indicate that seafood formed an important part of the Carthaginian diet. Curiously the tuna, the fish that bulks largest in the Tunisian diet today, is not in evidence. A wide variety of mollusk shells, both marine and terrestrial, are present. There is no evidence, however, that the land snails were eaten, and signs of water wear suggest that the marine snails and bivalves were not eaten either but were accidentally gathered along with beach sand that was used to make mortar.

Richard I. Ford and Naomi Miller of the University of Michigan have analyzed the botanical materials from the site. They include a variety of seeds and numerous fragments of burned wood. Olive pits are the most abundant of the seeds. Olives were eaten on the spot not only by the various inhabitants of our plot but also by the workers who prepared mortar for the stonemasons: pits appear frequently in the mortar of all the construction phases. Other seeds Ford and Miller have identified include wheat grains, grape seeds and a single peach stone. Meanwhile a representative collection of wood from various species of trees that grow in Tunisia today has been established, and Classical works are being searched for information about the main sources of timber in antiquity and the kinds of wood preferred for different purposes. The next step will be to see how the burned wood we have salvaged from our excavations

# Physics...analytical math... astronomy and cosmology... chemistry...computer sciences... structural geology...undoubtedly the best books in every scientific area that interests you.

## Take any 3 books, values to $67.50, all for only $3.95 as a member of

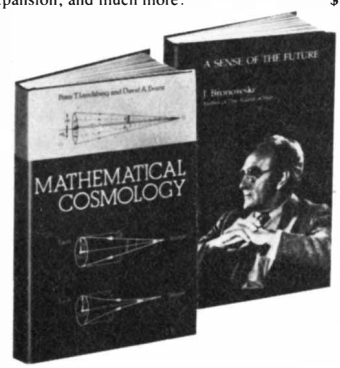
# The Library of Science

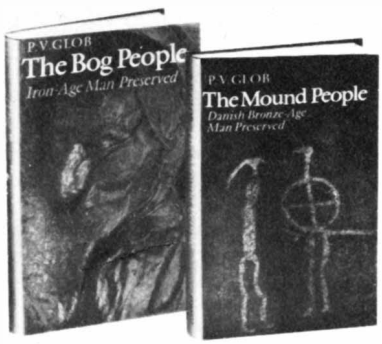
if you will join now for a trial period and agree to accept only 3 more books—at members' discount prices—over the next 12 months

(Publishers' Prices shown)

**36512-2. THE BOG PEOPLE/THE MOUND PEOPLE.** *P. V. Glob.* Perfectly preserved 2,000-year-old bodies buried in peat bogs, all murdered. .6-ft-long oak coffins hidden within mounds of earth...two ancient mysteries solved, revealing two cultures, quite different but equally fascinating. Richly illustrated. *The set counts as 2 of your 3 books.* **$25.00**

**60765. MATHEMATICAL COSMOLOGY.** *Landsberg and Evans.* A mathematicical approach to cosmology that finally bridges the gap between narrative introductions and advanced technical treatises. Covers Newtonian gravitation; particle, continuum, Friedmann, and steady state models; optical effects of the expansion, and much more. **$16.95**



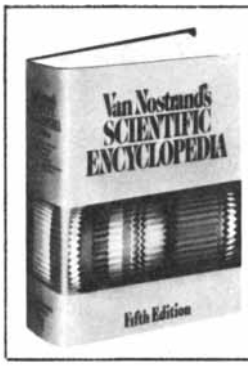




**43065-2. THE EARLY HISTORY OF THE EARTH.** *Edited by Brian F. Windley.* A highly readable, comprehensive look into our planet's past that covers a wide range of topics such as the early earth-moon system, geochronology, archaean tectonics, the atmosphere, the oceans, and life forms. Contains over 390 illustrations, photographs, and tables. *Counts as 2 of your 3 books.* **$45.00**

**77800. A SENSE OF THE FUTURE.** *Jacob Bronowski.* An eloquent defense of science, its methods and goals by the author of *The Ascent Of Man.* In this, his final work, Bronowski probes the creative process of the human mind, examines the limitations and values of science, and speculates on future developments in the sciences, especially biology. **$12.50**

**78235. SHELTER, SIGN, AND SYMBOL.** *Edited by Paul Oliver.* A guided, thorough illustrated tour of the wonders of such exotic locales as Kathmandu; the cave city of Cappadocia, Turkey; and the Greek island of Chios. Also provides a revealing look at the rituals, customs, ancient traditions, and religions of their inhabitants. **$22.50**

**61300. MATHEMATICAL MAGIC SHOW.** *Martin Gardner.* Delightful collection of brain teasers and inspired dissertations by the author of *Scientific American's* popular Mathematical Games column. Includes word squares, double acrostics, card tricks, sophisticated jigsaw puzzles and much much more. **$8.95**




# EXTRAORDINARY VALUE!

## The greatest single-volume source of scientific information available

**00470-3. VAN NOSTRAND'S SCIENTIFIC ENCYCLOPEDIA.** *Edited by Douglas M. Considine.* Nearly 200 experts have contributed to the thoroughly revised and expanded Fifth Edition of the most useful and authoritative single-volume source of scientific information available anywhere. 2 million words. Over 2,500 illustrations. 550 tables. *Counts as 3 of your 3 books.* **$67.50**

**50551. GRANTS: How to Find Out About Them and What to do Next.** *Virginia P. White.* A "must" for anyone who wants to get, keep, or increase grant funding. **$19.50**

**61610. MAYA.** *Charles Gallenkamp.* Mayan civilization from its glory to extinction. A compelling view of the life, arts, religion, and science of this unique civilization. **$12.95**

**65640. ON STONEHENGE.** *Sir Fred Hoyle.* Was Stonehenge man's earliest astronomical observatory? The mystery of the British Isles' most celebrated megalith is tackled by the world renowned astronomer, who proposes that the 5,000-year-old structure was built to predict lunar eclipses. Absorbing reading for both scientist and layman. **$9.75**

**69305. PLANETARY GEOLOGY.** *Nicholas M. Short.* A comprehensive, non-technical account of solar system geology. Data from the Apollo landings and recent unmanned probes provide clues to questions concerning lunar and planetary origin, evolution, and composition. Lavishly illustrated with more than 500 photographs, line drawings and graphs. **$21.95**


**50380. GLOSSARY OF CHEMICAL TERMS.** *Clifford A. Hampel and Gessner G. Hawley.* Over 2,000 concise definitions: all major chemical groups, important functional terms, basic phenomena and processes, all the chemical elements, and more. **$14.95**

**59800. MAN DISCOVERS THE GALAXIES.** *Richard E. Berendzen, Richard C. Hart & Daniel Seeley.* Topical, current look at modern astronomy dealing mostly with the discovery and arrangement of our Home Galaxy, the Milky Way System. **$15.95**

**61530. MATHEMATICS DICTIONARY. Fourth Edition.** *Edited by Robert C. James and Edwin F. Beckenbach.* Revised and expanded edition of this leading mathematics dictionary offers clear, accurate definitions of more than 8,000 terms. **$17.95**

**50265. GEOMORPHOLOGY AND TIME.** *J. G. Thornes and D. Brunsden.* An invaluable schematic guide that focuses on the time factor in current geomorphological research. Filled with hundreds of figures, diagrams, and tables, it reviews new ideas, concepts and approaches in the field. **$12.95**

**49267. FRONTIERS OF ASTROPHYSICS.** *Edited by Eugene Avrett.* A contemporary source of astrophysical topics from recent solar research to neutron stars, supernovae and intergalactic matter. **$20.00**

**55000. THE ILLUSTRATED ENCYCLOPEDIA OF ASTRONOMY AND SPACE.** *Ian Roxburgh.* Covers the full range of astronomical knowledge and space exploration, from ancient times to now. **$16.95**

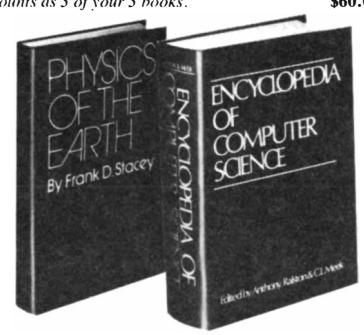
**66570. AN OUTLINE OF STRUCTURAL GEOLOGY.** *Bruce E. Hobbs, Winthrop D. Means and Paul F. Williams.* Up-to-date account of every aspect of structural geology—from the nature of stress and strain to the latest information on plate tectonics. **$16.95**

**34140. APPLIED ANALYTICAL MATHEMATICS FOR PHYSICAL SCIENTISTS.** *James T. Cushing.* Deals with topics ranging from spectral analysis of linear operations to complex variables, to group theory. **$19.95**

**82660. TECHNOLOGY MATHEMATICS HANDBOOK.** *Jan J. Tuma.* Instant access to concise summaries of all major definitions, formulas, graphs, tables of elementary and intermediate mathematics. **$15.95**


**36395. BLACK HOLES, QUASARS AND THE UNIVERSE.** *Henry J. Shipman.* Black holes, quasars, white dwarfs, neutron stars, supernovae, Seyfert Galaxies—explores the frontiers of astronomy. **$12.95**

**44900-3. THE ENCYCLOPEDIA OF COMPUTER SCIENCE.** *Edited by Anthony Ralston and Chester L. Meek.* A monumental 1,550 page volume, offering condensed, accurate information on essential topics ranging from theory to applications, from program structures to the logical design of central processing units, from Boolean algebra to microprogramming. *Counts as 3 of your 3 books.* **$60.00**




**54995. THE ILLUSTRATED ENCYCLOPEDIA OF ARCHAEOLOGY.** *Glyn Daniel, General Editor.* Worldwide coverage of the people, places, and things of the past with over 1,000 entries and more than 100 illustrations. The significant sites, cultures, civilizations, terms, tools, and techniques packed into one all-inclusive volume. **$17.95**

**34430. ARTIFICIAL INTELLIGENCE AND NATURAL MAN.** *Margaret Boden.* Brings you up-to-date on the latest work in artificial intelligence as it confronts the complex philosophical, psychological, and social issues raised by man's continued use of thinking computers. Informative and thoughtful reading for both scientist and layman. **$17.50**

**52292-2. A HANDBOOK OF NUMERICAL AND STATISTICAL TECHNIQUES.** *J. H. Pollard.* Here are all the statistical and numerical methods you need to solve the mathematical problems on a programmable calculator, minicomputer or interactive terminal. Examples are drawn mainly from the life sciences. *Counts as 2 of your 3 books.* **$24.95**

**40167-2. THE CONDENSED CHEMICAL DICTIONARY.** *Revised by Gessner G. Hawley.* Not just a dictionary, it covers technical data and descriptive information covering thousands of chemicals and chemical phenomena to meet today's needs. Over 18,000 clear, accurate entries. *Counts as 2 of your 3 books.* **$32.50**

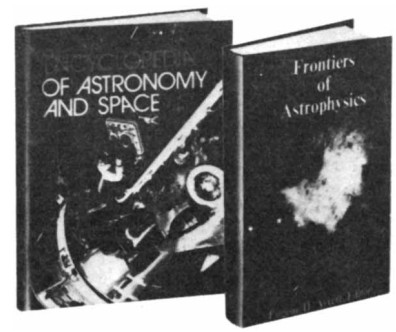
**61660. MECHANICS OF THE MIND.** *Colin Blakemore.* Reports discoveries now being made about the human brain: why we have an awareness of self, how we feel and see the world we live in, why we sleep, how we remember. 200 beautiful photographs and drawings. **$19.95**

**73240. QUASARS, PULSARS, AND BLACK HOLES.** *Frederic Golden.* A thorough review of the spectacular discoveries in astronomy. Illustrated with drawings and color photographs. **$7.95**

**75710. SCIENCE BRAIN-TWISTERS, PARADOXES AND FALLACIES.** *Christopher P. Jargocki.* Over 160 puzzles, questions, problems and situations, each based on scientific principles, with detailed explanatory solution. Great fun. **$9.95**

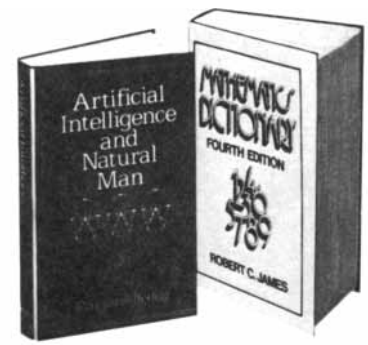
**40690. COSMOLOGY.** *Michael Rowan-Robinson.* Superbly illustrated with space photographs and numerous star charts and graphs, this fascinating blueprint brings together all the important ideas and developments in the new astronomy. Includes twenty selected controversies in cosmology today. **$13.95**

**56355-2. AN INTRODUCTION TO PROBABILITY THEORY AND MATHEMATICAL STATISTICS.** *V. K. Rohatgi.* Exceptionally thorough and well-balanced examination covers topics ranging from the functions of a random variable to the sequential probability test, from modes of convergence to the regression model. *Counts as 2 of your 3 books.* **$22.95**




**68780. PHYSICS OF THE EARTH.** *Frank D. Stacey.* Covering rotation, gravity, internal structures and plate tectonics, the completely rewritten second edition of this highly acclaimed work also includes the first comprehensive look at rotation and excitation of the wobble, tidal friction, earthquake prediction, thermodynamics of convection and melting and anelasticity. **$16.95**

**39746-2. COMPANION TO CONCRETE MATHEMATICS. Volumes I and II.** *Z. A. Melzak.* Mathematical techniques and various applications, with historical notes, solutions, on geometry, iteration, series and products, integrals, more. *The set counts as 2 of your 3 books.* **$46.45**




If reply card is missing, please write to
**The Library of Science, Dept. 2-AIA**
Riverside, N.J. 08075 to obtain membership information and an application.

fits into such a conceptual framework.

A similar effort has been launched with regard to the varieties of building stone used at Carthage. Both the immediate hinterland of the city and more distant areas of Tunisia and even Algeria are well known for abandoned quarries of both Punic and Roman times. What has been less apparent, however, is how the various quarry products were utilized.

Bullard, our geologist, has studied the question. He finds that the Romans followed Punic practices with respect to a porous coarse sandstone taken from Ghar el Kebir, a quarry some 60 kilometers northeast of Carthage. This stone, which could be brought to the city by water, was used as a general construction material. A local lithified sand and a quartz sandstone from the Hamilcar quarries just north of the city were used by the Romans as a filler rather than for surface masonry. Durable and easily worked limestone was available from quarries at Jebel Djelloud and Jebel er Rorouf, from which the stone could also be brought by water, and at Jebel Ayari, an overland haul of some 12 kilometers. The Romans used this excellent stone for posts, lintels and paving blocks and on occasion for wall courses.

