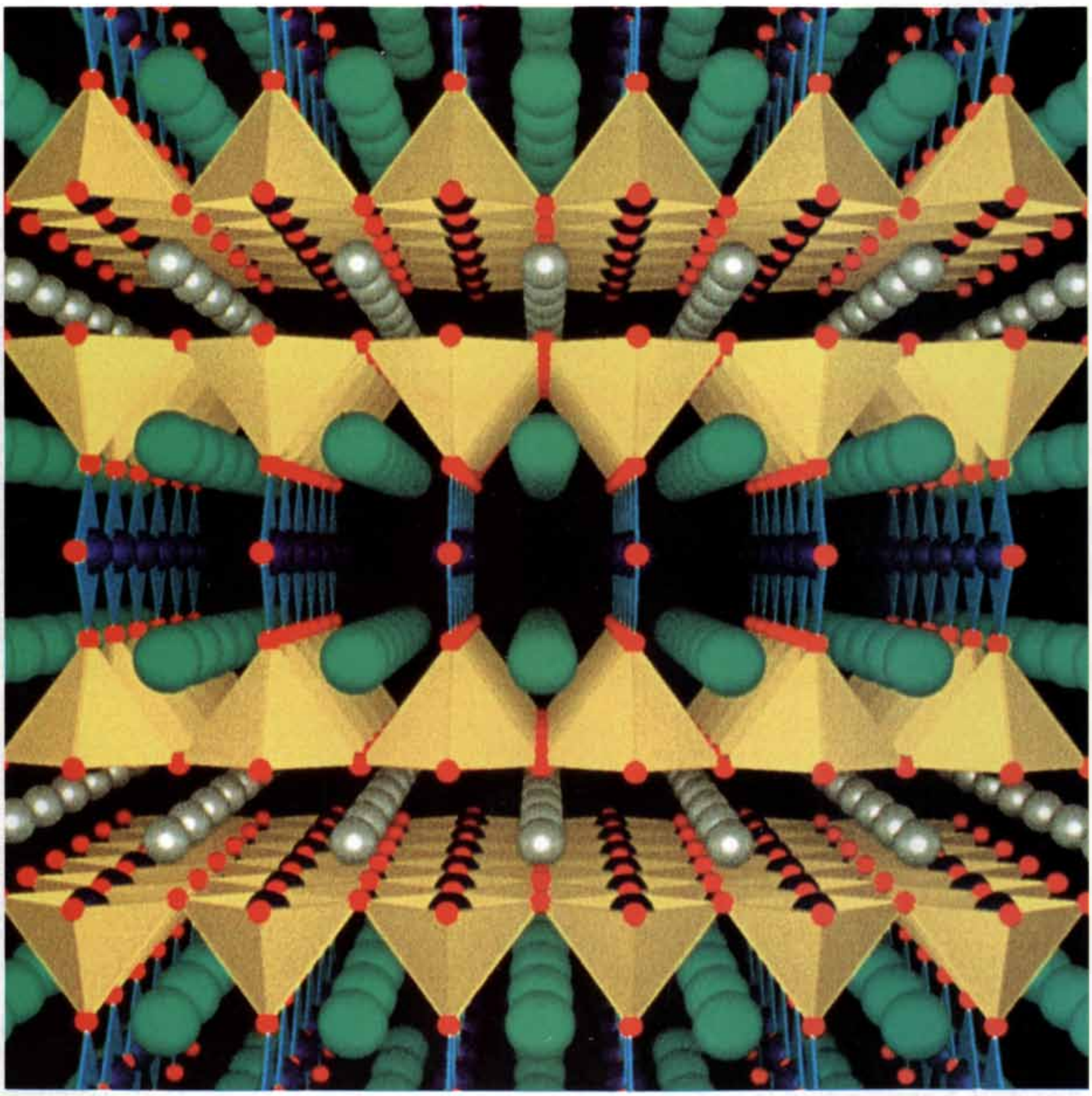
SCIENTIFIC AMERICAN

JUNE 1988
$2.50

How nonsense is deleted from genetic messages.

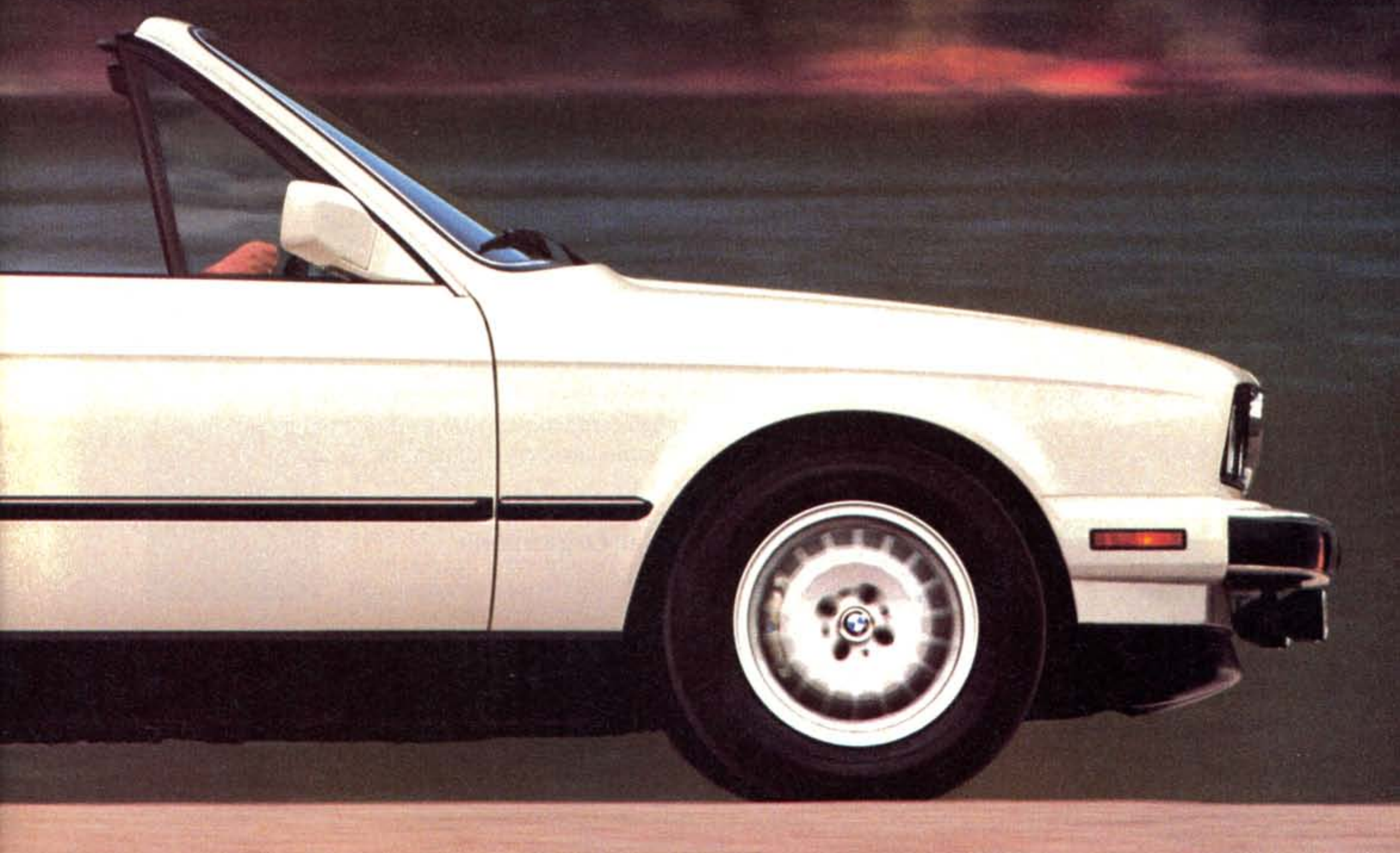
℞ for economic growth: aggressive use of new technology.

Can particle physics test cosmology?



***High-Temperature Superconductor** belongs to a family of materials that exhibit exotic electronic properties.*


A CONVERTIBLE BUILT FROM THE BOTTOM UP, NOT THE TOP DOWN.