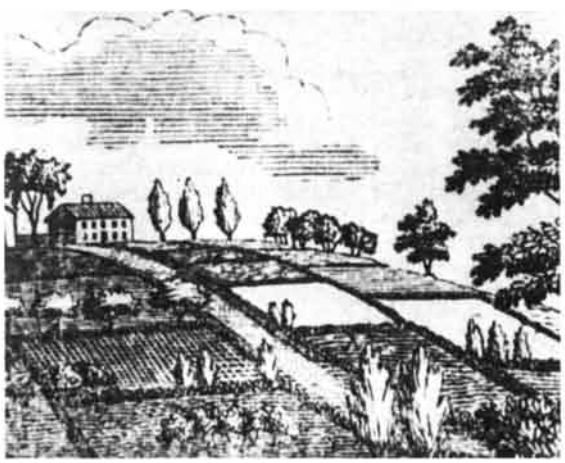
For elaborate facings the Romans quarried a near-marble, gray with pink and white veins, at an outcropping of Jurassic sediments some 60 kilometers southwest of Carthage near Thuburbo Maius. For massive columns they selected a white-and-brown-veined limestone from an adjacent deposit. Gypsum-rich clays from a deposit at Jebel Azeis, north of Thuburbo Maius, and quantities of gypsum available even closer to Carthage account for the frequent use of molded gypsum blocks in city construction. The marble quarry at Simitthu, 160 kilometers west of Carthage, was the Romans' source of *marmor Numidicum,* a mottled pink-and-yellow facing stone that was not only used in the city but also exported in large quantities to cities all over the empire.

As our work continues we expect to link still other Tunisian quarry sites to the building stones of post-Punic Carthage in the same way that our other interdisciplinary workers are adding to a growing sum of knowledge. An immediate example is the work conducted throughout our 100-by-400-meter field last year. Surveys using an electrical-resistivity meter in the southern half of the field have yielded important information on the presence and absence of cisterns, information that will allow value judgments with respect to such broader topics as architectural development, town planning and population density. Together with our colleagues in this international rescue effort we are with each season expanding the traditional bounds of Classical archaeology.

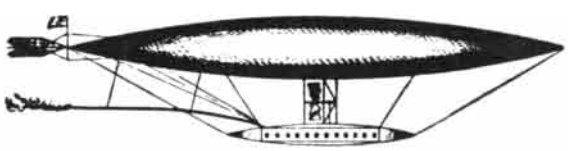
SCIENTIFIC AMERICAN
Filmstrips

# Rufus Porter
## *A Scientific American*


"We have been frequently solicited by enterprising mechanics to engage in the publication of a new scientific paper, for the advancement of several new, curious and useful arts which have recently been discovered and introduced . . ."



"It is not infrequently the case that the productions of self taught artists far surpass in excellence those of regular bred artists who have studied with the most popular masters . . ."



"It may be anticipated that within a few months these aerial machines may be seen soaring in various directions and at different elevations, some above the clouds, and others, like swallows, sailing leisurely just above the surface of the earth."

Rufus Porter—artist, inventor, journalist and founder, in 1845, of SCIENTIFIC AMERICAN—should be better known to his countrymen. His exuberant murals adorn 100 New England homes and are finding their way into museums as treasures of American art. His long list of inventions was crowned before 1850 by his demonstration of a self-propelled flying machine. His SCIENTIFIC AMERICAN spoke for the buoyant rationalism that committed this country to industrial revolution from early in the 19th Century.

The 20-minute filmstrip for DuKane projectors *Rufus Porter, A Scientific American* [$22, shipping included] brings this Yankee Leonardo back again to life. In his own colorful words and pictures it recounts the story of his untiring enterprise and places him in the context of the era in American history he so well personified.

Recommended for High School and Junior High School classes in American History, Social Studies and Art History. Available in filmstrip and tape cassette for DuKane projectors.

SCIENTIFIC AMERICAN, Inc.
415 Madison Avenue, New York, N.Y. 10017, Dept. RP

Please send me _____ copies of *Rufus Porter, A Scientific American,* a filmstrip presentation for DuKane projectors, at $22 (shipping included). I understand I may return same and receive a full refund if I am not completely satisfied.

Enclosed is my check for $ _____.
Please include appropriate sales tax.

Name (please print) _____

Institution (if any) _____

Address _____

City and State _____ Zip _____

# The Visual Characteristics of Words

*What are the visual cues that good readers use to recognize letters and words? Tests seeking such insights may provide a basis for instruction in more effective reading strategies*

by Peter Dunn-Rankin

Over the past 10 years or so I have conducted a series of experiments aimed at exploring an aspect of the general question of how people learn to read. My particular interest has focused on the visual characteristics of words, a subject that seemed at first to be narrowly defined and comparatively simple. In effect the purpose of this investigation has been to seek an answer to the specific question: What is it that a person sees when he looks at a letter or a word?

From the beginning my experiments have been based on the assumption that words are complex graphic stimuli containing a variety of potential cues that can be utilized in letter and word recognition. The difficulty many children have in learning the order and direction of letters within words indicates that the ability to recognize letters and words is not determined solely by innate visual capacity. For example, many children have not yet learned to distinguish between words containing different arrangements of the same letters. The words "saw" and "was" are often confused in an error known as a reversal. Some letters, particularly in a lowercase English type font, can be rotated or inverted to achieve similarity. Thus p can be obtained by inverting b. Words containing these letters, such as "pig" and "big," are sometimes puzzling to beginning readers. Many letters, of course, are already similar without rotation or inversion. For instance, letters such as h and n, or e and c, may be perceived incorrectly, and words such as "hot" and "not," or "cat" and "eat," may therefore be confused.

Most children are left on their own when it comes to learning how to recognize letters and words. Although some external agent—a parent or a teacher—usually reinforces the correct response, the particular features a child uses in remembering a given letter or word appear to be largely a matter of trial and error. Accordingly different children develop different strategies for recognizing graphic stimuli, and some strategies are clearly more effective than others.

It is easy to think of cases of inefficient visual processing. For example, suppose I am teaching a young child the lowercase letters. The letter b is printed on a white card. In the upper right-hand corner of the card a black smudge is all that remains of a spider that somehow lost its way into the stack of cards. I hold the card up and say: "This is a b. Can you say the letter b?" Assuming I get an affirmative reply, I might then say "Good" and go on to repeat the process with other letters. Why, however, does the child remember that this particular graphic symbol has the name b and not h or d? If he remembers from the mark of the dead spider, the strategy cannot be termed efficient for future use, because few other b's are likely to have this particular visual cue associated with them.

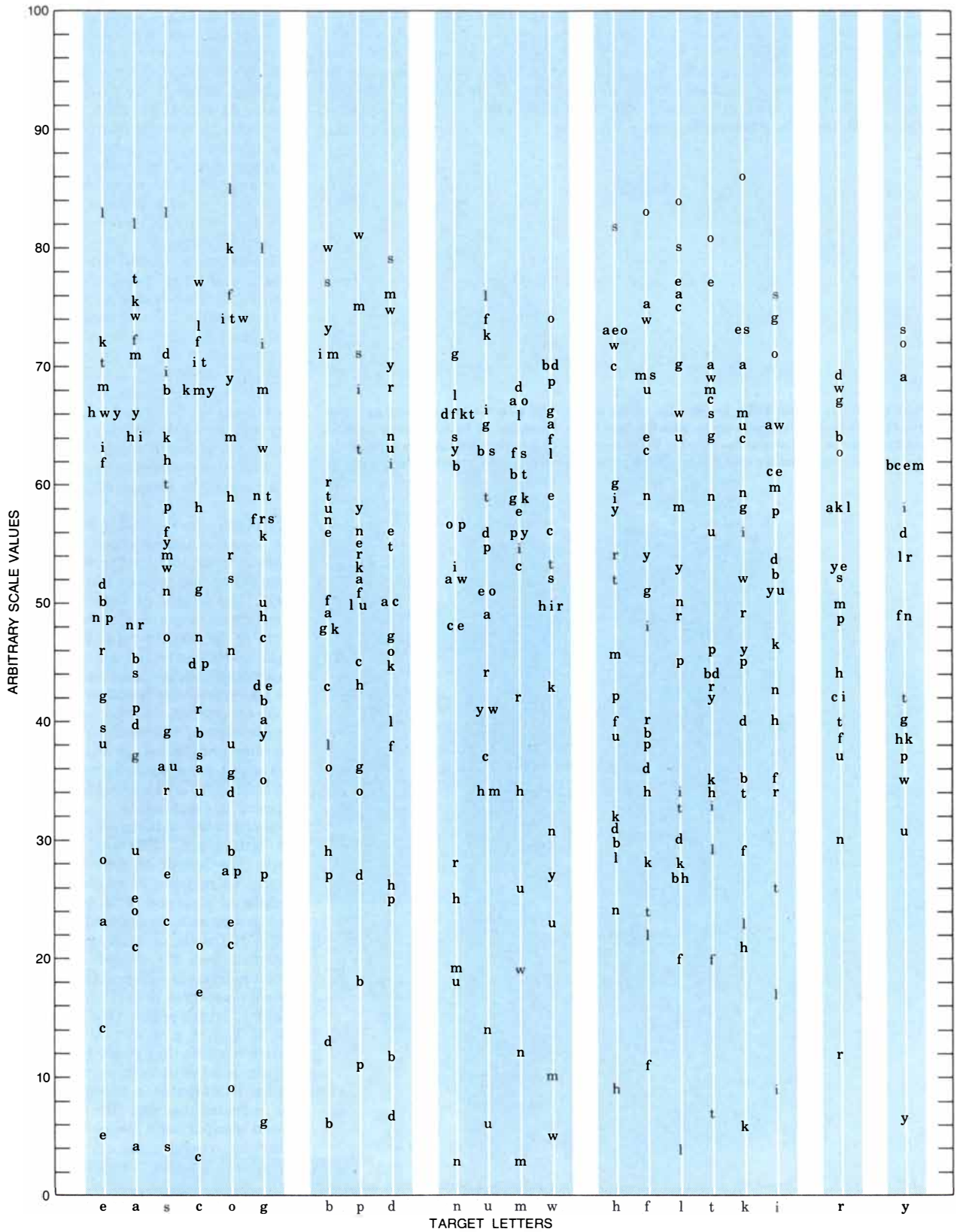
If good readers consistently notice certain attributes of letters or words and poor readers look elsewhere, the possibility arises of making instruction in the visual strategy of reading more effective. Before methods of teaching effective cues can be entertained, however, it is necessary to isolate the most significant visual features of letters and words. Unfortunately the strategies good readers use in making visual discriminations remain elusive, because they are formulated intuitively. Cues and rules are normally not verbalized, because most of them are not clearly understood as such by readers.

The basic visual features of the lowercase letters have been analyzed in many ways. Most investigators, however, consider such letters as being made up of five simple components. These features can be remembered by the mnemonic word "loves," which graphically represents a vertical line (l), a closed circular curve (o), an angular intersection (v), a horizontal line with a curve (e) and a cyclical curve (s). Almost all the English lowercase letters can be constructed by varying the direction of the basic modules represented by these five letters. Here simplicity can be a disadvantage for children who have not yet learned to distinguish directions, since the English lowercase letters present many rotational similarities. For instance, the four letters p, b, d and q are rotationally similar, as are n and u, f and t, m and w, and s, z and a.

In most studies of letter discrimination one of the letters in the test set is the same as the target letter. Such trials are called absolute discrimination tasks because the subject either matches the letter or does not. In experiments of this type the number of errors is normally very small. Hence it is possible to establish only a few gross categories of discriminability, based on an average rate of error of, say, zero percent and 5 percent.

In a study that I conducted in 1968 all 210 possible pairs of 21 English lowercase alphabet letters were tested by comparing their relative similarity to each target letter. By analyzing the cumulative choices made by 315 second- and third-grade children over all possible pairings, sets of linear scale values could be assigned to the lowercase letters in terms of their relative similarity to every other letter in the test alphabet [*see illustration on opposite page*]. The letters not included in this study, because of their comparative infrequency in English, were j, q, v, x and z. Each of the 21 scaled letters was plotted in a frame of reference running from zero to 100. Zero indicates no error in choosing the letter most like the target letter, and 100 indicates no error in choosing the letter most unlike the target letter. The scales are organized by similar letter groups.

The scales in this study have been

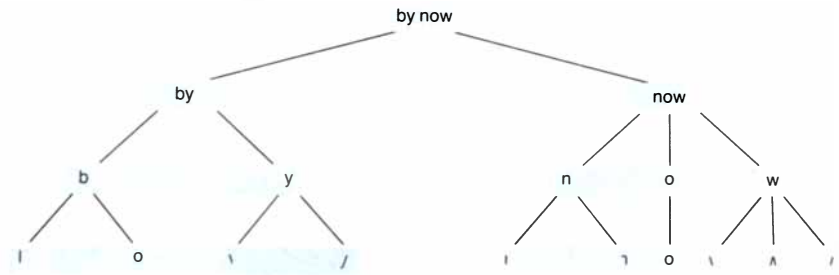



**LETTER-SIMILARITY SCALES** were constructed by the author on the basis of a letter-discrimination test in which 21 lowercase letters were matched in pairs to determine their relative similarity to one another. The cumulative choices made by 315 second- and third-grade children over all 210 possible pairings were then analyzed, and sets of linear scale values were assigned to the letters in terms of their relative similarity to each target letter. (The comparatively infrequent letters not included in the study were j, q, v, x and z.) In the arbitrary frame of reference in which the letters are plotted here zero indicates no error in choosing a letter most like the target letter, and 100 indicates no error in choosing the letter most unlike target letter. The scales are organized by similar letter groups (*colored bands*).

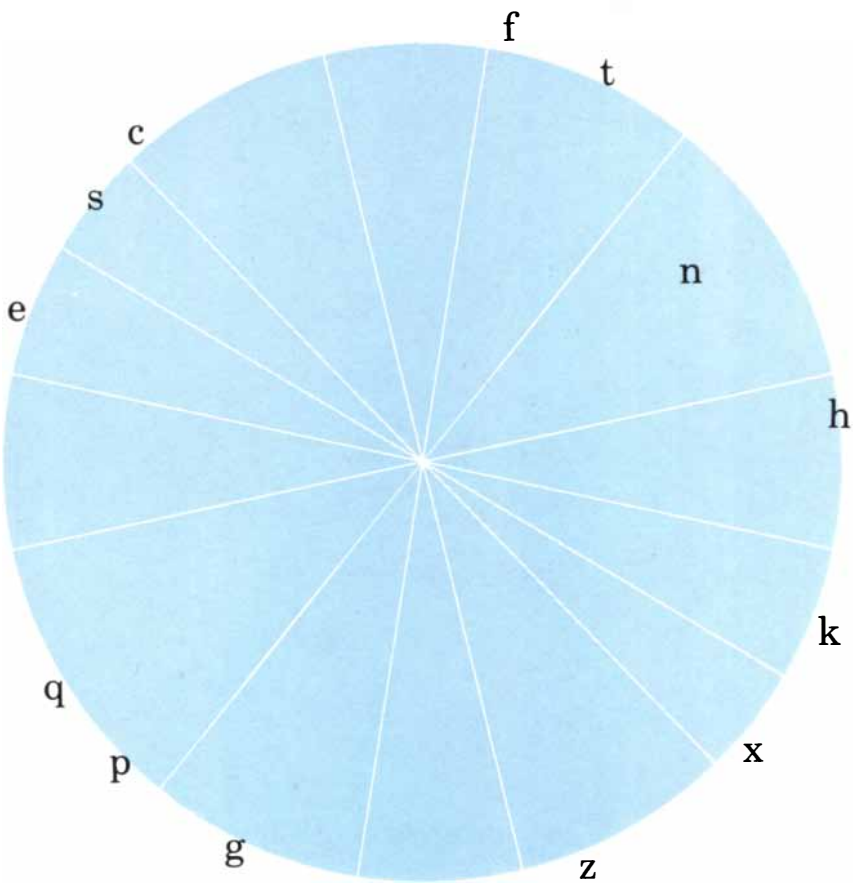
adopted in many other experiments on visual perception. They provide a basis for letter selection in discrimination-learning tasks and data for the construction of tests to diagnose the visual-discrimination ability of children learning to read. In addition they are useful for studies of word errors resulting from letter confusion.