SCIENTIFIC
AMERICAN

June 1988 Volume 258 Number 6

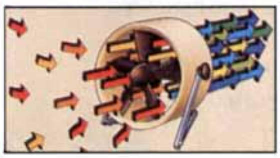
DEPARTMENTS

THE LIFESTYLE RESOURCE™

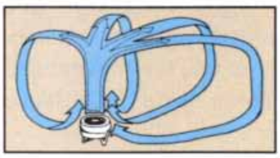
At The Lifestyle Resource we give you all the facts and details necessary to make an informed purchase. Your satisfaction is our primary concern. If your purchase doesn't meet your expectations, return it in original condition within 30 days for prompt refund.

AERODYNAMIC COOLING

The Turbo-Aire fan moves more *air and provides more cooling than ordinary fans* — you get more than twice the air delivered by oscillating room fans. Computer-designed to maximize efficiency and minimize noise and vibration, this fan creates an exceptionally strong, smooth jetstream column that sets a room stirring with secondary air currents. Refreshes better than the intermittent blast of air from an oscillating fan. Aerodynamic housing increases blade-tip efficiency over conventional fans. Adjusts to any angle. Floor, table or wall mount. Set in the hassock position, the five bladed, 12″ fan redistributes air in an entire room, makes it as useful in winter as in summer. Can reduce air-conditioning costs. 300% more energy-efficient than other fans this size. Weighs less than a case of cold soft drinks — and more easily moved! 2-yr. limited warranty. **$69.95** #2160.



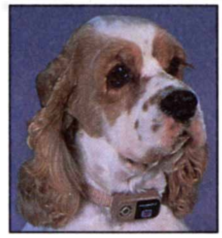
Turbo-Aire's jet stream



Hassock position — venturi effect

Conventional fan

BANISH FLEAS



Flea season is the bane of existence for most dogs and cats — and their owners. Pets suffer tremendous itching and pain from these pesky little varmints, not to mention the injury animals can cause themselves from continuous scratching and biting. Now there's a 100% safe, veterinarian-tested remedy that eliminates the use of poisonous chemicals and constant and expensive bathing, dipping, spraying and powdering your favorite pet. This clean and odorless electronic flea collar employs a pulse modulated burst circuit (PMBC) to create such a high frequency sound, inaudible to humans and pets, that the critters hastily abandon their host animal. This new and improved, water-resistant, Microtech-2® collar works on one long life lithium battery (included), and focuses on a four-foot zone of protection. Dog collar #2180 or safety breakaway collar for cats #2190; **$39.95** each.

FINGERTIP PHONE DIALER

It's like taking your desktop with you. The new 8K Sharp Dial Master holds up to 400 names and numbers with fast-search ability by name or company. Full two-line, 24-character, LCD display with automatic tone dialing at two speeds. One-touch redial. Credit card number and long distance computer access keys. Secret key requires password for access to protected information. Two separate memo files. 10-digit calculator with memory. Three telephone files let you separate business, personal and other numbers. 80-character name entry, 80-digit number entry. Easy-touch raised keys. Backup battery. 1-yr. warranty. Let the pocket size Sharp Phone Dialer put your desktop at your fingertips — anywhere, anytime. **$89.95** #2150.

BRING MOUNTAIN TOP FRESHNESS INDOORS

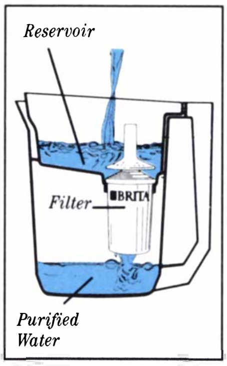
The new Bionaire 700, no bigger than a table-model radio, will clean and recharge your stale indoor air to virtually mountaintop freshness. Get relief from breathing allergy causing dust, pollen, tobacco smoke, animal hair and dander, cooking odors, soot, and mold spores. The Bionaire 700 will clean and rejuvenate the air in a 12x12x8 ft. room 4 times an hour, while the filtering system removes up to 99% of all particulate pollutants as small as .01 microns in the process. The filtering process begins with an activated carbon pre-filter that helps removes odors and large particles; next, with the patented electret main filter, the Bionaire removes particles as small as 1/10,000th the diameter of a human hair. Finally, Bionaire's unique negative ion generator—which not only precipitates any remaining particles, but also generates millions of negative ions—reproduces the stimulating effect of fresh mountain air. Two year limited warranty, UL approved, weighs 5.2 lbs., **$149.95** #2070.

ZONE OF CALM

Noise produces stress in the human organism. Today the volume of civilization nearly everywhere seriously interferes with our abilities to relax, read, sleep, concentrate or otherwise function at optimum efficiency. But you needn't be victimized by noise pollution. The *new* Marsona 1200A Sound Conditioner electrically synthesizes a variety of pleasing natural sounds that help mask out and reduce the annoyance of unwanted noise. You can simulate sounds of ocean surf, summer rain, mountain waterfalls. You control not only the volume but also wave pattern, wave or rain rhythm, the seeming nearness or distance of the source. The new Marsona is the finest instrument of its kind, through a 5 inch speaker it creates sounds as close to nature as you will find. UL listed. **$149.95** #2200.

THE LIFESTYLE RESOURCE™

WATER, WATER EVERYWHERE



Now Brita makes it taste good enough to drink. There are more elements in the water than just H_2O — like lead, copper, iron, rust, and chlorine. Until now your alternatives for quality water were expensive — bottled waters or a costly filtration system hooked up to your plumbing. Brita, with two decades of water filtration experience, has an elegant, easy and inexpensive solution — the Brita double pitcher system with a unique patented filter. Just add tap water and in minutes you have almost a half gallon (2.0 litres) of clear, clean, fresh tasting water. The E.P.A. registered bacteriostatic filter will remove the chlorine and 90% of the lead and copper, reduce water hardness and inhibit bacterial growth. Depending on composition of your local tap water, each replaceable filter can process up to 35 gallons of water . . . that's less than 2¢ per glass. Readily available replacement filters are of top quality carbon and resin and will not increase sodium content of water. Remarkably improves the taste of food prepared and cooked with Brita. Coffee and tea never tasted so good. Drop for drop Brita makes it better. Plastic pitcher is dishwasher safe. **$29.95** #2100; set of 3 replacement filters **$24.95** #2110.

SOOTHING SOUND MASSAGE

Tested more than 30 years, the Novafon sound massager is used in clinics and health centers around the world. Novafon's sound waves penetrate up to 2¼", and they help a variety of conditions because of their mixed frequencies. They bring relief from pain, loosen joints and soothe tired muscles, help circulation, speed recovery from exercise and over-exertion. The Novafon is lightweight (8 oz.) small (8″length). Adjustable intensity control, choice of massage heads (disc-type and ball-type). It comes in a fitted plush case, perfect for carry along. 1 year warranty. A precision made instrument with no interacting parts to wear out, the unit will give many years of service. **$169.95** #1750.

WORTH CUTTING THE CORD FOR

Southwestern Bell FF-1700 Offers—

- *1000-foot range*
- *Base/handset intercom*
- *Memory dialing*
- *Hearing aid compatible*
- *10-channel select*
- *Digital security code*
- *Hold button*
- *Tone/pulse switchable*
- *Jack for answering machine connection*

A leading consumer magazine writer likens a person's first conversation walking around talking on a cordless telephone to the exhilaration of that first bike ride minus training wheels. Later, that article rates Southwestern Bell's FF-1700 Cordless Phone tops for range in controlled tests among 21 brands and models. By handling incoming and outgoing calls to range of 1000′ (the article rated a maximum of 1500 ft.), with outstanding speech quality and convenience features, the FF 1700 ended on top of the consumer magazine ratings reports. Base unit serves as freestanding speakerphone with dialpad, so you get two phones in one. Plus intercom, paging and 10-channel selection. Digital security code protects line from outside access. **$179.95** #2130.

To take your freedom a step further, Southwestern Bell's FA-450 Telephone Answering Machine gives you the latest technology and newest features at a most attractive price. Single cassette operation, call screening, household memo function, voice-activated record, one-touch playback. Two-way record for messages or conversations. Beeperless remote lets you retrieve messages from any pushbutton phone, also allows remote announcement changes. These new-generation Freedom Phones connect you in without tying you down. **$99.95** #2140.

THE LIFESTYLE RESOURCE™ — CREDIT CARD ORDERS CALL TOLL-FREE 24 HOURS **800-872-5200**

DEPT. SFAF08; 921 EASTWIND DR. SUITE 114: WESTERVILLE, OH 43081

	ITEM NO.	QTY.	DESCRIPTION	ITEM PRICE	TOTAL PRICE
SEND TO (PLEASE PRINT)					
ADDRESS					
CITY STATE ZIP					
☐ CHECK OR MONEY ORDER ☐ MASTERCARD ☐ VISA ☐ AMEX SIGNATURE					
ACCT. # EXP DATE				SUB TOTAL	
				SHIPPING (see table at left)	
				TOTAL	

Shipping Charge covers UPS, handling and insurance for guaranteed delivery.

UPS 2ND DAY AIR

Up to	$20 .... $ 3.95	$50.01 to $60 .... $ 7.95
$20.01 to	$30 .... $ 4.95	$60.01 to $70 .... $ 8.95
$30.01 to	$40 .... $ 5.95	$70.01 to $100 .... $10.95
$40.01 to	$50 .... $ 6.95	Over $100 .... $12.95

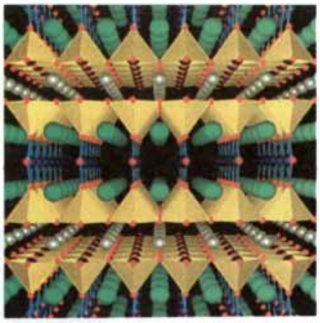
UPS Second Day available for an additional $7.50 per order.

ORDER WITH CONFIDENCE

- *We ship most orders within 48 hours.*
- *Credit Card orders billed only upon shipment.*
- *No risk 30-day return privilege.*

THE MEDIA DEVELOPMENT GROUP

©1988 G38 ADVERTISEMENT



THE COVER shows a computer-generated model of the "1-2-3" high-temperature superconductor, which, like the other recently discovered high-temperature superconductors, is a member of a crystallographic family known as perovskites (see "Perovskites," by Robert M. Hazen, page 74). The compound consists of atoms of yttrium (*gray*), barium (*green*) and copper (*blue*) in a 1 : 2 : 3 ratio, along with oxygen (*red*). The image was made with WINSOM software developed by the IBM Corporation.

THE ILLUSTRATIONS

Cover photograph courtesy of the IBM Corporation

Page	Source
45	The Bettmann Archive (*top*), Ford Motor Company (*bottom*)
46-47	Jerome Kuhl
48	General Electric Company
50-51	Jerome Kuhl
57-58	Ian Worpole
59	Yvonne N. Osheim, University of Virginia School of Medicine (*top*); Ian Worpole (*bottom*)
60-63	Ian Worpole
67-72	Andrew Christie
75	Quesada/Burke
76-78	George V. Kelvin, Science Graphics
79	Robert M. Hazen (*top*); Ross J. Angel, Geophysical Laboratory, Carnegie Institution of Washington (*bottom*)
80	Robert M. Hazen
81	George V. Kelvin, Science Graphics
83	Hans Reichenbach
84	James A. Shapiro
85	Jeffrey C. Burnham, Susan A. Collart and Barbara W. Highison (*top*), Gabriela M. Vasquez, Frank Qualls and David White (*bottom*)
86	James A. Shapiro
87	James A. Shapiro (*top*), S. A. Sturdza (*bottom*)
88	James A. Shapiro
89	Martin Dworkin
91	Josefino C. Comiso, National Aeronautics and Space Administration Goddard Space Flight Center
92	Ian Worpole
93-96	Josefino C. Comiso, National Aeronautics and Space Administration Goddard Space Flight Center
99	Courtesy of Academy of Motion Picture Arts and Sciences
100	Tom Prentiss
101	Tom Prentiss (*top and bottom*), Charles M. Lent (*middle*)
102	Charles M. Lent
103	Tom Prentiss
105	Royal Museum for Central Africa, Tervuren, Belgium
106	Michael Goodman
107	Francis Van Noten (*top*), Michael Goodman (*bottom*)
108	Francis Van Noten
110-111	Michael Goodman
112	Cetus Corporation (*top*); Philip Leder, Harvard Medical School (*left*)
117-119	Michael Goodman
120-123	Andrew Christie

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc, 415 Madison Avenue, New York, N.Y. 10017. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Subscription rate: $24 per year in the U.S. and its possessions; $33 per year in all other countries. Postmaster: Send address changes to Scientific American, P.O. Box 953, Farmingdale, N.Y. 11737-0000.

SCIENTIFIC AMERICAN

Established 1845

How Macintosh II came to be one of the safest decisions in business.

First, we meet your standards.

Business computing has recently undergone some fairly radical changes. Consequently, so has the idea of "safety."

Because, at present, only one computer offers all the promise for the future, and delivers on it today: the Macintosh® II personal computer.

Rest assured, it has the power to meet every standard that serious business demands.

Like breakneck speed. Full expandability. A wide choice of sophisticated programs for every business use. And the capacity to store even the most intimidating mountain of information.

It has memory that grows beyond megabytes, all the way to gigabytes. And can display your work in a choice of millions of colors, on monitors that show up to four pages at once.

Its own power aside, Macintosh II is perfectly prepared to meet the standards of other machines as well. Whether they speak MS-DOS, UNIX® or assorted dialects of mainframese, from IBM to DEC.

So it can work with files from–and run–DOS programs like Lotus 1-2-3 and WordPerfect.

Connections are painless via the AppleTalk® network system, using a variety of cabling, including Ethernet®.

Keep in mind, you can use Macintosh technology no matter what size your thirst for power. There's a whole family of Macintosh computers, all designed to connect to existing equipment.

Macintosh II just happens to make a perfect place to start, because you can add so many options down the road.

Except, of course, the warm feeling of security.

That comes standard.

Compatibility. *Easily digests data from–and even runs–MS-DOS programs.*

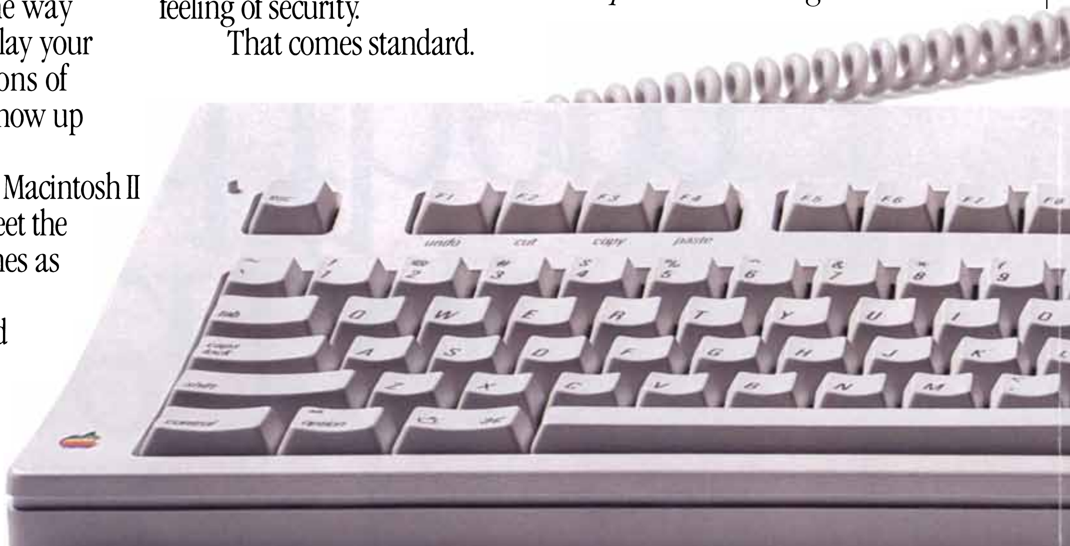
Display. *All sizes and shapes, including color and large-screen.*

Applications. *A wide choice for every business use.*

Connectability. *Via AppleTalk, using cabling such as Ethernet.*

Expandability. *Six expansion slots for future growth.*

Raw Power. *Math co-processor standard. Vast memory 1 MB and up. Internal storage to 80 MB.*




File Edit Search Format Font Special Windows
Lotus 1-2-3
'Confidential
READ
HD40
Eighth National Savings and Loan Fact Sheet
Confidential
STATEMENT OF INCOME AND EXPENSE SUMMARY
CASH ASSETS
VersaTerm Pro-DEC VT100 Emu
PAYROLL EXPENSE
LEASEHOLD IMPROVEMENTS
RENT
Doc 1: WordPerfect for Macintosh™
Dear Mr. Morrison:
Per your request, attached please find the complete operating budget for Fiscal Years 88/89. Of particular interest is Third Quarter/88, in which HiSP's acquisition of ENS&L will be complete.
In the following graph, I've isolated the impact of the acquisition. Combining all available data, (i.e., that from our DEC database and ENS&L's PC files), it's clear that total cash assets will be somewhat lower than normal at closing. But by the end of the First Quarter/89, ENS&L will be contributing substantially to higher net profits.
$30M
$25M
$20M
$15M
$10M
$5M
$0M
Q3'88
Q4'88
Q1'89
Payroll
Leasehold Improvements
Rent
Cash Assets
Macintosh II

Then, we raised them.