Starting with the five basic letter modules, a hierarchical learning process can be visualized in which "modules of form" (curve, angle, vertical line and so on) are used as the feature cues for letter recognition. In this case letters are the cues for the recognition of syllables and words, and syllables and words are in turn the basic features necessary for recognizing phrases. According to this model, the recognition of the phrase "by now" can be thought of as being developed in such a way that at any one level of integration no more than $3 \pm 2$ units of a lower level can be combined [*see top illustration at left*]. Phrases rarely contain more than $3 \pm 2$ words; words rarely contain more than $3 \pm 2$ syllables; syllables rarely contain more than $3 \pm 2$ letters, and letters rarely contain more than $3 \pm 2$ letter features.

This model of the hierarchy of visual features is incomplete because it lacks a representation of how the primary features are integrated into larger units. Several workers have suggested that the processing of the visual features of letters and words is serial, or sequential, but recent research suggests that familiar stimuli are perceived and processed in a parallel, or integrated, fashion.




**HIERARCHICAL ANALYSIS, in this case of the phrase "by now," views reading as a progressive feature-recognition process in which a few basic "modules of form" (curves, angles, vertical lines and so on) serve as the cues for the recognition of letters, which in turn serve as the cues for recognizing syllables and words, which in turn serve as cues for recognizing phrases.**




**LETTER WHEEL represents the results of a multidimensional letter-discrimination analysis in which a record was made of the reaction time of 45 children and adults to a set of 13 letters paired in all possible ways. In this experiment the two paired letters were first hidden by a shutter. When they were exposed, a clock was started. As soon as the subject being tested pressed a switch indicating that the letters were "different" the clock stopped. The length of time it took for the subject to respond was then used as a measure of letter similarity. The results suggest that the important visual features of letters are perceived not as being immutable but rather as melding into each other in a continuous way. In this circular representation extremely unlike letters tend to be opposite each other. The letter n is not on the circumference of the wheel because it shares certain characteristics with e, s and c; its left vertical component and its striking similarity to h, however, place it closer to the letters with vertically ascending components.**

Evidence for the latter hypothesis is found in an analysis of the reaction times of individual subjects to a complete set of letters paired in all possible ways. The length of time it takes for someone to respond to such stimuli appears to be closely related to reading level, particularly for mature readers. Important findings have resulted from the analysis of the reaction times of 45 children and adults who responded to the question of whether two letters were the same or different. In this experiment two letters, placed side by side on a card, are hidden by a shutter. When they are exposed, a clock starts. As soon as the subject presses a switch indicating that the letters are "different," the clock stops. The length of time it takes the subject to respond is recorded in hundredths of a second. The reaction time then serves as a measure of letter similarity. A matrix of these similarities is analyzed with the aid of a technique called individual-differences scaling [*see bottom illustration at left*].

The 13 letters selected for this study (f, t, n, h, k, x, z, g, p, q, e, s and c) were chosen because they could be combined in various ways to form pairs containing similar letter features. More letters were not included because the increase in paired comparisons makes the task too tiring for young children. The multidimensional scaling analysis takes the form of a circular pattern of letters much like a color wheel. This representation suggests that the dimensions of letters are not immutable but integrative; in other words, one letter melds into the next in a continuous way. A general division can be suggested for the 13 letters (in terms of angle v. curve or ascender v. descender). The letter k is opposite e, and t is opposite g, for exam-

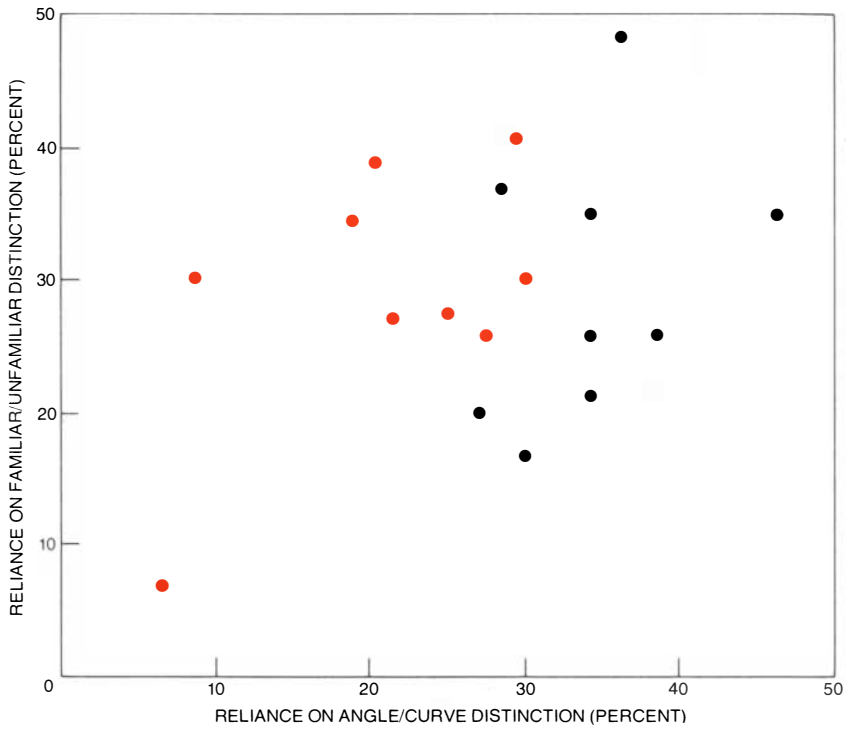
ple. At any point on the wheel, however, both factors must be considered.

The letter wheel indicates that a parallel processing approach to the perception of familiar letters is preferable to a serial model. Most of the 13 letters contain two or more of the basic constituents, and an analogy between primary-color combinations and letter-feature combinations is strongly suggested. Just as orange is seen as a unique color even though it is a combination of red and yellow, so are the letters of the letter wheel seen as integrated units that are combinations of basic features. In this model x combines with l to produce k, and l combines with n to produce h, yet h and k are perceived as wholes.

Further reinforcement for an integrative perception of familiar units comes from the fact that the two basic dimensions indicated (curve v. angle, ascender v. descender) are in general relied on equally by all subjects. A unique by-product of the newer scaling methods is that individual weights can be determined for each subject on each dimension found. When I plotted the positions of the 45 subjects with respect to the use they made of these two dimensions, it was clear that none of the subjects, young or old, good reader or poor, chose one feature exclusively. Some subjects, however, recognized the combination of the two basic features more readily than others. These subjects were also the most mature readers in the sample. Such a result might be an artifact of consistency of response or might reflect the importance of these dimensions to perception in reading, or both.

Immature and mature readers can also be distinguished in terms of their reliance on letter frequency for recognition. Where adults will distinguish between a and z because one is curved and the other is angular, very young children make the distinction because a is familiar and z is unfamiliar, confirming the idea that frequent exposure to stimuli such as letters provides a level of satiation necessary for automatic response. It follows that the identification of simple stimuli such as letters should not serve as a variable in tests aimed at identifying good and poor readers beyond the second or third grade. More complicated stimuli (such as words), however, do provide pattern-discrimination tasks at a level that is complex enough for determining good and poor readers in all elementary grades.

I first tested my developmental hypothesis for the visual discrimination of words with a device called the word-preference inventory. In this experiment I would show the child a target word such as "sore" and present two pseudowords similar to "sore" but containing specific kinds of errors. For the target word "sore" I might present "ssore" (in which the extra s is called an addition)



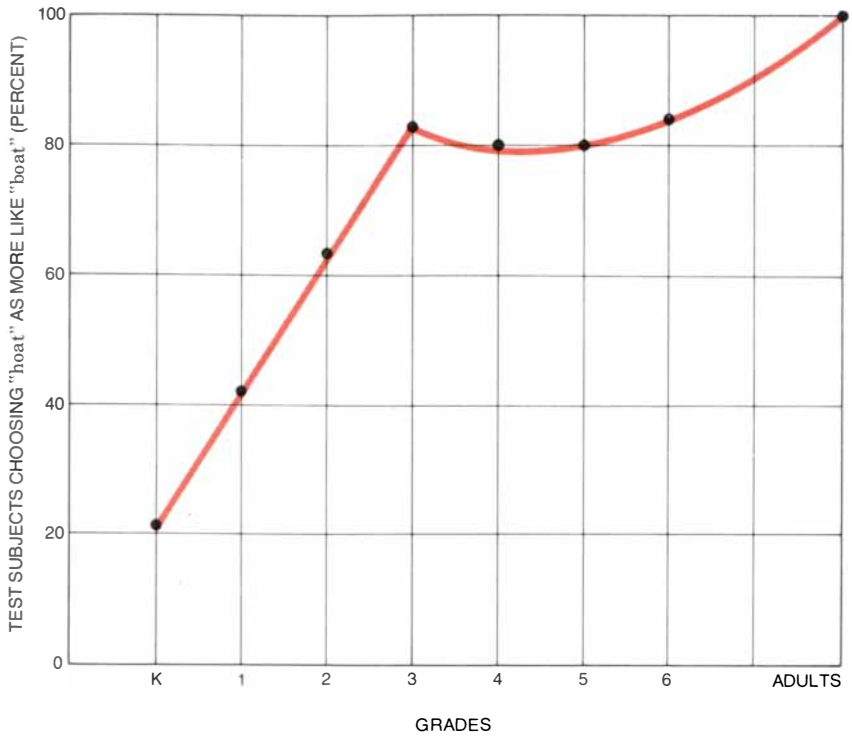

**IMMATURE AND MATURE READERS can be distinguished in terms of their reliance on letter frequency for recognition. As this two-dimensional graph shows, adults (*black dots*) tend to distinguish the letters a and z because one is curved and the other is angular, whereas very young children (*colored dots*) make the distinction because a is familiar and z is unfamiliar.**

and "core" (in which c is a misperception of s). Mature readers usually choose "ssore" as more like the target, whereas younger readers choose "core" as more like it. When adults and kindergarten children were asked which pseudoword, "hoat" or "qoat," looked more like the target word "boat," 85 percent of the adults preferred "hoat," whereas 80 percent of the children chose "qoat." (In these pseudowords h is defined as a misperception of b in the target word, whereas q is a rotation of b.)

The distribution of choices in this test for children in elementary school suggests that an adult view of word similarity only begins to be approximated by the time a child reaches the third grade and that even in the sixth grade there are differences in pseudoword choice between adults and children. Somewhere between the second and fourth grades it appears that most students change their strategy of reading: they start guessing what the words are from the context and start using phonetic cues to help them guess. For a lower socioeconomic group the change might be observed between the third and fifth grades. Approximately 10 percent of the children tested in the fourth and fifth grades revealed immature and probably ineffective visual strategies. These children were also found to be below-average readers.

The Swedish psychologist Benny Brodda, in the course of testing whether or not prospective trade names in Sweden are similar to existing ones, has found that the letter order of the two words being compared is the most important factor in word similarity. He uses letter order as the primary basis for his test of whether or not two words are similar. I have also used letter order in test items in my word-preference inventory. Some poor readers will consistently choose words that are not in order, that is, words that represent reversals or permutations of the original target word, whereas good readers choose other errors instead.

In order to assess the phonetic bias of the word-preference inventory, the 51-item test was also administered to 240 native Japanese children in the fourth and fifth grades. When the responses of the Japanese children were compared with those of the American children, the basic factors of letter order and word unity were found in both cultures. A phonetic bias was not found. The distinction between rotation and misperception, a factor noticed in the American population, was not made by the Japanese children. An explanation for this difference is that the traditional Japanese script has no characters that can be made equivalent by rotation. The Japanese have been modernizing their type fonts, however, and in at least one of the new fonts some of the characters are rotationally equivalent. This ty-




**DIFFERENT READING STRATEGIES employed by children and adults are reflected in this graph, which shows the distribution of responses to a question in one of the author's tests, called the word-preference inventory. Both the children and the adults were asked which pseudoword, "hoat" or "qoat," looked more like the target word "boat." At one extreme 80 percent of the kindergarten children tested chose "qoat" (in which q is a rotation of b), whereas 85 percent of the adults preferred "hoat" (in which h is a misperception of b). The adult view of word similarity only begins to be approximated by the time a child reaches the third grade. At about that stage most children appear to change their reading strategy: they begin to guess what the words are from the context. For below-average readers the change is observed somewhat later.**

pographical simplification may create some directional reading difficulty for Japanese children where there has been none.

A different kind of study of word similarity involves taking approximately 100 words from a newspaper article and having people group words they think look most alike. Analyses of this kind indicate that there are four basic features adults rely on in determining whether words are visually similar. The first factor is whether or not the beginning letter or letters are the same. The second factor is similarity in word length. The third factor is similar word endings such as "-ing," "-ed," "-ly" and "-igh." The fourth factor is similar internal letters or letter combinations. The combinations could include letters with ascending components (such as l, d and t) or those with descending ones (such as g and y). Letter combinations, either adjacent or separated in the word, are important visual cues used in word recognition. The beginning letter, however, is the dominant visual characteristic of a word. Its importance is also confirmed by other studies.

Word length has been questioned as an effective feature in word perception and recognition because of its high correlation with infrequency of usage: long words are less familiar. In a recent study, however, I was able to confirm the importance of word length as a visual feature utilized in estimating word similarity. I conducted an individual-differences scaling of 10 words that begin with the same letter, using measurements both of reaction times and of percent overlap to determine similarity, and the results were almost identical. The dominance of word length was apparent from this study.