There's more to good business than simply meeting standards. Which is why we've put so much effort into exceeding them.

Macintosh II is, first and foremost, a Macintosh–using the graphics-based interface we pioneered.

Beyond that, it's a new generation of computer, working at full strength with an operating system that exploits every bit of its power.

The latest part of that system, MultiFinder,™ adds multitasking capabilities.

So you can go about your work even more naturally, switching back and forth between jobs as the situation requires. Or doing a number of things at once.

Practically speaking, that means you can print a document while you compose a letter. Or receive E-Mail while you recalculate a spreadsheet.

It means you can use Macintosh to move information effortlessly between programs, merging graphics and text. Even link data dynamically, so changes in one document are automatically incorporated in another.

This higher standard adds new value to existing Macintosh software (and to existing Macintosh computers).

But even more intriguing, it creates a world of new possibilities.

For example, now you can use advanced programs for Apple® Desktop Publishing–the standard we created over two years back–right alongside your business programs. So it's even easier to integrate a number of different efforts into a single document.

And, in the Macintosh tradition, what

Graphical Interface. *We pioneered the operating system that others are only now beginning to imitate. Macintosh programs offer consistency, with clear graphical options instead of mysterious commands. So training costs are cut drastically.*

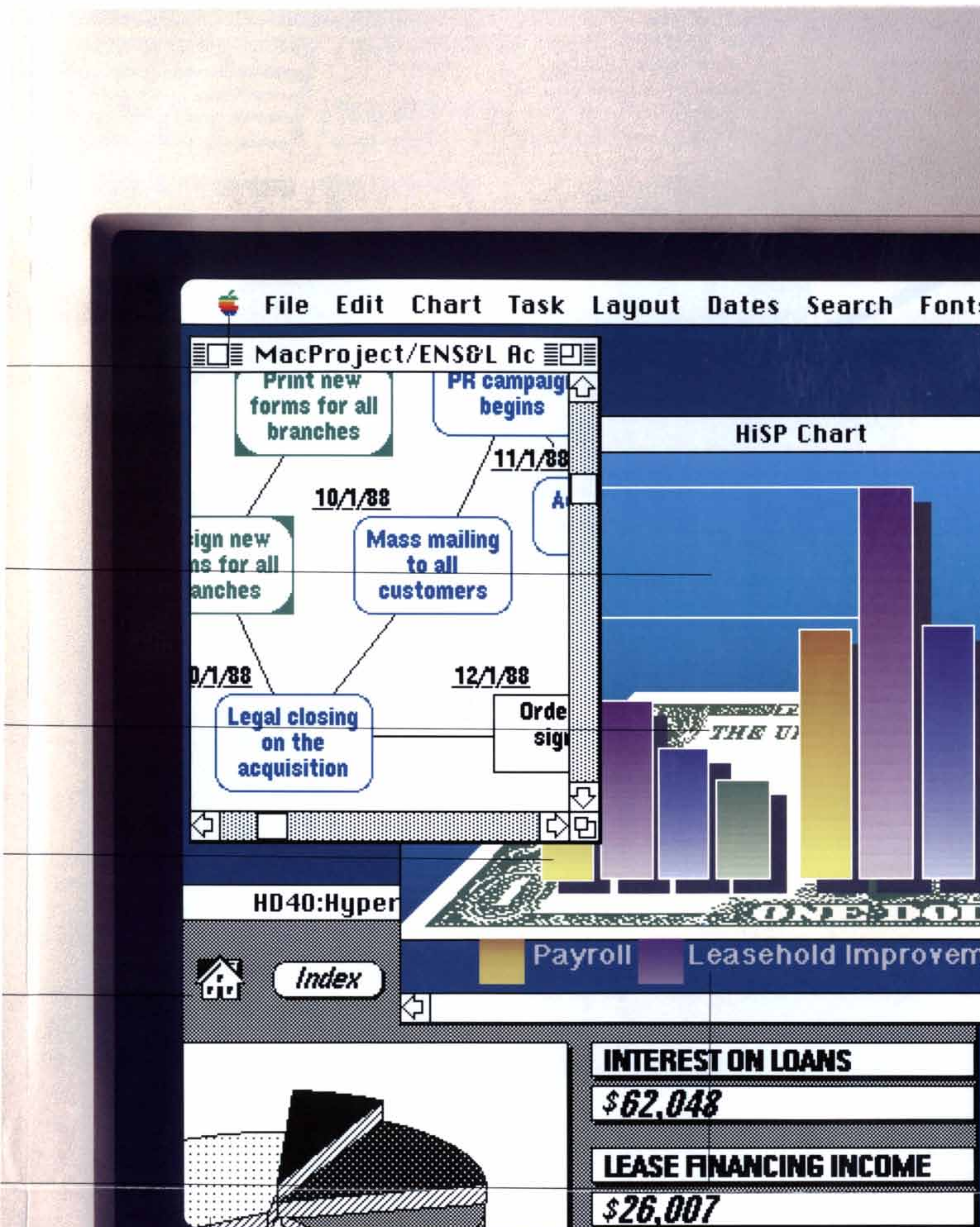
True WYSIWYG. *As in "what you see is what you get." Macintosh is designed for graphics from the ground up. You don't have to guess about what's going to appear on paper.*

Multitasking. *It's a fact, not a promise. Thanks to MultiFinder,™ you can run multiple programs simultaneously. Move back and forth, or do several things at once.*

Integration. *Cut and paste information freely between applications, even those made by different software companies. Combine text and graphics as you please.*

HyperCard.™ *A revolutionary tool that lets you link "stacks" of information by simple association, instead of complex command. Without extensive training, you can create your own powerful programs to manage ideas and information.*

Applications. *Just what you'd expect from a new-generation computer: new-generation programs in every*


File
Edit
Chart
Task
Layout
Dates
Search
MacProject/ENS&L Ac
Print new
forms for all
branches
PR campaig
begins
11/1/88
10/1/88
Mass mailing
to all
customers
12/1/88
Legal closing
on the
acquisition
HiSP Chart
HD40:Hyper
Index
Payroll
INTEREST ON LOANS
$62,048
LEASE FINANCING INCOME
$26,007

even easier to integrate a number of different efforts into a single document.

And, in the Macintosh tradition, what you see on screen is what you can expect to see on paper.

Like all Macintosh programs that have come before, the new generation offers a consistency that's unique to computing.

Use any one and you quickly gain a feel for the others – something that can translate very quickly into reduced training costs.

The best part is, this technology lets you raise the basic standard of how you get your information.

On your desk, Macintosh II becomes a window to all your computers, be they PC or mainframe. And lets you access and analyze different kinds of information in the same, intuitive Macintosh way.

Which means not only higher standards for the people using Macintosh, but higher standards for the equipment you already own.

Now, if you'd like to find out more about raising the quality of your work, it might help if you do a little sight-seeing.

Visit an authorized Apple dealer, and examine Macintosh II in detail, along with the new family of LaserWriter® II printers. (Call 800-446-3000, ext. 333, if you need the address of one nearby.)

Then you'll understand why so many business people feel the same way about getting a Macintosh II:

Better safe than sorry.

from a new-generation computer: new-generation programs in every area of business use, designed to take full advantage of the power of Macintosh II. Not next year. Today.