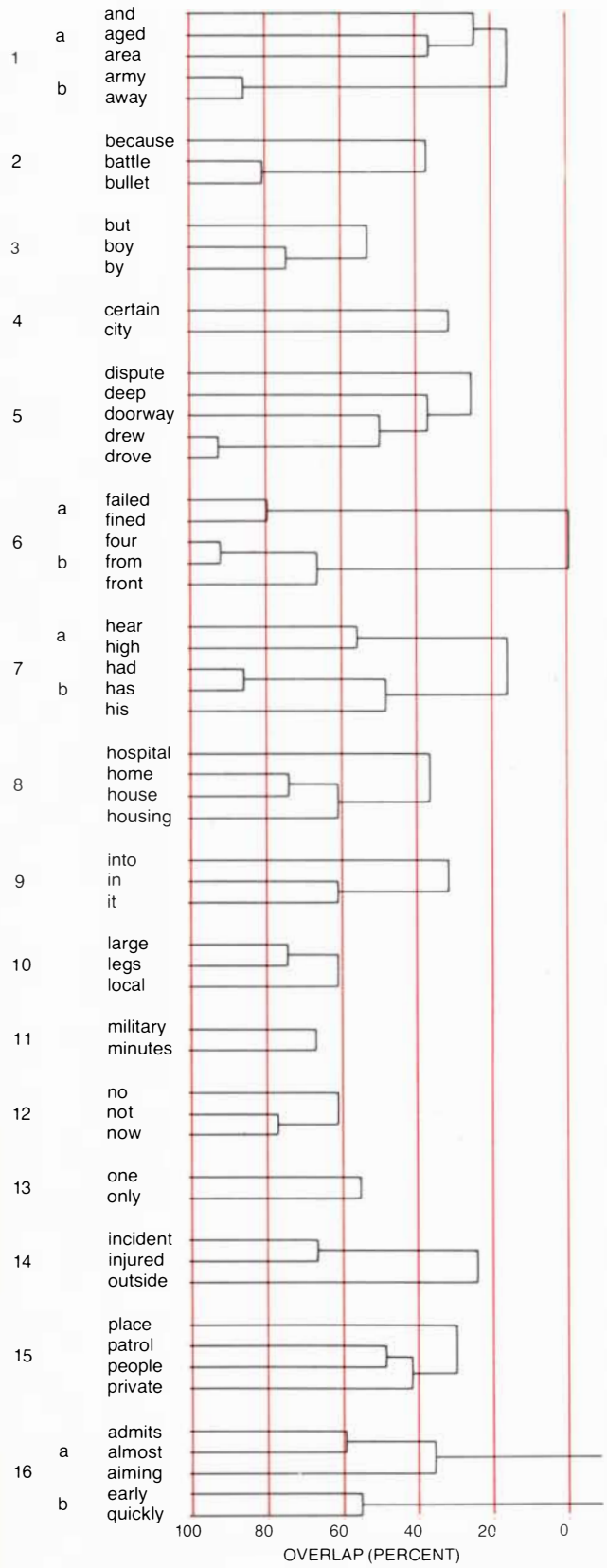
Since there are many words that begin with the same letter and many words of the same length, initial letter and length cannot be independently useful as cues to word recognition. I therefore made a survey to determine the frequency with which both features are found in elementary-school textbooks. The results showed that fewer than 1 percent of the possible pairs of different words have both features in common, suggesting that their combination can be an important visual characteristic in word perception. One has only to take any newspaper article and count the number of word pairs having the same length and also beginning with the same letter to realize that the total is less than 1 percent of the possible pairs.

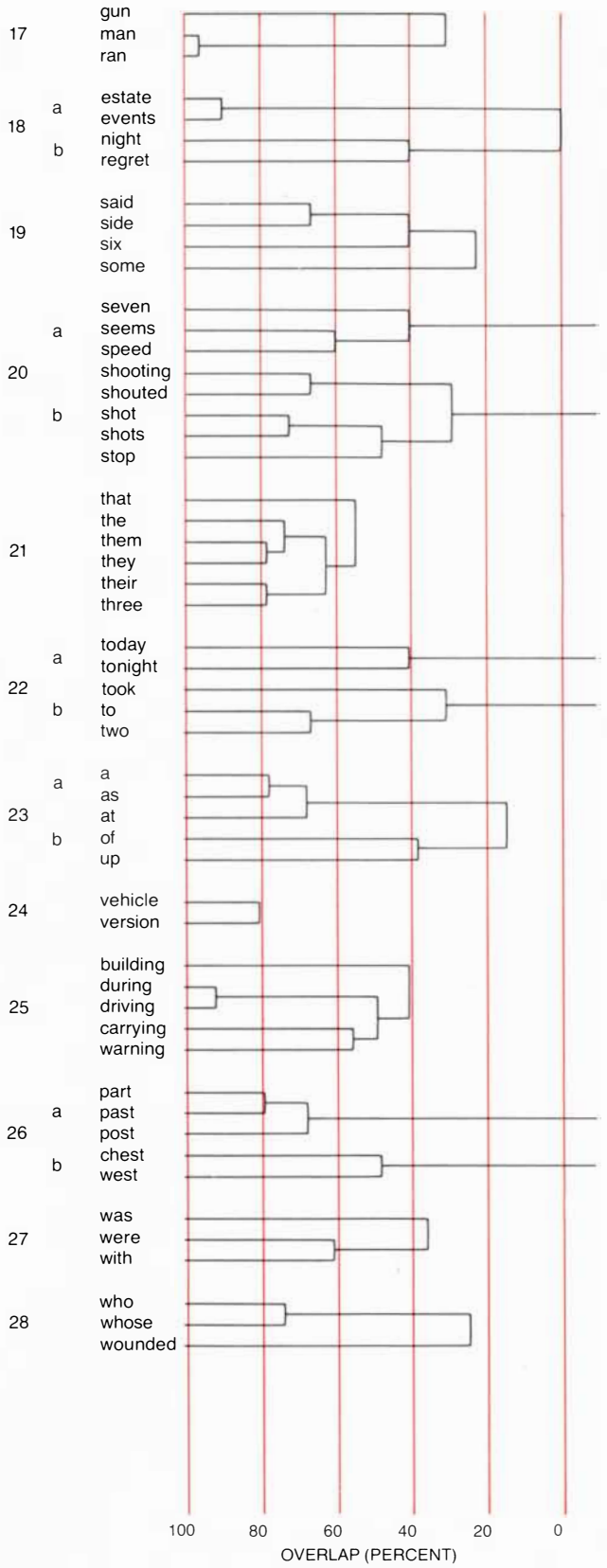
Some of my experiments have involved the use of afterimages: visual images that persist after the stimulus has ceased. I first noticed the usefulness of the afterimage as a representation of the visual focal point during a camping trip in 1969. A gas lamp was hung in the tent where I was reading, and I was drawn to stare at it for some reason, perhaps because it was not working well. When I looked back at the page, I saw the afterimage of the bright gas mantle clearly superposed on the printed material at each visual fixation point. I immediately began dotting the fixation points as I read.

Four years later, during a sabbatical at the University of Stockholm, I began to study this phenomenon systematically and to use it in the search for effective cues to letter and word identification. (I was unaware that Hermann von Helmholtz had used a similar technique almost a century earlier.) At that time I discovered by accident that the apparent size of the image is a function of the distance at which it is projected. By creating afterimages from small light sources at a distance of two or three meters the apparent afterimage can be made quite small when one looks at a surface held a short distance away. The image can, in fact, be made small enough to be accurately pinpointed within a line set in a common textbook type font.

In experiments exploiting the afterimage phenomenon the subject is asked to fixate on a screened 40-watt lamp projected from a dark, unlighted background through a hole one centimeter in diameter located at a distance of approximately three meters. The subject gazes at the source for about a minute and a half, or until the light appears to "swim" or "shimmer" and a gray afterimage appears whenever the subject switches his gaze to a white sheet of paper. As an aid in locating the afterimage the subject is asked to close his eyes for a moment. A bright dot or flash of light is invariably reported by subjects at the same position as the afterimage. The stimuli are usually typed letters or drawn black figures on nonreflective white cards or paper.

The projected afterimage varies among individuals but is commonly described as consisting of two types, one following the other. First, a dark gray spot a millimeter or two in diameter is seen superposed on a white or light gray background. Wherever the subject looks the center of the gray dot is taken as the focal point. This gray image lasts for



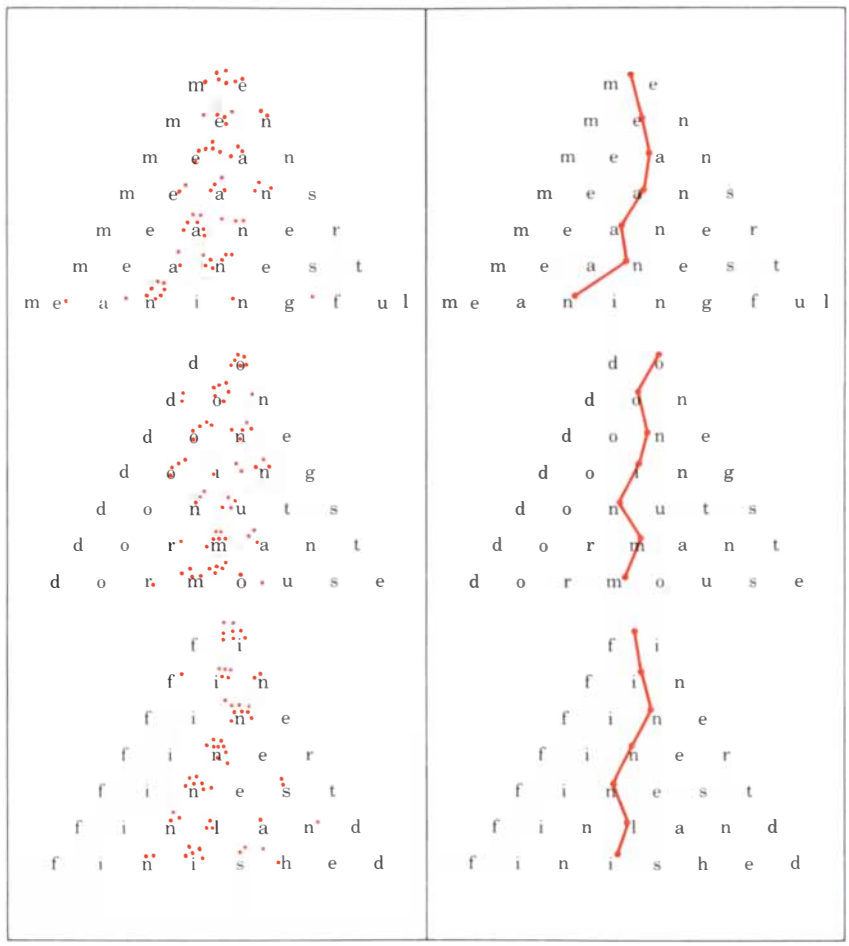




**CLUSTERING of visually similar words results from a different kind of study, in which the test subjects were given a newspaper article containing approximately 100 words and were asked to group together words they thought looked most alike. Such findings, organized into hierarchical tables such as this one, indicate that there are four basic features adults perceive as being important in determining whether words are visually similar: similar beginning letters, similar word length, similar word endings and similar internal letters or letter combinations. Of these four factors similar beginning letters appear to be the most important, a finding that is confirmed by other studies.**

about 30 seconds and is followed by a pink or red circular image of approximately the same size, which persists for as long as three minutes. Several positions of the afterimage are easily remembered by the subject as he looks at the test image.

Preliminary studies of this kind indicate that as words increase in length the subjects' focal point moves progressively to the right of the initial letter. The Dutch psychologist Herman Bouma has shown that subjects can see farther to the right of a fixation point than they can to the left. It seemed reasonable to Bouma that in an English word such as "today" the ideal focal point should be located slightly to the left of center, say over the o, since this would be an efficient place for viewing the entire word. As it turns out, among mature readers this point is indeed the place where the word is most frequently fixated.

On returning from Stockholm to my laboratory at the University of Hawaii I continued my experiments on afterimages. One such study involved three different groups of words. Each group began with similar spellings but varied in length. The words were randomized and were typed in elite type on white paper. After inducing afterimages of the words five graduate students in educational psychology at the university marked their focal points for each word on two separate occasions. The results for all the focal points on these occasions were consolidated for the words that had similar beginnings [*see illustration on this page*]. When the median focal points were connected by straight lines with the words arranged in a triangular fashion, the results showed that the focal point is fairly central. The variation of the position, however, increases markedly with longer words, suggesting a greater choice of stimuli that are effective to look at and also indicating that longer words may cause some subjects to adjust their focal point to the left as a strategy for clearly seeing the initial letter. In almost all cases one focal point suffices for each word. The focal points are also characteristically found on the upper borders of the lowercase letters.



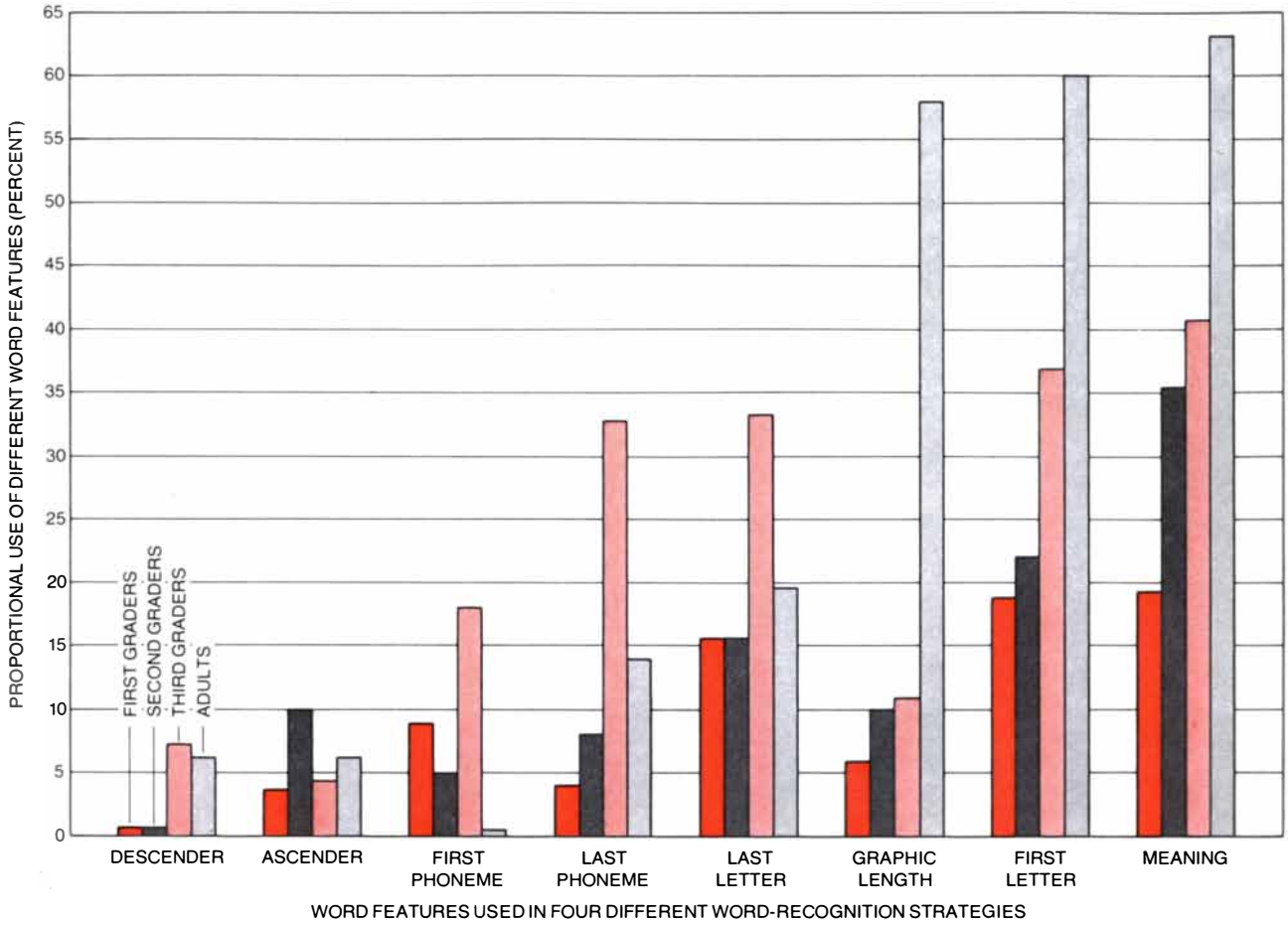

**FOCAL POINTS of test subjects looking at words can be accurately estimated by first generating a strong point afterimage in each subject and then having the subject mark the position of the afterimage when different words are presented as stimuli. In this demonstration of the afterimage technique the colored dots over the words at left represent all the focal points recorded for five different subjects on two separate occasions when the subjects were presented with three sets of words that had similar beginnings but varying lengths. The words have been arbitrarily arranged in a triangular fashion in the illustration and the spacing between the letters has been exaggerated. The colored zigzag lines superposed on the word triangles at right connect the median focal points recorded for each set of words. The results show that the focal points tend to be located fairly centrally in each case, usually on the upper borders of the lowercase letters. The variance in the position of the focal points increases with longer words, although in almost all cases one focal point suffices for each word. There is a slight tendency on the part of some subjects to adjust the focal point to the left with longer words, perhaps as a strategy for clearly seeing the initial letter. Subjects were graduate students in pyschology.**

This study and others indicate that for stimuli such as words and short phrases the afterimage can be an accurate approximation of the focal point. The results show that the focal point does not fall haphazardly on the stimulus and that it is not evenly distributed. This finding suggests a mechanism for subconscious visual preprocessing. It may be that the preprocessing, although it is an innate capacity of human beings, is susceptible to development by frequent exposure to such symbolic stimuli as letters and words. In the mature reader the preprocessing usually results in efficient placement of the focal point. In the studies reported here almost all the focal points were observed on the upper contours of the stimuli centered around (and in a sense attracted by) unique letters, usually internally placed vowels or consonants with ascending letter features. All focal-point studies suggest a peripheral unconscious control of focalization in which the focal points of mature readers are such that vision is efficiently utilized.

It is unwarranted to assume that the recognition of visual features is solely responsible for reading skill. For material in which the context is familiar, glances may be cursory and mostly confirmatory. The following sentence, for example, is printed backward: ".rat eht saw tac ehT"

Read through the sentence once from right to left. Careful reading reveals the nonsensical statement "The cat was the tar." Most subjects, however, read "The cat saw the rat" or "The cat was the rat." Here habit and context are so strong that they misinform. Where words or their context are unfamiliar the processing of smaller units of stimuli becomes more apparent, and the letters of an extremely unfamiliar word may even be studied individually.

To cite another example, read the following sentence and count all the f's: "The first fine fishing day of the year we finally flew to Alaska for five days of




**ANALYSIS of the proportional use made of eight different word features in the preferred word-recognition strategies of children and adults was conducted by Selvin Chin-Chance of the University of Hawaii. The eight objective measures of word similarity, identified at the bottom, included semantic features, phonetic features and meaning along with visual features for a set of 24 carefully chosen words. In addition the eight objective measures of word similarity were correlated with each subject's own subjective estimate of word similarity. The results of this experiment showed that whereas phonetic cues are popular among children, particularly third-graders, they are not used extensively by adults, perhaps because "sounding out the word" is too slow a process for an efficient recognition strategy. Instead mature readers tend to place greater emphasis on meaning, first letters and graphic length in establishing the similarity between word pairs.**

real fishing." Most people count eight f's, the ones that begin the words. They fail to notice that the word "of" also contains an f. Partly because one expects to see f at the beginning of a word, but also because "of" is phonetically pronounced "ov," the specific letter is not noticed.

Something similar happens when one is asked to read the following phrase in this form:


The familiar content and a guessing strategy make one overlook the second "the" in "the spring." In such experiences meaning dominates the visual system. When one is asked to read the following two sentences, "Loveisnowhere," "Theytoldhimtobeatthefrontdoor," one can read either "Love is now here" or "Love is nowhere" in the first case, and either "They told him to be at the front door" or "They told him to beat the front door" in the second. To a large extent one's personal experience dictates which statement is the more probable one. Context, sound and meaning are used as nonvisual cues to word recognition by skillful readers. Along with visual cues they constitute a complex of features employed in an effective reading strategy.

In a recent doctoral dissertation one of my students, Selvin Chin-Chance, analyzed semantic features, phonetic features and meaning along with visual features for a set of 24 carefully chosen words. A separate quantitative measure of similarity between each of the 276 pairs of the 24 words was determined for eight different word features: ascenders, descenders, first phoneme, last phoneme, last letter, graphic length, first letter and meaning. After breaking down the 24 words into two separate lists subjects at various grade levels were asked to rate the similarity between each pair of words. The words were embedded in phrases, and the phrase and then the word were read aloud by the experimenter. At the same time they were also available to the subject in a printed text.

For each word pair a correlation was obtained between the eight objective measures of similarity and the subject's own estimate of word similarity. For each subject a set of 524 correlations over 68 word pairs was available. The questions to be answered were: Which objective measures of similarity most agreed with the subject's estimate? Could a subject's personal strategy for estimating word similarity be determined objectively?

The results show that in general the word-similarity strategies adopted by adults rely primarily on meaning, first letter and graphic length. The develop-

mental processes involved in word perception therefore appear to be similar to those found in my other studies. The most important aspect of these results, however, is that they demonstrate that phonetic cues were not extensively used by the mature reader, perhaps because such processing is too slow; it may be that "sounding out the word" is too time-consuming and unnecessary. Phonetic cues, on the other hand, are frequently used by third-grade children. Although many elementary-school children lack consistent strategies for estimating word similarity children whose responses reveal a consistent strategy also score highest on tests of reading comprehension.

It is reasonable to assume that reading begins with the perception of graphic stimuli. Although some visual material appears to be unnoticed or irrelevant, it may be wrong to assume that familiar material goes unseen or unattended. For example, the Australian psychologist John Ross has shown that the visual system of a mature person can process depth information in less than .0002 second [see "The Resources of Binocular Perception," by John Ross; SCIENTIFIC AMERICAN, March, 1976]. This processing is at least 100 times faster than the average time it takes a person to shift from one visual focal point to another. Processing at this rate is therefore essentially subliminal, or unconscious. It is possible that familiar material is processed at a similar level by some kind of automatic process.

The automatic-preprocessing model of reading states that as the subject's exposure to any given stimulus increases, there is a point at which the stimulus can be preprocessed prior to a conscious awareness. Although some stimuli may be quite complex for the immature subject, they can be at the automatic level for the experienced reader. Thus fewer and fewer focal points are needed as the reading material becomes increasingly familiar. Familiarity can be attained by frequency of meaningful exposure to the actual visual stimulus or by the familiar content and semantic order in which it is embedded; in other words, common phrases could be automatically preprocessed by means of context cues or phonetic structure as well as by visual familiarity. This model indicates that for the young or inexperienced subject most graphic stimuli are complex and must be handled in feature units that are processed through conscious awareness. What is complex for the inexperienced reader may be at the preprocessing level for the more mature reader. Since such processing is learned, concentrating on such useful word features as beginning letter and word length should help in attaining the automatic level of word recognition and reading.

# Some of the Saddest Children in the World are Starting to Smile...

Hungry Children.

Sick Children.

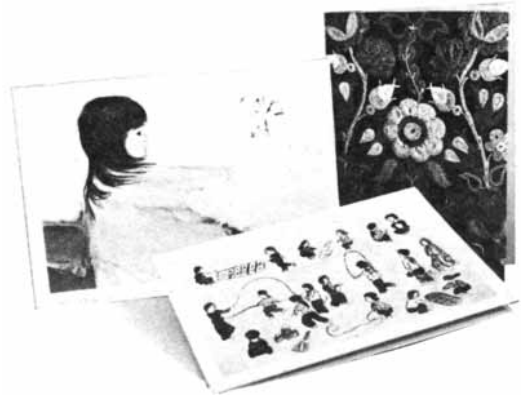
Children with little chance of education.

Millions of children in the poorest countries of Africa, Asia and Latin America.

In over 100 countries, UNICEF, the United Nations Children's Fund, is helping to make these children smile—with more and better food, improved health care and increased educational opportunities.

You can help too, by buying UNICEF Cards, Stationery and Gifts. Proceeds from just one box of cards can provide 185 PAS tablets to help cure children of tuberculosis, a stationery portfolio can mean a new stethoscope for a rural health center, a UNICEF puzzle may supply a box of chalk for a classroom teacher. And since the world of tomorrow will be run by the children of today, supporting UNICEF's efforts to give them a better start in life can help to insure a more peaceful world for future generations.

The truly international collection of UNICEF Cards, Notes and Stationery are perfect for birthdays, anniversaries, congratulations, thank you, or for just keeping in touch. UNICEF gifts—puzzles, games, story books, coloring books—are an educational treat for children any time of the year.



Write for a free color brochure:
U.S. Committee for UNICEF, Department GC
331 East 38th Street, New York, N.Y. 10016

**UNICEF**

*Cards • Stationery • Gifts*

# THE AMATEUR SCIENTIST

*More about polarizers and how to use them, particularly for studying polarized sky light*

by Jearl Walker

One of the natural sources of polarized light is the sky. As Lord Rayleigh wrote in 1871, polarization results from the scattering of sunlight by the molecules of the air (the same phenomenon that is responsible for the blueness of the sky). On a clear day much of the light from the sky is linearly polarized, a phenomenon that bees and some other animals exploit in navigation. There is reason to believe the ancient Vikings also navigated across open water by means of polarized sky light.

Imagine a light ray from the sun that encounters a molecule of one of the gases making up the air. Consider the unpolarized light as having two senses of polarization, one lying in the plane that includes the sun, the gas molecule and the observer on the ground and one lying perpendicular to that plane. The two senses of polarization (respectively termed parallel and perpendicular) represent the directions along which the electric fields of the light oscillate.

Both senses of oscillation cause electrons in the molecule to oscillate and reradiate light. The oscillation of an electric field along a particular axis, however, cannot cause an electron to radiate along the same axis. The electron radiates most strongly perpendicular to the axis and less strongly in directions lying progressively closer to the axis along which the field oscillates.

What does this pattern mean for the polarization of the sunlight scattered from an air molecule to an observer? Consider an observer directly below a molecule illuminated by a setting sun. The perpendicular sense of polarization in the incident light is scattered strongly to the observer but the parallel sense is not scattered at all. The net result for the observer is that the incident unpolarized light of the sun is polarized by the scattering from the molecules.

You can map the polarization of various points in the sky with a polarizing filter. First orient the filter by examining the light reflected from a street or certain other flat surfaces. (I described the procedure last month.) When the glare from an appropriate surface is blocked by the filter, the sense of polarization being passed by the filter is perpendicular to the surface.

Now examine part of the sky. Depending on the position of the sun, the sky is polarized mainly perpendicular to the plane containing the sun, that portion of the sky and you. The degree of polarization is largest directly overhead when the sun is near the horizon; the polarization drops off as you scan either toward or away from the sun or toward the horizon on either side of the sun. (Never look at the sun, even with what you might think is a dark filter. The filter may be dark in the visible range of light but may transmit dangerous amounts of radiation in the infrared or the ultraviolet, both of which can damage your eyes before you are even aware of the destruction.)

The sky near the horizon at a right angle to a low sun (say the northern horizon for a setting sun) should be polarized as strongly as the sky directly overhead. The light from that horizon has to traverse more of the earth's atmosphere, however, so that it is likely to be scattered more than once or to be scattered from airborne particulate matter or aerosols. All these processes decrease the polarization.

Clouds also destroy the polarization because of multiple scattering of the light going through them, an effect often exploited by cloud photographers. With a polarizing filter oriented properly in front of the lens of a camera the polarized background sky light is diminished but the light from the cloud is diminished less, so that the cloud contrasts better with the background sky.

A more convenient device for determining the polarization of the sky can be made from several small polarizing filters or one large polarizing sheet. In either case cut 12 isosceles triangles in such a way that the apex is 30 degrees and the sense of polarization is along the base. Lay the triangles in a dodecagon on a clear piece of glass. Tape or glue the edges to the glass. Do not fasten down entire area of the filters, because the adhesive may alter the polarization of the light transmitted through the device.

When you look at the sky through the dodecagon, you will see light and dark triangles corresponding to the sky's polarization. You no longer have to rotate a single piece of polarizing filter. To make the polarization even more apparent, or at least more colorful, put one or two layers of clear cellophane tape on the triangles before you cut them out. Orient the tape at 45 degrees to the sense of polarization of the filters so that each triangle can then act as a half-wave plate of the type I described last month. The tape must be on the side of the device that will be toward the sky. The linearly polarized light from the sky enters the tape and is shifted in phase between the tape's fast and slow axes. It emerges with a different polarization depending on its color.

The filter to which the tape is attached determines what color is seen. Through this array of half-wave plates the sky appears in different colors in the different triangles. With a single layer of cellophane tape on the filters I find a variation in colors around the device from clear (thus showing the sky color itself) to deep blue. With two layers the variation is from yellow-green to blue.

Today polarizing filters can be employed to determine the polarization of the sky for the purpose of navigation. The first such polarizer, however, was developed (by Edwin H. Land) only in 1928. Before that time the polarization of the sky could be checked only by observing Haidinger's brushes (the curious pattern some people see when they view polarized light) or by means of the birefringent or dichroic crystals I described last month.

The device that may have enabled the Vikings to navigate across the Atlantic with the aid of linearly polarized sky light may well have been a dichroic crystal. Not having a compass, they would have been lost when the sun was below the horizon, as it often was in the high latitudes of their homeland. Several types of crystal were available to serve as "sun stones," but the Danish archaeologist Thorkild Ramskou believes a pure single crystal of cordierite was the most likely one.

Ramskou acted on the suggestion of a 10-year-old boy whose father was chief navigator for the Scandinavian Airlines system. The boy noted a connection between old stories of sun stones and the instrument called a twilight compass by which his father locates the sun even when it is blocked by clouds or is somewhat below the horizon. With a polariz-

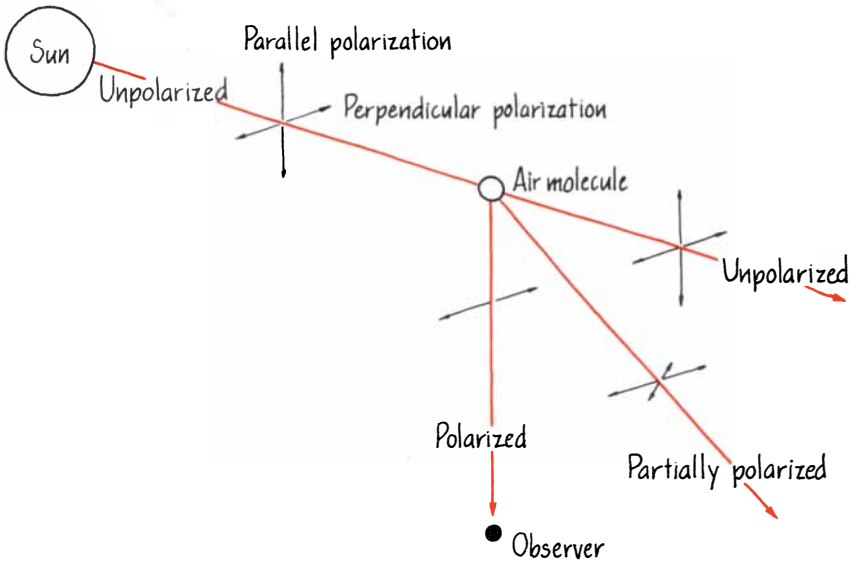
ing filter a modern navigator detects the sense of polarization of a clear part of the sky; the instrument then points in a direction perpendicular to that sense and hence toward and directly away from the sun. Knowing roughly which direction is which, one can use the twilight compass to locate the position of the sun fairly accurately.