Transparent Access. *Use Macintosh II as your window to the world, and all your information is accessed the same, intuitive way. Whether it's in a Macintosh, PC, or in minis and mainframes from IBM and DEC.*

NuBus™ Technology. *Add vast memory, run alternative operating systems or tie into networks and mainframes. Unlike other computers, Macintosh II offers 32-bit expansion slots based on a published architecture – so future development is uninhibited.*

$26,007
Macintosh II
F1
F2
F3
F4
F5
F6
F7
F8
undo
cut
copy
paste

LETTERS

To the Editors:

In "The Ancestry of the Giant Panda" [SCIENTIFIC AMERICAN, November, 1987] Stephen J. O'Brien gives the impression that the relationships of the giant panda have persistently defied elucidation by comparative anatomical methods. We believe that this is a misleading portrayal of the debate's historical development. Although O'Brien's molecular data provide important new evidence supporting the close relationship of the giant panda to bears, this proposition is not without firm precedent. A stable notion of panda affinities did elude earlier workers, as O'Brien states. Controversy about the ursine ancestry of this genus ended, however, with publication of D. Dwight Davis' meticulous anatomical description of the giant panda in 1964.

O'Brien regards the poorly substantiated objections of R. F. Ewer and Desmond Morris as having raised serious doubts about the validity of Davis' conclusions among carnivoran systematists. The objections raised by Ewer and Morris took two forms: they listed features of the giant panda not found in other bears, and they speculated in behavioral and functional terms about patterns of transformation and adaptation. Both of these issues are irrelevant from the standpoint of phylogenetics.

Like Ewer and Morris, O'Brien emphasizes ways in which the giant panda differs from other bears: the bleating of the panda (as opposed to the roaring of other bears), the extension of the radial sesamoid bone on the panda forepaw into an opposable thumb, its largely herbivorous diet, its ambling gait and so on. These workers failed to recognize that in systematics, unique traits of an organism do not provide information about its evolutionary affinities. The unique characters of the giant panda neither support nor contradict its relation to any other carnivoran. Even placing the giant panda within its own family because of its unusual specializations, as some workers have done, does not conflict with Davis' proposition that the giant panda is most closely related to bears.

The behavioral "evidence" O'Brien cites as having cast doubt on the proposition that the giant panda is a bear was first adduced in 1966, in the book *Men and Pandas,* by Ramona and Desmond Morris. It can hardly be considered authoritative (nor was it intended to be). The authors state quite casually: "If we are forced to cast an intuitive vote it must be for the raccoon school. We cannot avoid a feeling that somehow the giant panda is a counterfeit bear rather than a true one." Such offhand comments are scarcely cause for rejecting sound evidence.

Without debating the systematic utility of the DNA-DNA hybridization technique, we note that O'Brien's data provide surprisingly little resolution of the relations among species of ursine bears. Also significant is the inability of this allegedly powerful molecular technique to solve the truly perplexing question in panda relations: the evolutionary position of the lesser panda.

As a final point we question the declarative tone of the article, which makes common use of such phrases as "proved," "now been solved" and "finally settled." Just as Vincent M. Sarich viewed his 1973 immunodiffusion results as corroborating Davis' conclusions, so we regard O'Brien's findings not as a demonstrated "answer" to a particular phylogenetic question but as lending further credence to a hypothesis carefully formulated by Davis nearly a quarter of a century ago.

JOHN J. FLYNN

ANDRÉ R. WYSS

Department of Vertebrate Paleontology
American Museum of Natural History
New York

To the Editors:

Since our original publication in *Nature* of molecular data relating to the phylogeny of the giant panda there have been a number of published comments, largely from morphologists, affirming our primary conclusion (that the giant panda diverged from the ursid, or bear, line several million years after the split between the ursids and the procyonid, or raccoon, line) but in several cases suggesting that the issue had been settled when Davis published his monograph *The Giant Panda: A Morphological Study of Evolutionary Mechanisms* (in *Fieldiana: Zoology Memoirs*). I am pleased that a consensus seems to have emerged between molecular and morphological disciplines but surprised at the suggestion by Drs. Flynn and Wyss that "controversy about the ursine ancestry of this genus ended...with publication of D. Dwight Davis' meticulous anatomical description in 1964." Although I agree that Davis' study was masterly, I cannot agree that it ended the controversy. Even a cursory review of the post-1964 scientific literature on the topic reveals that the debate has flourished and the controversy has continued unabated until the present.

When the article was originally submitted to *Scientific American,* it included a complete table of the articles addressing the question of the giant panda's phylogeny that have appeared since Père Armand David first described the animal in 1869. The table was dropped from the published article in order to save space (both the editors and I were persuaded of the extent of the controversy), but most of the citations can be found in Davis' monograph; in "A Bibliography of the Giant Panda, *Ailuropoda melanoleuca,*" by D. Jarofke and H. Ratsch (in *Proceedings of the International Symposium on the Giant Panda*), and in *The Giant Panda,* by Ramona and Desmond Morris. Of the 42 citations in the table, 18 supported placement of the giant panda in Ursidae (the bear family), 13 in Procyonidae (the raccoon family) and 11 in a separate family, Ailuropodidae or Ailuridae. Nineteen of the studies appeared after Davis' monograph. Of this latter group, four supported Procyonidae, seven supported Ursidae and eight concluded that the giant panda was sufficiently divergent to belong in a separate family.

I do not mean to disagree with Davis' anatomical study; in fact, our own data support his primary conclusions. I do mean to state that, contrary to Flynn and Wyss's assertions, Davis' conclusions were not widely accepted in the scientific or popular literature. Graphic displays at the National Zoological Park in Washington, D.C., and at the British Museum of Natural History today emphasize the extent of the controversy and do not affirm Davis' or our interpretation. Stephen Jay Gould is one who was persuaded by Davis' monograph, but even he has stated (in the February 1986 issue of *Discover*) that "Davis' personal tragedy must reside in his failure to persuade his colleagues."

The primary criticism leveled at Davis' conclusion is that his anatomical analysis, although lengthy, did not follow standard principles of systematics. The critics pointed out that many of the characters Davis suggested are homologous in pandas and bears are irrelevant since the same traits are found in other carnivores as well. Ewer and Morris each gave a comprehensive review of the phylogenetic evidence presented by Davis and other workers and concluded the giant panda is a procyonid. Giles T. MacIntyre and Karl F. Koopman dismissed Davis' study as neglecting comparative measurements on red pandas and raccoons in favor of extensive comparisons of giant pandas and bears. By his own admission (see page 11 of his monograph) Davis simply concluded a priori that the giant panda is a bear and made little attempt to present comparative data because "this became so difficult I gave up."

A more recent scientific contribution to the question comes from George B. Schaller and his colleagues, who reviewed the anatomical and morphological features of pandas in the book *The Giant Pandas of Wolong,* published in 1985, and concluded that the giant panda and lesser panda are closest relatives and belong to a separate family, Ailuridae, equally distant from bears and raccoons. The notion of a separate family for pandas has also been cogently argued by such authorities as John F. Eisenberg, V. E. Thenius, Doris H. Wurster-Hill and several Chinese paleontologists. I am not persuaded that placement of the giant panda in a separate family "does not conflict with Davis' proposition that the giant panda is most closely related to bears," since virtually every author's stated reason for this taxonomic assignment was to resolve the bear-raccoon family conflict and not to ignore it.

One week before our *Nature* article appeared, Brian Bertram noted the differences between our results and Schaller's deductions in his review of Schaller's book, also in *Nature.* Even more recently, two authorities on carnivore evolution and systematics, Björn Kurtén and Leigh M. Van Valen, have constructively continued the discussion in *Nature* by emphasizing the consistency of the fossil record with our molecular topology. Implicit in their comments was the notion that the issue profited substantially from the molecular and cytological analysis.

Finally, in 1986 a text on the anatomy of the giant panda just as daunting as Davis' was published in China by a consortium of zoologists from the Beijing Zoological Gardens and the associated universities. These authors presented 600 pages of anatomical data based on 27 specimens (Davis had one!). They concluded that "the giant panda is different from the bear...we are in favor of assigning an independent family for the giant panda."

The declarative tone of the *Scientific American* article was unintentional and perhaps resulted from the shortening of the manuscript in the editing process. In retrospect, however, it does reflect my own impressions for two reasons. First, the work I describe involved several talented collaborators and colleagues (Raoul E. Benveniste, Janice Martenson, Mary A. Eichelberger, William G. Nash, David E. Wildt, Mitchell E. Bush and Vincent M. Sarich); I am personally rather proud of their work and convinced by the consensus of conclusions. Second, these days nearly everyone seems to agree with the derived phylogeny, including the proponents of the present discussion. As Gould remarked in *Discover:* "The proof involved an array of independent techniques, all using the latest methods of molecular biology. Genes don't lie; the issue is closed."