Cordierite is dichroic when its fast and slow axes are perpendicular to the incident light from the sky. Rotating the crystal while keeping it perpendicular to the direction of the light causes a series of different axes to lie along the sense of polarization of the incident light. With the fast axis aligned the crystal is either clear or slightly yellow when you look through it. With the slow axis aligned it is blue. Once a Viking navigator understood how his crystal had to be oriented to give blue light, he could find the sun when it was below the horizon or hidden behind clouds by again finding the blue orientation in some part of the clear sky. (If the entire sky was cloudy, none of the light was polarized and the Viking was out of luck.)

Ramskou flew along with the Scandinavian Airlines navigator to compare locations of the sun using the cordierite and the twilight compass. He found that his tracking of the sun with the sun stone was good to within about three degrees. He could still find the sun when it was as much as seven degrees below the horizon. If you have access to a single crystal of pure cordierite or some other dichroic crystal such as tourmaline, you might try a similar experiment.

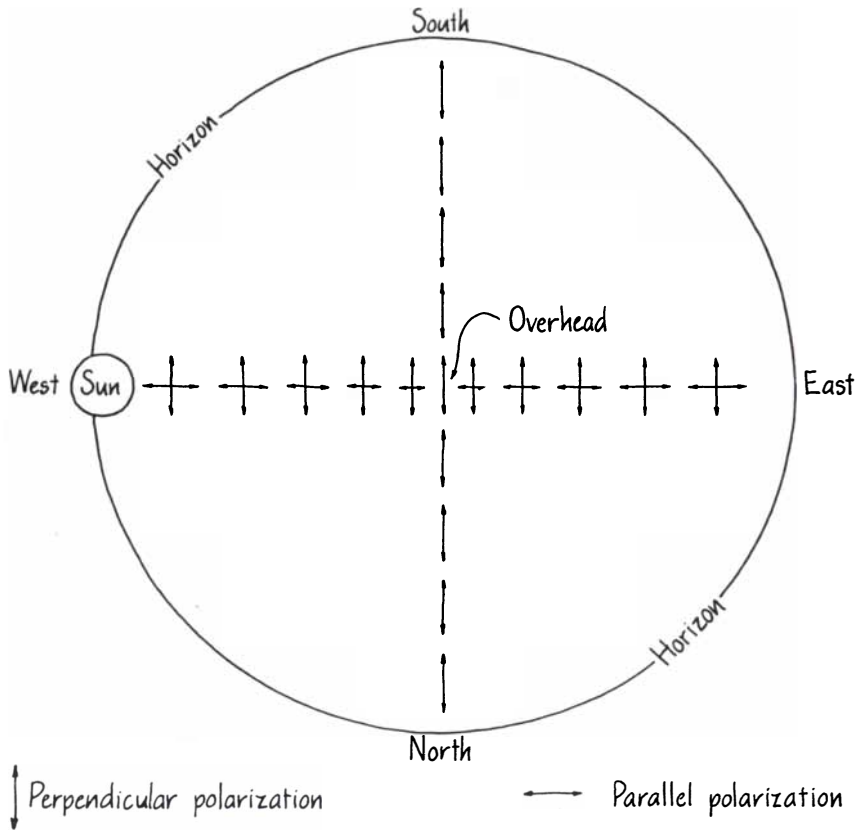
Much of what is known about the polarization of sky light is based on work by Lord Rayleigh. His model for the polarization of the sky breaks down in two interesting cases. First, if the air is significantly polluted, the multiple scattering of the light and the scattering by particles nearly the size of (or larger than) the wavelength of visible light destroy the sense of polarization in the Rayleigh model. You will see this destruction of polarization on days of high air pollution from either man-made sources or natural ones such as volcanic eruptions.

The other departure from the Rayleigh model is more curious. If you carefully examine the sky near the sun or near the antisolar point (the exact opposite point in space), you will find several points that are unpolarized and also a small region with exactly the wrong sense of polarization, known as negative polarization. The points are called neutral points and are named after their various discoverers. The cause of the neutral points and of the region of negative polarization is still somewhat controversial, but it appears to be due to both multiple scattering of sunlight and scattering by aerosols, primarily in the lower atmosphere.




*The polarization of sunlight when it is scattered by gas molecules of the air*

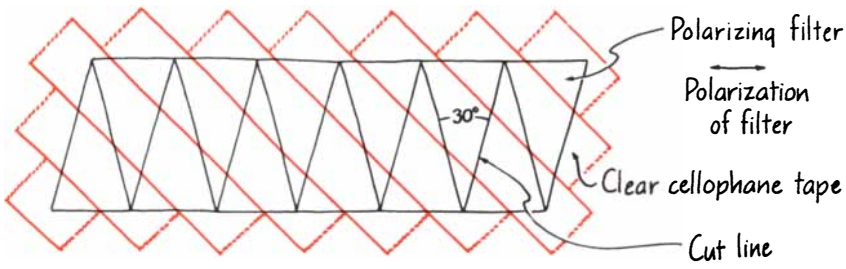
The neutral point that divides the area of negative polarization from the area of positive (normal) polarization is known as Arago's point (after D. F. J. Arago, the early-19th-century French astronomer and physicist). I can locate Arago's point by waiting until the sun is near the horizon and then finding the region of negative polarization near the antisolar point. The position of Arago's point (and thus the angular extent of the region of negative polarization) depends on the wavelength of light you consider, the elevation of the sun and the extent of the reflection of light from the earth's surface. For example, if you place a

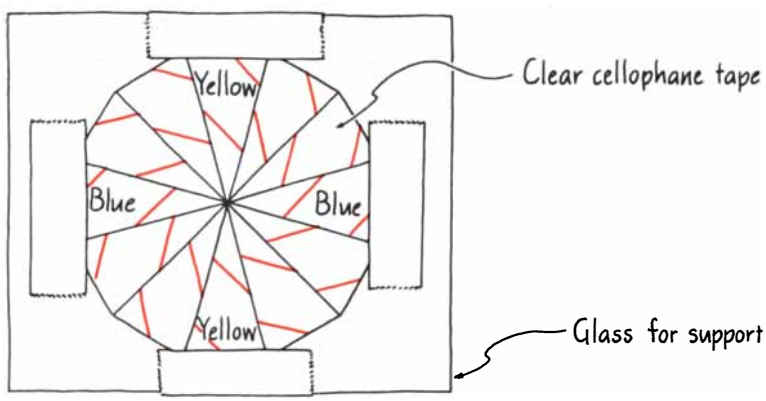



*A sky map showing the polarization of light according to ideal Rayleigh scattering*

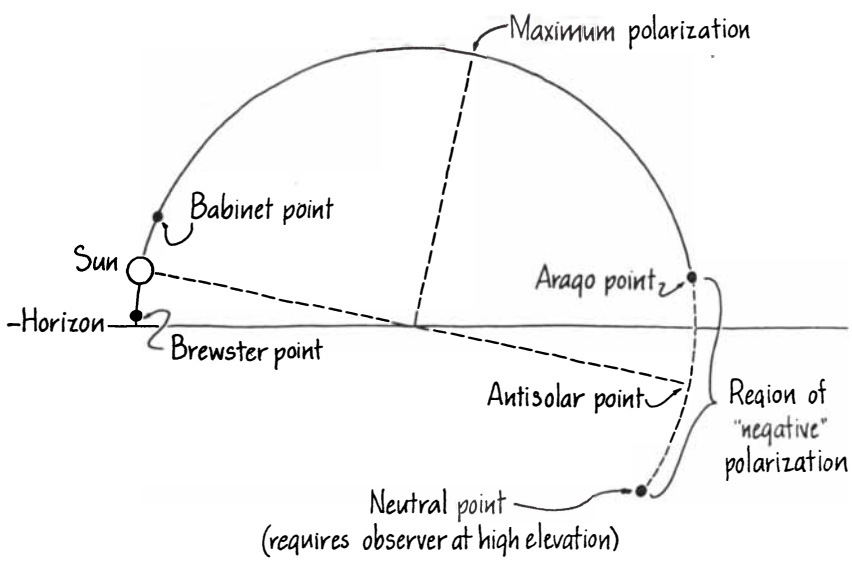
blue or a red filter over your polarizing filter (you could use blue or red cellophane), you will find that Arago's point is higher in the sky in the blue than in the red. The position of the point also depends on the sun's elevation; as the sun sinks, less of the lower atmosphere is illuminated.

If you start with Arago's point by looking at the antisolar point as the sun is setting at your back, Arago's point at first moves slightly away from the antisolar point and then moves back toward it as the sun reaches the horizon. At sunset the lower atmosphere, where the multiple scattering and aerosol scattering responsible for the negative polarization occur, is no longer well illuminated, whereas the upper atmosphere still is. Hence the region of negative polarization is diminished and the positive polarization is increased, moving Arago's point closest to the antisolar point. With the sun below the horizon Arago's point again moves away from the antisolar point. You might find the total movement of the point with respect to the antisolar point to be approximately five degrees. The third effect on the position of Arago's point is attributable to the light reflected from the earth's surface. With greater reflection, such as would be the case over a large body of water, the point descends by a few degrees.




*Making a dodecagon of half-wave plates for examining the sky's polarization*




*Neutral points and "negative" polarization*

Whereas light scattered from molecules can be strongly polarized and light scattered from large particles can be unpolarized or only weakly polarized, the intermediate-size particles of haze result in an intermediate degree of polarization. This intermediate polarization has interesting effects on haze-obscured distant objects. For example, the San Gabriel Mountains in California are often obscured by the smog of Los Angeles. If you are wearing polarized sunglasses, however, or are using some other polarizing filter, you can make the mountains come in and out of visibility by tilting your head or the filter you are using. Warren H. White of Meteorology Research, Inc., explains the difference in visibility by the degree of polarization by the smog and so is able to determine the approximate size of the smog particles. For one particular orientation of the filter the mountaintops are distinguishable against the background sky, but the features on the mountain slopes are not visible. Rotating the filter by 90 degrees to change the sense of the transmitted polarization makes the mountaintops disappear and the surface features become apparent.

If the smog particles were very small compared with the wavelength of visible light, say with diameters of less than about .05 micron, they would polarize light just as individual molecules do in the Rayleigh model. If the particles were much larger than the wavelengths of visible light, with diameters of more than about five microns, the scattered light would be weakly polarized. Consider sunlight incident on the mountains at an angle of 45 degrees from the vertical. In the first orientation of the filter the mountaintops are distinguishable because the background light of the distant clear sky is more polarized than the smog-scattered light between the peaks and the observer. The light from the sky is therefore brighter than the light coming from the peaks, and the contrast is noticeable.

When the filter is rotated by 90 degrees, the light from the sky and from the peaks is blocked but some of the features on the slopes are made visible. They are the features that are appropriately oriented to polarize the reflected sunlight in the same way that a road surface does. Although this light is only weakly polarized, it is distinguishable

because the light from the rest of the mountain is polarized by the smog in such a way that it is blocked by the new orientation of the filter. Because the smog polarizes the light more strongly than the surface features but not as strongly as the more distant clean air the diameters of the smog particles must lie in the range from .05 micron to five microns, a conclusion consistent with more sophisticated research that finds the range is from .1 to one micron.

I have seen a similar effect when watching ships disappear into the Cleveland smog as they move out onto Lake Erie. If you live in an environment with a comparable level of air pollution, you might want to repeat White's experiment and also to correlate the aerosol size with the season. The same opportunity should exist with the blue hazes that cover some mountain ranges, such as the Blue Ridge Mountains in West Virginia and the Blue Mountains in Australia. These hazes are believed to result from the release by vegetation of small particles of wax as a result of the relatively high electrostatic fields near the tips of the foliage. The particles are less than .6 micron in diameter and therefore should contribute some polarization as they scatter sunlight.

The partial polarization in the light scattered by an aerosol can be simulated in the kitchen by a simple experiment. Block off most of the face of a flashlight (or some similar light source) so that a narrow rectangle of light is emitted. Set up the flashlight to shine through a clear glass of water, turn off the room lights and examine the glass for light scattered perpendicular to the beam. (A restricted beam is better than the full beam because it makes it possible to avoid illuminating the walls of the glass more than necessary.)

If the water is relatively clean, it will be too transparent to scatter light perpendicularly and the glass will look dark. Each water molecule is scattering light in all directions into the room, but the light scattered by one molecule interferes destructively with the light scattered by another in all directions except the strictly forward one. The beam is therefore reconstructed in that direction and no net scattering of light occurs in other directions.

Now add a few drops of milk. Even before you check the perpendicular scattering notice the colors of the milk. Blue is scattered from the forward direction more strongly than red, and the colors between blue and red in the spectrum are scattered at intermediate angles. If you look at the milk from the side from which the light is shining, the milk appears bluish. If you look from the other side (into the beam), the milk appears reddish.

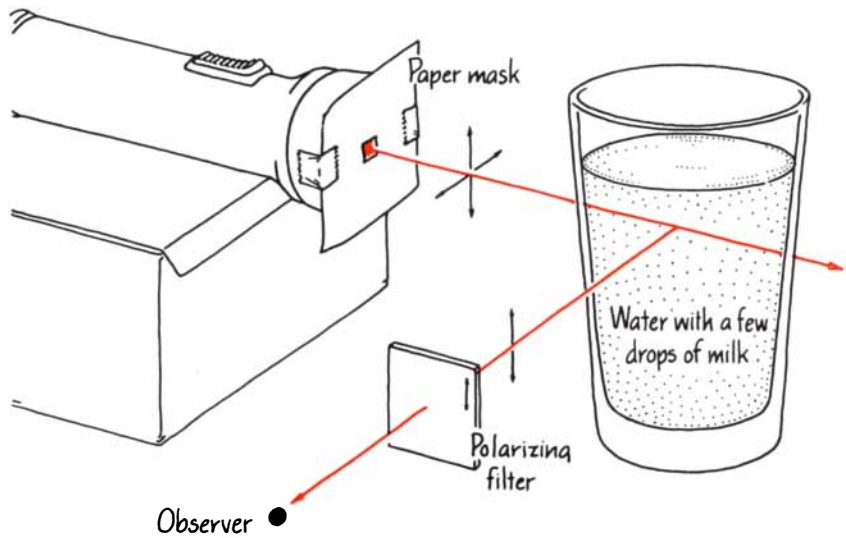
This distribution of colors is similar to the general appearance of the sky during the day, although the scattering is not exactly the same. When light is scattered by air molecules according to the Rayleigh model, blue light is scattered more strongly than red. The result is that the entire sky is blue for most of the day. When the sun is setting or rising, the blue is scattered out of the light from near the horizon, leaving mostly red, which is why a setting or a rising sun is red. The light scattered by the milk is not scattered by individual molecules because the milk is not dissolved in the water. Rather the milk forms a colloidal suspension of tiny globules.

This suspension scatters light at angles to the direction of the flashlight beam and thereby polarizes the scattered light. With a polarizing filter check the polarization of the light scattered perpendicular to the beam. If the beam is horizontal, the filter will transmit scattered light with its sense of polarization vertical. As you add more drops of milk the degree of polarization in the scattered light decreases.