In summary, I am delighted that our molecular/cytological phylogeny agrees with Davis' conclusions. But in deference to other morphologists and taxonomists who have different opinions and in the light of well-documented discrepancies between morphological, molecular and cytological conclusions about phylogeny for this and other taxa, the time has long since passed when taxonomic issues can be resolved using a single set of criteria. The giant panda provides an excellent example of how a multifaceted approach can produce a consensus phylogenetic tree that is consistent with molecular, morphological, cytological and paleontological findings. It is my own opinion that prospects for future advances in evolutionary biology lie in large part in the synthetic interpretation of molecular and morphological results.

STEPHEN J. O'BRIEN

Laboratory of Viral Carcinogenesis
National Cancer Institute
Frederick, Md.

To the Editors:

I read the article "Helium-rich Supernovas," by J. Craig Wheeler and Robert P. Harkness [SCIENTIFIC AMERICAN, November, 1987], with considerable interest, in particular because we in New Zealand have had quite a grandstand seat for supernova 1987A.

Let me correct one error. Early in the article the authors refer to Albert Jones as "another Australian amateur." Actually Jones is a New Zealander, who has just made his first trip outside the country—to Australia, as it happens.

To confuse New Zealanders with Australians raises hackles on both sides of the Tasman Sea. It is as much a faux pas as confusing Canadians with citizens of the U.S.

RODNEY R. D. AUSTIN

New Plymouth
New Zealand

To the Editors:

We did not know the facts, and we apologize to all our friends in New Zealand and Australia whose sensibilities this slip may have offended. We offer as a token of our good faith the translated version of the article in some of the foreign editions of *Scientific American,* where the error was corrected.

J. CRAIG WHEELER

ROBERT P. HARKNESS

Department of Astronomy
University of Texas
Austin

Corsica, Chevrolet's sporty new front-drive sedan, carries a lot more under its smooth shape than seating capacity for five passengers.

Corsica also carries over 30 U.S. Government patents on everything from the fuel injectors to the sun visors.

Some of these patents help provide the instant starts and smooth, powerful operation of Corsica's standard 2.0-liter fuel injected engine. Others help give Corsica its reliability, durability and convenience. Many require the mind of an engineer to understand. But isn't it reassuring to know that engineering minds achieved so many breakthroughs in designing an affordable, contemporary sedan? Corsica–so advanced it's patented.


Let's get it together... buckle up.
GM 6/60
QUALITY COMMITMENT PLAN
The smoothest Chevrolet sedan ever built: 0.33 drag coefficient.
Available 2.8 Liter Multi-Port Fuel Injected V6 power.
Most passenger room of any compact sedan in America.
Standard base-coat/clear-coat paint for added luster.
6-year/60,000-mile powertrain warranty.
See your Chevrolet dealer for terms and conditions of this limited warranty
THE Heartbeat OF AMERICA
TODAY'S CHEVROLET

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JUNE, 1938: "A private newspaper with any spot in your home as the press room is available today to anyone in the United States possessing an ordinary radio receiving set. No thundering press will deafen you while your newspaper is being printed; instead, equipment contained in a small, attractive box will silently print your 'latest edition' while you sleep. The name of this service now available is facsimile. It involves the conversion of illustrations or other copy, such as printed matter, into electrical signals, which can be sent over radio or telephone circuits. At the receiver the signal is automatically converted back into visible form, appearing as a recorded replica of the original copy."

"In cell division, chromosomes are seen to reproduce themselves. The gene, the foundation of the chromosome's architecture, must do likewise. Here is an almost unbelievable, and a wholly novel, ability of a molecule: to create its like out of the lesser molecules of a suitable surrounding medium."

"New patrons of a bank at Suffern, New York, are puzzled these days by a curtain of pale bluish light that falls between them and the tellers, along the grilled windows. For some reason bank patrons try to get as close to tellers as possible, and bank employees suffer more than an average number of colds. Invisible ultraviolet radiation emitted by slender 30-inch tubes at the tops of the grilles protect tellers in this bank from infection."

"The personalities of the elements are determined by their electrons—'genes' of the atom. At all times their behaviors are predictable and consistent, for genes are constant. Standing at one end of the line of elements are the noble gases, exemplified by neon, with their symmetrical structures of eight electrons and their rugged, self-contained atoms. In the center of the line is carbon, greatest exponent of communistic sharing; each atom is joined to four fellows through the medium of its four electrons with rigid and unyielding bonds. At the other end of the line stand copper, silver and gold, soft and yielding, adaptable and flexible, poor in electron genes but rich in mobility."

"We have good reason to believe that ordinary stars derive their energy from the gradual transformation of hydrogen into other elements, and have still a long life of luminosity before them. But the white dwarfs, with their enormous density, undoubtedly are throughout most of their substance very nearly in the degenerate state that marks the final stage of a star's history. They should not be in this state unless they have exhausted almost all their internal energy and lived through all but the final stages of their lives."

"Power for transmitting telephone conversations across certain parts of the desert in the Southwest will be supplied by windmills, as a result of tests recently completed by engineers of the Bell Telephone Laboratories. The windmills will be installed on the new trunk route recently put in operation across New Mexico and Arizona. This important line to the Pacific coast traverses wild country in spots, into which it is impracticable to run power lines. The windmills drive generators that charge the batteries supplying current to vacuum-tube amplifiers, or 'repeaters.'"

SCIENTIFIC AMERICAN

JUNE, 1888: "The latest proposal for utilizing the wasted water power of Niagara Falls is to tap the Niagara River at some distance above the falls by means of a tunnel driven along the side of the river. The water would be distributed by means of lateral underground conduits to turbines placed on the bank below the falls. These could give power direct to mills, factories, etc., and by electrical transmission furnish light and power to Buffalo and neighboring towns."

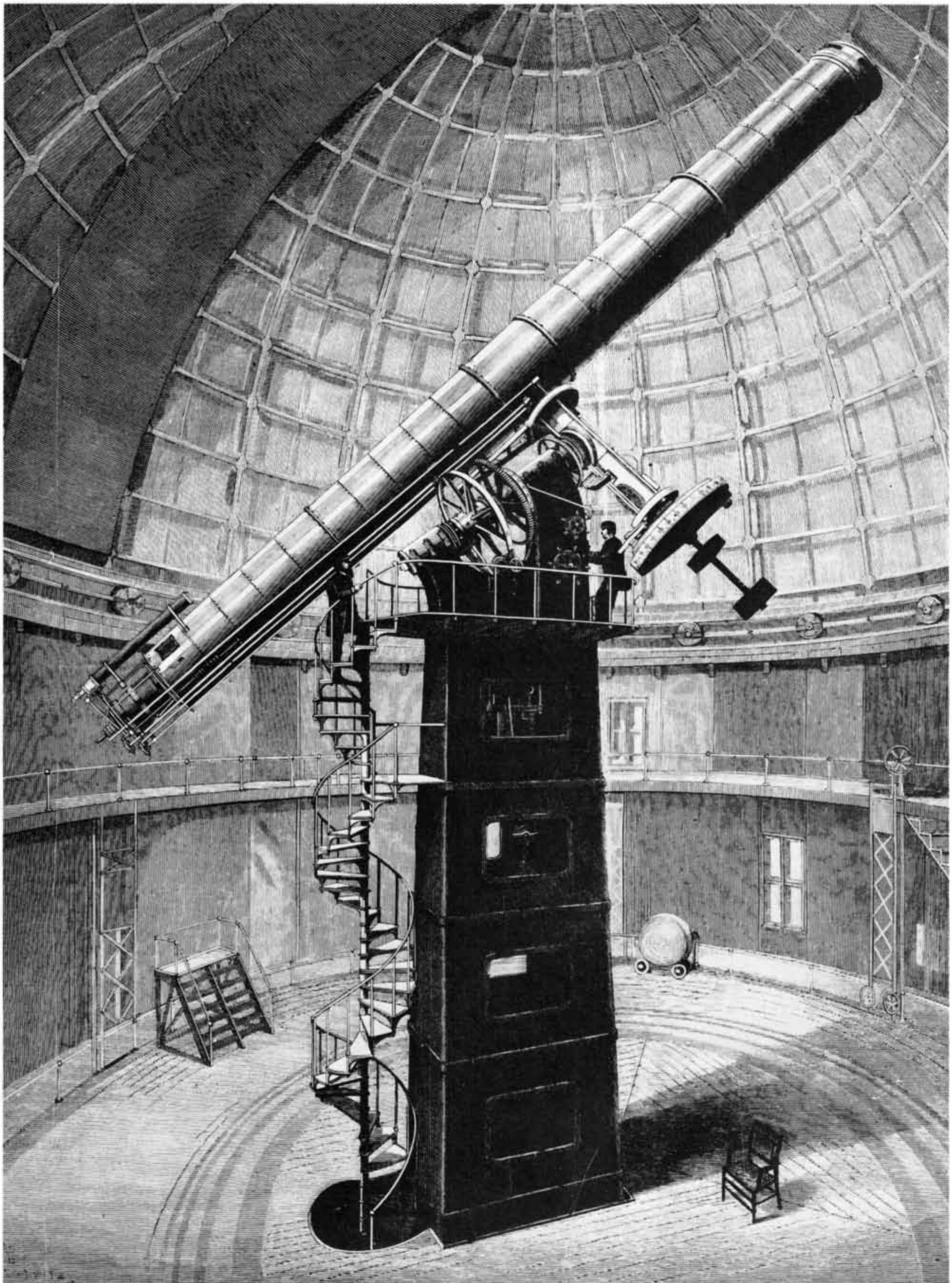
"A laundryman in the vicinity of Paris has discovered a very ingenious method of cleaning linen without soap. He uses no soap, nor lye, nor chlorine, but replaces these substances by boiled potatoes, with which he rubs the linen."