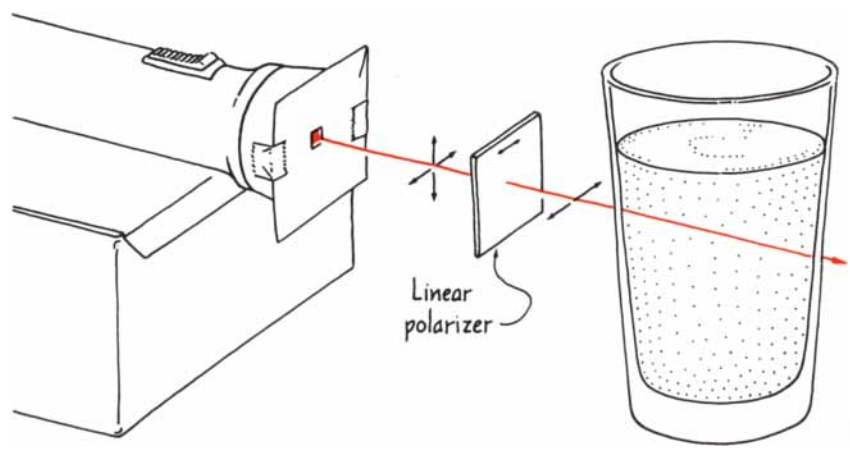
When the milky water is noticeably white and the red and blue disappear, the scattered light is no longer polarized. By then the density of the suspended globules of milk is so high that most of the scattered light has been multiply scattered by the time it reaches the eye. Just as with atmospheric scattering, the multiple scattering destroys the polarization.

I could find no difference in the results from skim milk and whole milk in spite of the difference in fat content. You might want to try other colloidal suspensions. You might also want to try sugar to see the difference between a solution (in which the light is scattered by individual molecules) and a colloidal suspension.

With a few drops of milk in the water you can also demonstrate a result I discussed above: Light is not scattered along the axis of polarization of the inci-




*Scattering effect of a colloidal suspension of milk in water*




*Effect of making the incident light horizontally polarized*

dent light. Attach a polarizing filter to your flashlight and with another filter again examine the light scattered perpendicular to the beam. If the flashlight beam is polarized vertically, the eye is receiving polarized scattered light. If the beam is polarized horizontally, the eye is receiving no scattered light because it lies on the polarization axis of the incident light.

Last month I discussed how circularly polarized light can be produced. Linearly polarized light is passed through a quarter-wave plate arranged so that the axis of the polarization lies midway between the fast and slow axes of the plate. At the wavelength for which the plate is precisely a quarter-wave one circularly polarized light emerges. The sense of rotation of the emerging polarization vector depends on the orientation of the incident polarization and the two plate axes. By convention the rotation is termed right circular if the rotation is clockwise as you would view it while you were looking toward the source of the light and left circular if the rotation is counterclockwise.

If a particular circular polarizer produces, say, right-circular light, the light can also travel through the device in the opposite direction, although the light finally emitted is linearly polarized by the linear filter. If left-circular polarized light is first sent through a right-circular polarizer from the quarter-wave-plate side, the light is simply absorbed. This effect has been exploited to reduce the glare on television screens, oscilloscopes, airplane windows and similar places where glare is a problem.

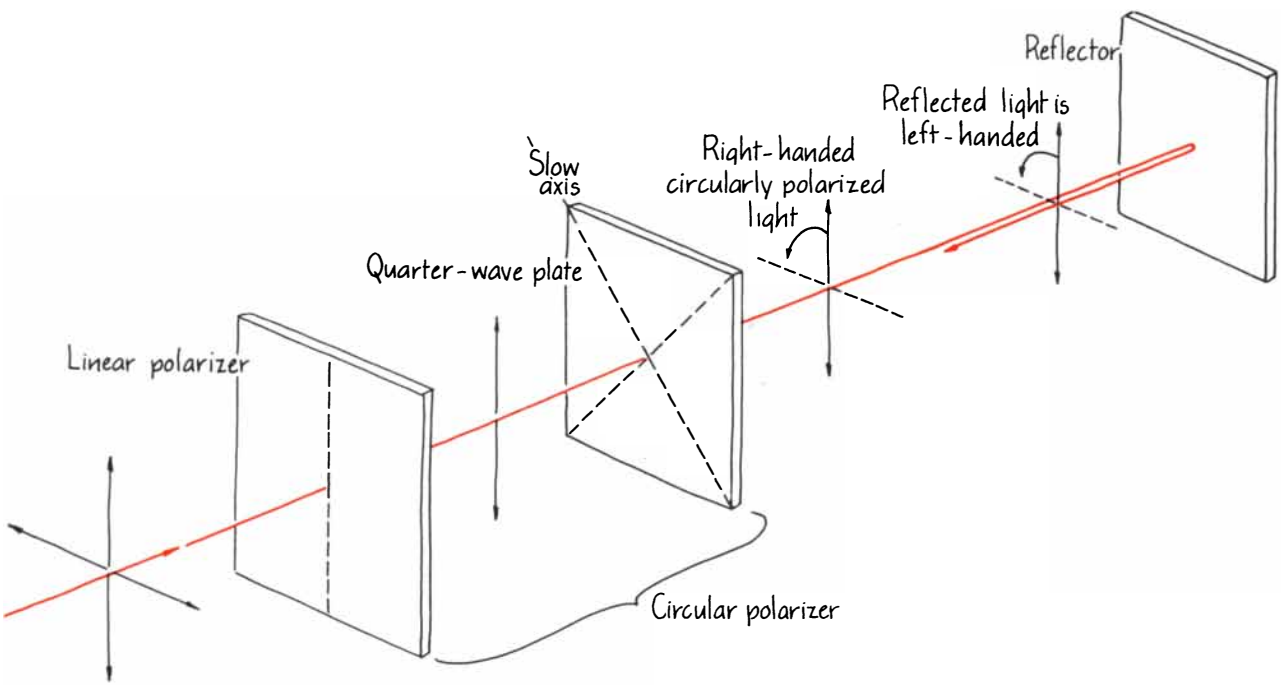
For example, a commercially produced circular polarizing layer can be put in front of your television screen. The layer is a package of a linearly polarizing filter glued at 45 degrees to the fast and slow axes of a quarter-wave plate, with the linear filter on the outside. Room light passing through the sheet becomes polarized in, say, the right-circular sense. When the light is then reflected from the television screen, the reflection converts the light's sense of polarization into left-circular, but the sheet absorbs that sense. Hence the room light that would otherwise have been reflected from the screen to interfere with viewing is now absorbed by the circularly polarizing sheet.

To demonstrate this kind of absorption construct a circular polarizer from the quarter-wave plate I described last month. Lay the polarizer on a dime or some other shiny object. If the linear polarizer that is part of the circular polarizer is on the bottom of the polarizer, the dime is shiny when you look down through the device. If the linear polarizer is on the top, the dime is dark because of the absorption of the reflected light. Lift the polarizer away from the dime. As light reaches the dime without having to go through the polarizer, the dime once again becomes shiny.

Replace the dime with a smooth piece of aluminum foil and repeat the demonstration. You will get the same results. Now press a *V*-shaped crease into the foil and lay the circular polarizer over the *V* with its linear filter upward. The smooth part of the foil is again dark, but the crease is shiny because it is acting as a double mirror. Suppose right-circular polarized light is produced by your circular polarizer. The component of the light that strikes one side of the crease is reflected to the other side before reemerging, its polarization thus being switched from right-circular to left-circular and then back to right-circular before it reaches the polarizer again. Its return as right-circular polarized light means that it passes through the circular polarizer without absorption, thereby giving you a shiny *V*.

You can play further with your circular polarizer by making more of the quarter- and half-wave plates. If left-circular polarized light is passed through a half-wave plate, you will see right-circular polarized light. If instead it is passed through a quarter-wave plate, you will see linearly polarized light. You might repeat the absorption demonstration involving a reflector by inserting a half-wave plate between the circular polarizer and the reflector. Does the circular polarizer still absorb the reflected light?

A sample circular polarizer, along with a short explanation of its physics, is available to readers of this department without charge from the Polaroid Corporation (Polarizer Technical Products, 20 Ames Street, Cambridge, Mass. 02139). Ask for the card "Polaroid Circular Polarizers for Contrast Improvement."




*Absorption of returning light by a circular polarizer*

# This is Coby 1.™

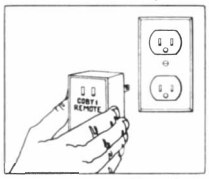
## A brand new electronic home control center that will change your way of life.

**You can *automate* your home with the Coby 1 System.**

- **Coby 1 Control plugs in anywhere to give you computerized ON/OFF control over *electrical devices in your home.***
- **It's the most sophisticated timer you can buy. It turns things on or off at precise times, at the preset intervals you select, and can be programmed *up to 11 months in advance.***
- **It's an instant control center for electrical devices right from your bedside—or wherever you and Coby 1 Control happen to be.**
- **Its elegant digital clock gives you the month, day, hour, minute, and second, with accuracy to within *five seconds per month.***

### COBY 1 CONTROL NEEDS NO WIRING —PLUGS IN ANYWHERE.

The compact control panel, which we call Coby 1 Control, plugs into any ordinary wall outlet, sending computer-coded pulses through your present wiring. The pulses trigger **Coby 1 Remotes**—small remote switches to go between plug-in appliances and wall sockets. Soon (by March) we will also have Coby 1 Remotes to replace wall switches and Coby 1 Remotes to take care of built-in appliances like water heaters and air conditioners. It's safe, simple, and sure. There's nothing like it.



### A REMARKABLE APPLICATION BREAKTHROUGH IN MICROELECTRONICS AND PULSE-CODE COMMUNICATION.

The Coby 1 System™ is the result of brilliant engineering by a team of aerospace electronics people. Coby 1 Control includes an Intel 8085 Microprocessor—**a complete tiny computer**—plus control circuitry, power supply electronics, coding and signal-generating circuits, an emergency power cell, and memory. The memory contains 2048 words of low-power, programmable random access memory (RAM) and 2048 bytes of read-only memory (ROM). It stores device numbers, commands, and status information for **up to one hundred Coby 1 Remotes.**

The coding and signal-generating hardware translates commands and distributes them through your home wiring (but without interfering with any of your other appliances) to the Coby 1 Remotes, which decode the signals and turn things ON or OFF.

As you enter commands through the keyboard, the display lights up to confirm. It can also be used to review commands stored in memory. All programs are entered, stored, and modified through 12 function and control keys and a 10-key numeric pad (plus AM and PM keys).

### ITS ACCURACY IS BLACKOUT-PROOF.

Power blackout? Built-in battery power keeps Coby 1's memory fresh. Unlike a conventional timer or clock radio, Coby 1's clock won't lose a second. When power comes back on, the batteries automatically recharge. This feature also lets you **unplug your control unit and plug it in again anywhere.** Its handsome digital clock gives you the year, month, day, hour, minute and second with accuracy to five seconds per month. The calendar will show the correct date until 2021.

We've protected Coby 1's sophisticated brain with a handsome, precision-aluminum package that is spillproof. Since Coby 1 has no moving parts, it requires no maintenance.

### IT'S FUN TO USE COBY 1.

Each Coby 1 Remote is assigned an identification number. If the front hall lamp is Number One, you simply tell Coby 1 Control to turn Number One on or off—now, Tuesday, any day or every day, if you like. And if you've ever operated a pocket calculator, you'll have no problem whatsoever with Coby 1.

### COBY 1 IS AMONG THE GREATEST LUXURIES YOU CAN OWN.

Picture yourself on a frigid winter morning. Coby 1 can wake you with your hi-fi system and a lamp.

You rise to a warm bathroom. When you come out, the coffee is ready to pour. Coby 1 turns the hi-fi off and the TV on, while you enjoy your coffee and paper.

Coby 1 has warmed the car engine for you, so it starts readily and warms up quickly.

At bedtime, with Coby Control now plugged in at bedside, you turn off all the lights and switch off the TV—**without getting up.** You go to sleep knowing things will be ready for you again in the morning. If, during the night, you want to turn on the outside or living room light, you have the comfort and security of being able to do so from your bedside table.

### IT'S A CARETAKER WHEN YOU'RE GONE.


Now you can go away for a week (or a month) and leave Coby 1 in charge. Consider a potential thief watching your house: lights (Nos. 1–6) go on and off as if people were moving around. The TV (No. 7) goes on; then goes off. Finally, the bathroom (No. 8) and bedroom lights (No. 9) go out. **You can repeat the pattern daily or vary it for up to a year** in cycles as short as a second or as long as 100 hours. Yet **it will use less energy** and suggest more activity than leaving lights or a radio on constantly.

When you come home, Coby 1 can have the house warm (or cool, in summer), the porch light on, the sofa lamp on, and the hi-fi on to welcome you.

### CONSIDER THE ENERGY SHORTAGE.

Coby 1 can do wonders for your electrical bill. It never forgets to turn things off. It can turn car heaters, air conditioners, or electric heaters on just far enough ahead to make things comfortable—no need to have them on constantly. You can change the times from your easy chair. No mechanical timers; no wasted power.

### WE'RE INTRODUCING THE COBY 1 SYSTEM AT A SPECIAL LOW PRICE.

We're anxious to get the first factory run into the hands of users as quickly as we can because we're interested in how you put Coby 1 to work. **So until February 15, 1978, we'll accept advance orders for a Coby 1 Control at $399.00, the price to include a free Coby 1 Remote.** Other remotes will be extra.

Simply fill out the coupon and send it with your check, money order, or credit card data. You can also call in your order or get more information by calling (505) 526-3358. We'll ship your Coby 1 along with full instructions and suggestions on its use, after our first production run in January. We'll also include our 90-day parts-and-labor limited warranty.

### WHO ENERGY TECHNOLOGY IS:

The company was started by the three of us, Brook Reece, Phil Reed, and Keith Burn. We developed the system ourselves. We're excited about Coby 1 because everyone we've talked to has expressed real interest in the product and sees a need for it.

We've been working on Coby 1 for months. Development and testing of production models is now complete. They'll be ready to ship in January.

**Dealer inquiries invited. Energy Technology, Incorporated, 1601 South Main St., P.O. Box Q, Las Cruces, NM 88001. Phone: (505) 526-3358**

Mail to:

## Energy Technology, Inc.

1601 South Main St., P.O. Box Q, Las Cruces, NM 88001

Put me down for one of the first Coby 1's. I understand that this is an advance order, and that shipment is expected after January 15, 1978.

**Ship me one Coby 1 Control and one 10-amp plug-in Coby 1 Remote at the *Special Introductory Price* of $399.00, shipping included.** $399.00

In addition, I want to order the following:

___ more 10-amp plug-in Coby 1 Remotes @ $39.95 _____

___ 25-amp plug-in Coby 1 Remotes @ $49.95 _____

Total _____

New Mexico residents: please add Gross Receipt tax _____

☐ Check ☐ BA/V ☐ MC ☐ Money Order Total _____

Card No. _____

Interbank No. _____ Card Exp. date _____

Ship to _____

Address _____

City _____ State _____ Zip _____

SA 1/78

Signature _____

# BIBLIOGRAPHY

*Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.*

## MATHEMATICAL GAMES

100 PUZZLES: HOW TO MAKE AND SOLVE THEM. Anthony S. Filipiak. A. S. Barnes and Co., Inc., 1942.

LA SCULPTURE DE BERROCAL. Giuseppe Marchiori. La Connaissance Bruxelles, 1973.

## THE CARBON DIOXIDE QUESTION

THE CARBON CYCLE. Bert Bolin in *Scientific American,* Vol. 223, No. 3, pages 124–132; September, 1970.

BROOKHAVEN SYMPOSIA IN BIOLOGY NO. 24: CARBON AND THE BIOSPHERE. Edited by George M. Woodwell and E. V. Pecan. U.S. Atomic Energy Commission, Division of Technical Information, Oak Ridge, Tenn., 1973.

PRIMARY PRODUCTION OF INLAND AQUATIC ECOSYSTEMS. Gene E. Likens in *Primary Productivity of the Biosphere,* edited by H. Lieth and Robert H. Whittaker. Springer-Verlag, 1975.

## THE SURGICAL REPLACEMENT OF THE HUMAN KNEE JOINT

ARTHROPLASTY OF THE KNEE USING AN ENDOPROSTHESIS: EIGHT YEARS' EXPERIENCE. Börje Walldius in *Acta Orthopaedica Scandinavica,* Vol. 30, No. 2, pages 137–148; 1960.

POLYCENTRIC KNEE ARTHROPLASTY: PROSTHETIC SIMULATION OF NORMAL KNEE MOVEMENT. Frank H. Gunston in *The Journal of Bone and Joint Surgery,* Vol. 53-B, pages 272–277; May, 1971.

GEOMETRIC TOTAL KNEE ARTHROPLASTY, I: CONCEPTION, DESIGN, INDICATIONS, AND SURGICAL TECHNIC. Mark B. Coventry, Jackson E. Upshaw, Lee H. Riley, Gerald A. M. Finerman and Roderick H. Turner in *Clinical Orthopaedics and Related Research,* No. 94, pages 171–176; July–August, 1973.

THE SPHEROCENTRIC KNEE: BIOMECHANICAL TESTING AND CLINICAL TRIAL. David A. Sonstegard, Herbert Kaufer and Larry S. Matthews in *The Journal of Bone and Joint Surgery,* Vol. 59-A, pages 602–616; July, 1977.

## THE THREE-DIMENSIONAL STRUCTURE OF TRANSFER RNA

THREE-DIMENSIONAL STRUCTURE OF YEAST PHENYLALANINE TRANSFER RNA: FOLDING OF THE POLYNUCLEOTIDE CHAIN. Sung Hou Kim, Gary J. Quigley, Fred L. Suddath, Alexander McPherson, Daryll Sneden, Jung-Ja Park Kim, Jon Weinzierl and Alexander Rich in *Science,* Vol. 179, No. 4070, pages 285–288; January 19, 1973.

THREE-DIMENSIONAL STRUCTURE OF TRANSFER RNA. Sung Hou Kim in *Progress in Nucleic Acid Research and Molecular Biology,* Vol. 17, pages 181–216; 1976.

TRANSFER RNA: MOLECULAR STRUCTURE, SEQUENCE, AND PROPERTIES. Alexander Rich and Uttam L. RajBhandary in *Annual Review of Biochemistry,* Vol. 45, pages 805–860; 1976.

## THE STRUCTURE OF THE INTERSTELLAR MEDIUM

A MODERN LOOK AT "INTERSTELLAR CLOUDS." Carl Heiles in *International Astronomical Union Symposium No. 60: Galactic Radio Astronomy,* edited by F. J. Kerr and S. C. Simonson III. D. Reidel Publishing Company, 1974.

THE EVOLUTION OF SUPERNOVA REMNANTS, I: SPHERICALLY SYMMETRIC MODELS. Roger A. Chevalier in *The Astrophysical Journal,* Vol. 188, No. 3, Part 1, pages 501–516; March 15, 1974.

SOFT X-RAY EMISSION IN ERIDANUS: AN OLD SUPERNOVA REMNANT? S. Naranon, S. Shulman, H. Friedman and G. Fritz in *The Astrophysical Journal,* Vol. 208, No. 3, Part 1, pages 718–726; September 15, 1976.

## HOW BACTERIA STICK

STRUCTURE AND FUNCTION OF THE CELL ENVELOPE OF GRAM-NEGATIVE BACTERIA. J. W. Costerton, J. M. Ingram and K.-J. Cheng in *Bacteriological Reviews,* Vol. 38, No. 1, pages 87–110; March, 1974.

DENTAL CARIES. R. J. Gibbons and J. van Houte in *Annual Review of Medicine: Selected Topics in the Clinical Sciences,* Vol. 26, pages 121–136; 1975.

FINE STRUCTURE AND DISTRIBUTION OF EXTRACELLULAR POLYMER SURROUNDING SELECTED AEROBIC BACTERIA. Gerald D. Cagle in *Canadian Journal of Microbiology,* Vol. 21, No. 3, pages 395–408; March, 1975

## THE EFFICIENCY OF ALGORITHMS

ON COMPUTABLE NUMBERS, WITH AN APPLICATION TO THE ENTSCHEIDUNGSPROBLEM. Alan M. Turing in *Pro-*

# Introducing something new for kids who don't like to read.

That's right. You can help. And we can help you do it.

We're Reading is Fundamental, a non-profit, reading motivation program with hundreds of local programs in communities like yours all across the country.

The idea is simple. RIF programs get books to kids. Books they can choose for themselves. And keep for their own. Books that'll get them to start to read, to like to read, to want to keep on reading.

Reading is thinking. And learning. And growing. That's why we want to start more programs, distribute more books, get more kids into reading.

That's why we need you—and people like you—or your organization—to start a RIF group in your community. We'll help you get started—or put you in touch with a RIF program in your area. Because we know RIF works. And you will, too.

Write today for more information. Write to RIF, Box 23444, Washington, D.C. 20024.

And thanks. From a lot of kids you don't know yet.

READING IS FUNDAMENTAL

When it comes to getting kids to read, RIF wrote the book.

## You.

This is a public service message on behalf of Reading Is Fundamental and this magazine.

*ceedings of the London Mathematical Society,* Series 2, Vol. 42, Parts 3 and 4, pages 230–265; November/December, 1936.

THE COMPLEXITY OF THEOREM-PROVING PROCEDURES. Stephen A. Cook in *Proceedings of the Third Annual AMC Symposium on Theory of Computing,* pages 151–178; 1971.

REDUCIBILITY AMONG COMBINATORIAL PROBLEMS. Richard M. Karp in *Complexity of Computer Computations,* edited by Raymond E. Miller and James W. Thatcher. Plenum Press, 1972.

TOWARDS FEASIBLE SOLUTIONS OF THE TAUTOLOGY PROBLEM. Bradford Dunham and Hao Wang in *Annals of Mathematical Logic,* Vol. 10, No. 2, pages 117–154; December, 1976.

## ROMAN CARTHAGE

EXCAVATIONS AT CARTHAGE, 1974: FIRST INTERIM REPORT. Henry Hurst in *Antiquaries Journal,* Vol. 55, pages 11–40; 1975.

EXCAVATIONS AT CARTHAGE CONDUCTED BY THE UNIVERSITY OF MICHIGAN, 1975: I. Edited by John H. Humphrey. Tunisian National Institute of Archaeology and Art, 1976.

CARTHAGE 1977: THE PUNIC AND ROMAN HARBORS. L. Stager in *Archaeology,* Vol. 30, pages 198–200; 1977.

EXCAVATIONS AT CARTHAGE, 1975: SECOND INTERIM REPORT. Henry Hurst in *Antiquaries Journal,* Vol. 56, Part 2, pages 177–197; 1977.

## THE VISUAL CHARACTERISTICS OF WORDS

THE SIMILARITY OF LOWER-CASE LETTERS OF THE ENGLISH ALPHABET. Peter Dunn-Rankin in *Journal of Verbal Learning and Verbal Behavior,* Vol. 7, pages 990–995; 1968.

VISUAL INTERFERENCE IN THE PARAFOVEAL RECOGNITION OF INITIAL AND FINAL LETTERS OF WORDS. Herman Bouma in *Vision Research,* Vol. 13, pages 267–282; 1973.

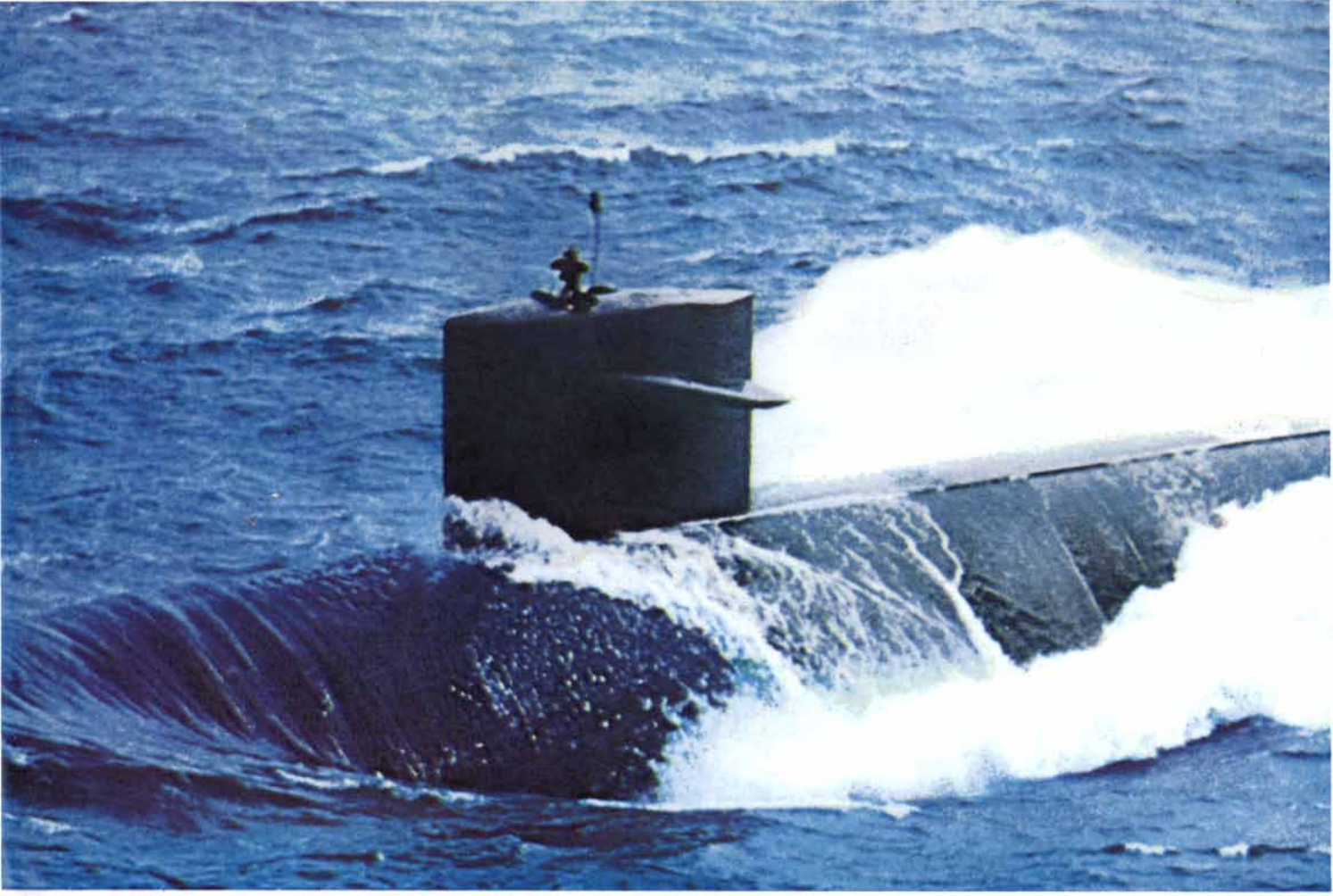
THE RESOURCES OF BINOCULAR PERCEPTION. John Ross in *Scientific American,* Vol. 234, No. 3, pages 80–86; March, 1976.

## THE AMATEUR SCIENTIST

A SIMPLE DEVICE FOR OBSERVING THE POLARIZATION OF LIGHT FROM THE SKY. Ulf Halldén in *Nature,* Vol. 182, No. 4631, page 333; August 2, 1958.

HOW BEES NAVIGATE. Allan W. Snyder and Colin Pask in *Nature,* Vol. 239, No. 5366, pages 48–50; September 1, 1972.

ESTIMATING THE SIZE RANGE OF SMOG AEROSOL PARTICLES WITH A PAIR OF SUNGLASSES. W. H. White in *Atmospheric Environment,* Vol. 9, No. 11, pages 1036–1037; 1975.



# HERE'S ONE ENGINEERING OPPORTUNITY YOU WON'T GET IN PRIVATE INDUSTRY.

If you're thinking about a technical position after graduation, think about this.

How many companies can offer you a nuclear submarine to operate? The answer is none. Equipment like this is available only in one place – the Navy.

The Navy operates over half the nuclear reactors in America. So our training is the broadest and most comprehensive. We start by giving you a year of advanced technical education. In graduate school, this would cost you thousands, but in the Navy, we pay you.

Once you're a commissioned Nuclear Propulsion Officer, you'll earn a top salary. Over $24,000 a year after four years. And you'll be responsible for some of the most advanced equipment developed by man.

The Navy also has other opportunities in surface ships and aviation assignments. If you are majoring in engineering, math or the physical sciences, contact your placement office to find out when a Navy representative will be on campus. Or send your resume to: Navy Officer Programs, Code 312-B382, 4015 Wilson Blvd., Arlington, VA 22203.

The Navy. When it comes to nuclear training, no one can give you a better start.

# NAVY OFFICER. IT'S NOT JUST A JOB, IT'S AN ADVENTURE.

# Last year, Technics gave you everything you wanted in direct drive. This year we're giving you less.

Introducing three new Technics turntables: The SL-1600 automatic, the SL-1700 semi-automatic and the SL-1800 manual. All with the Technics direct-drive system. The system FM stations use and discos abuse. And all with performance specs even better than last year's.

## Less Acoustic Feedback.

This year we did more to give you less acoustic feedback. With Technics double-isolated suspension system. A floating system that not only damps out external vibrations at the base but also from the platter and tonearm. Combined with our new sensitive gimbal-suspension tonearm, it dramatically reduces feedback. Even at high music levels.

## Less Wow & Flutter.

Amplified, even the smallest fluctuations in platter speed can be your biggest problem. So this year with a B-FG speed stabilizer built into our 321-element one-chip IC, you get infinitesimal wow and flutter, equal to Technics professional level. An incredible 0.025% WRMS.

## Less Rumble.

The less noise your turntable makes, the more music you hear. That's why professionals loved Technics last year. This year, we're even more lovable, because there's even less rumble, −73 dB (DIN B) to be exact.

## For Less Money.

Even with features like oil-damped cueing. A computer-analyzed S-shaped universal tonearm with anti-skate control. Pitch controls variable by 10%. And detachable dust cover. Technics turntables cost less this year than last.

Technics. We're giving you less. And that means you're getting more.

**Technics**
by Panasonic