"A speed trial between the telegraph and the telephone from New York to Boston was lately undertaken at the *Sun* newspaper office in this city. The contest lasted for ten minutes; 330 words were delivered in Boston, ready for the printer, by the telegraph, and 346 words by telephone. But many of the telephone words were incorrectly received. And so the telegraph was the winner."

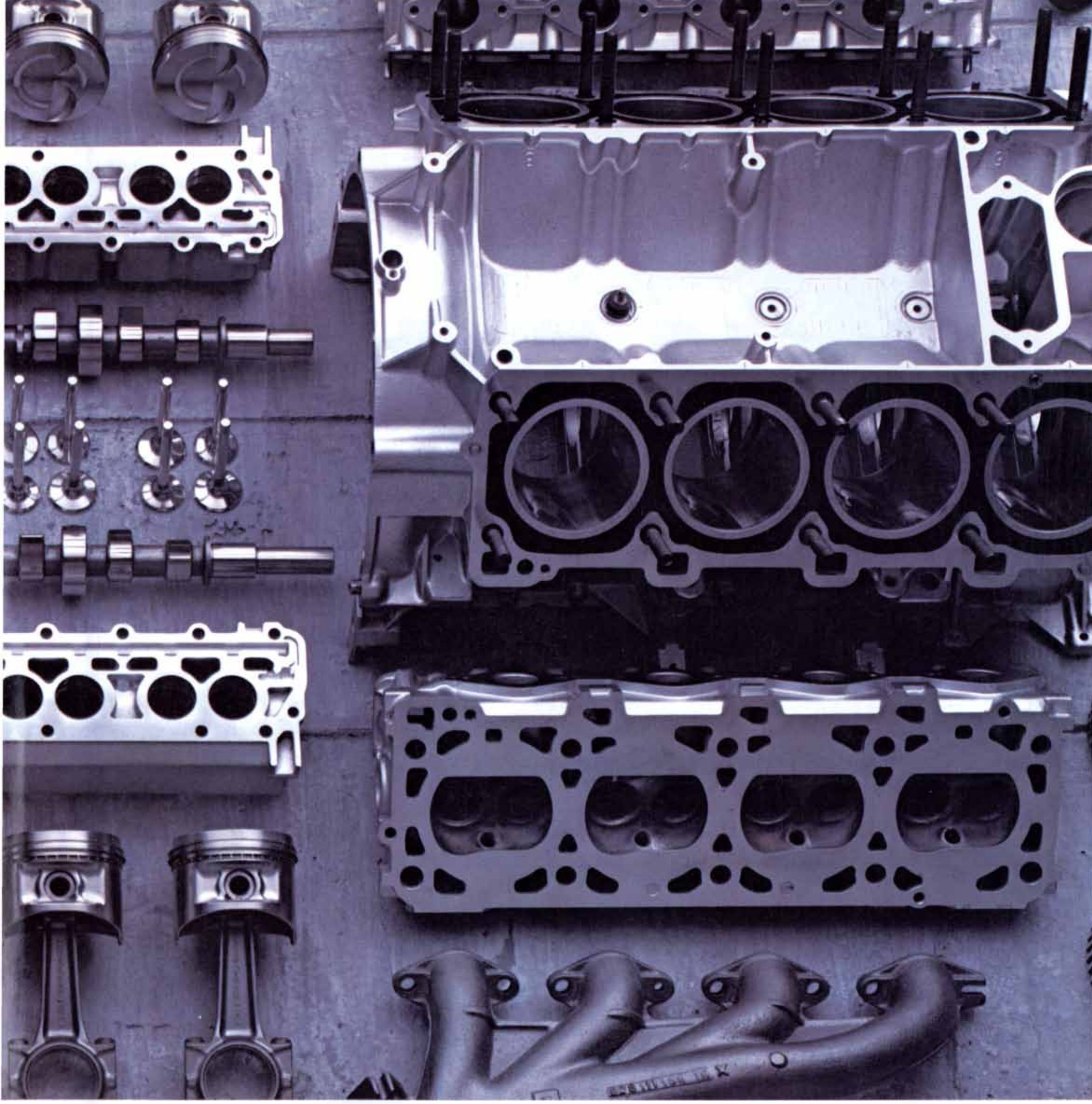
"Thomas A. Edison, the inventor, has been interesting himself with his new baby and a phonograph at his home. When the baby crowed with glee, the crow was registered on the phonograph; when it got mad and yelled, its piercing screams were irrevocably recorded on the same machine. That phonograph is now a receptacle of every known noise that is peculiar to babyhood."

"The State of New York may pride herself in the fact that the gallows is to be banished and a more humane and scientific method of executing criminals is to be instituted. On June 4 Governor Hill signed the bill autho rizing that criminals should be put to death by an electric shock. The subject of execution by electricity has long been argued before the public, but New York State stands as the first government that has undertaken to make the experiment of its practicality. The failure and barbarity of the old system have been amply demonstrated."

"The great telescope of the Lick Observatory was mounted in the south dome on Mt. Hamilton in the early part of the present year, and the observatory is now practically completed. The interior walls are of California redwood, handsomely finished with a dead surface to prevent annoying reflections; overhead the dome is painted pale pea green, the edges of the girders and intercostals and the square tie plates salmon pink, giving an appearance of airiness and lightness to the structure which is in harmony with its movable character. The somber black with which the great instrument in the center is painted, relieved from absolute deadness by the polished brass work of the fittings, increases the ponderous aspect of the telescope and asserts the dignity of its purpose. The only actual work which has been attempted so far is a series of micrometric measurements of the satellites of Mars; the brightness with which these ordinarily difficult objects appear in the great telescope attests the extraordinary light-gathering power of its objective."



The 36-inch telescope of the Lick Observatory



SHOULD HAVE A FUEL THAT IS EQUALLY EXACTING.

tankful of gas. So not only is Exxon 93 Supreme able to prevent this problem, it is also able to repair it.

To keep your engine from being at the mercy of the weather, Exxon fuels are fine-tuned for climates as well as engines. Fine-tuned every month for 21 geographic regions to ensure the fuel delivers the proper performance, regardless of weather, altitude, or season. So on a hot day the gasoline won't evaporate in the fuel lines causing vapor lock. And on cold days the gasoline will vaporize properly so your car will start quickly and accelerate when you need it to.

So, if you're driving an engine designed to aerospace tolerances, fill it up with the precision equipment that respects that fact. Exxon 93 Supreme.

Just turn the key.

EXXON 93 SUPREME.

Precision equipment for precision engines.

SCIENCE AND THE CITIZEN

Earning Its Wings

Hypersonic flight is on the way, but don't make reservations yet

When President Reagan announced the National Aerospace Plane (NASP) in his 1986 State of the Union Address, he called it the "Orient Express": a craft that by the end of the next decade would take off from a commercial airport such as Washington's Dulles International and either accelerate to 25 times the speed of sound to enter a low orbit or fly to Tokyo in two hours.

Today the NASP is a classified research project managed at Wright-Patterson Air Force Base in Ohio. The planned budget is $3.3 billion; about 70 percent comes from the Air Force, the rest from the National Aeronautics and Space Administration. The goal is to build by 1995 two experimental manned aircraft, designated X-30, that will use air-breathing engines and will explore single-stage ascent to orbit from a runway.

Whether that goal is reached depends on several breakthroughs, particularly in the power plant. Rockets, the traditional route to space, carry their own oxidant, which is usually heavier than the fuel; jet engines can use readily available air. Existing jet engines, however, will not work above about Mach 5 (five times the speed of sound): slowing the inrushing airstream enough to allow fuel to burn causes overheating. Above Mach 4 the X-30 will use supersonic combustion ramjets, or scramjets. A scramjet injects hydrogen—the only fuel that burns fast enough—from a series of cooled struts directly into a supersonic airstream. Robert R. Barthelemy, director of the NASP joint program office, says he is "leaning toward" slush hydrogen, a mixture of solid and liquid hydrogen.

But until the X-30 is built no one will know whether air-breathing engines can reach orbital velocity, about Mach 25. Wind tunnels cannot realistically test engines above about Mach 8. Furthermore, the X-30 requires a radically new design approach. The airframe and engines cannot be developed separately and mated later. In order to maximize thrust they must be entirely integrated, and so extensive modeling on supercomputers will be essential.

Subscale scramjets have shown useful thrust in tests at up to Mach 7, Barthelemy says; full-scale tests later this year are planned by the two engine contractors, Pratt & Whitney and Rocketdyne. Simulations suggest that scramjets can provide thrust at up to Mach 24, but Barthelemy notes that "scaling laws don't necessarily apply" in fuel mixing and combustion.

Four kinds of engine will probably power the X-30, although two modes might be combined. Conventional turbojets will operate at low speeds. Ramjets would cover the low supersonic regime. At Mach 4 the scramjets would fire up. For orbital flight the X-30 would rely on rockets.

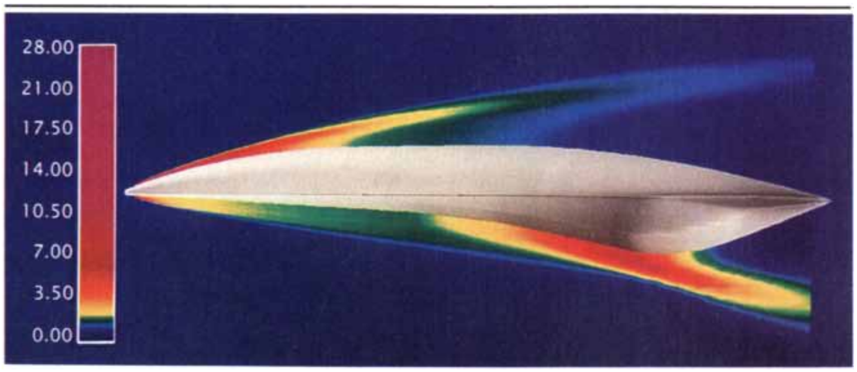
Before the X-30 can fly, materials must also be developed that can survive the heat of Mach 25 flight and the chemical ravages of liquid hydrogen. The nose cone and parts of the engines might reach 3,000 degrees Fahrenheit, and weight is at a premium. The hottest structures will be cooled by "heat pipes" containing circulating liquids. Large parts of the craft will be actively cooled by liquid hydrogen circulating in interior channels.

The NASP contractors are cooperating on materials research. Carbon-carbon composites will be used for the hottest surfaces; a central question, says Barthelemy, is whether they can be coated with a ceramic to prevent oxidation. If they can, wing tips and cowls—hard to cool actively because of their sharp edges—might not need active cooling. Other composites and rapidly solidified titanium-aluminum alloys will also be essential.

Three contractors, General Dynamics, McDonnell Douglas and North American Aircraft, are working on the X-30 airframe. Six designs are being considered. One gives the two pilots a rotating cockpit that provides a view at low speeds but recesses into the fuselage for high speeds; another depends on pop-up mirrors.

What is the ultimate goal? Barthelemy suggests that hypersonic commercial transport might be an early spin-off. Others disagree. "The earth isn't big enough for anything over Mach 5," says James P. Loomis of the Battelle Memorial Institute's Center for High Speed Commercial Flight. McDonnell Douglas is studying a Mach 5-range aircraft—unlikely to be in service before the year 2015—and S. Fred Singer, chief scientist at the Department of Transportation, has proposed unmanned hypersonic freight aircraft, but neither concept would use scramjets. Nevertheless, materials developed for the X-30 could improve the

The fundamental question is whether air-breathing engines can reach orbital velocity



***COMPUTER SIMULATION** of airflow around a proposed X-30 aircraft flying at Mach 12.4 shows the formation of compression zones. Highest pressures are magenta, lowest are blue. The work was done at NASA's Ames Research Center.*

efficiency of low-supersonic commercial flight, according to Michael L. Henderson of the Boeing Company.

Whatever the fate of civil-aviation applications, an aircraft based on X-30 technology might provide assured access to space from a variety of sites. Such an aircraft could tend and even launch satellites, Barthelemy says; the Strategic Defense Initiative Organization is among those interested. *G* forces on the pilot would limit its use for interception, but some people see a role as a strategic bomber, antisatellite-weapons carrier or spy plane. Tokyo, it seems, will remain 14 hours away—for now. —*Tim Beardsley*

Yucca Mountain, Nevada

Can it contain our deadliest nuclear waste for 10,000 years?

On a sunny afternoon in early April, William W. Dudley of the U.S. Geological Survey stands on the crest of Yucca Mountain, a six-mile-long ridge that rises 1,000 feet above a remote Nevada desert. As he points out geologic landmarks to a visitor, Dudley squints. Perhaps it is the wind, or perhaps the effort of trying to peer 10,000 years into the future. Could the fault underlying the valley to the east rupture and trigger a major earthquake in that period? Might the cinder cone to the south erupt again? Dudley and other Government scientists must answer these questions before Yucca Mountain can become the burial site for the most poisonous issue of the nuclear age.

Called high-level nuclear waste, it consists of spent fuel from reactors and of solid by-products from the manufacture of nuclear explosives. A major component of the waste is plutonium, an extremely toxic and long-lived radionuclide (its half-life is about 24,000 years). More than 20,000 metric tons of the literally hot waste have accumulated at power and weapons plants around the U.S., and the amount is increasing by 2,000 metric tons annually.

In 1983 Congress charged the Department of Energy with selecting a site where the waste could be stored underground. By last year the DOE had narrowed the search to three candidates: Yucca Mountain, which lies on the western border of the Nevada Test Site, where the DOE tests nuclear weapons underground, and sites in Texas and Washington. The repository faced stiff political opposition in all three states. In December, Congress abruptly eliminated the Texas and Washington sites from further consideration and designated Yucca Mountain the sole candidate for further feasibility studies. Nevada officials, including the governor and members of Congress, contend the site was singled out not for scientific reasons but because the state is politically weak. Defenders of the congressional decision cite the high cost of studying all three sites and the fact that DOE officials have long expressed a preference for Yucca Mountain.

The safety standards the repository must meet are established by the Environmental Protection Agency and the Nuclear Regulatory Commission. In essence they require that the waste be prevented from contaminating the environment for 10,000 years. Carl P. Gertz, manager of the DOE's waste-management program, notes that the repository's designers must consider the possibility that American society might not survive undisrupted (able to provide maintenance and security) for that period. "We assume continuity for about 300 years," he says.

Over the next seven years the DOE, with the help of Geological Survey specialists such as Dudley, will study the Yucca Mountain site. If DOE officials conclude that the site is suitable—and if the Nuclear Regulatory Commission, which must grant the construction license, agrees—by 1998 workers will begin drilling a 1,500-acre grid of tunnels in the heart of the mountain. Five years later workers will start placing canisters filled with waste in the tunnels. After the repository is filled to capacity (about 70,000 metric tons), perhaps 50 years from now, it will be permanently sealed.

The canisters are expected to contain the waste for 300 to 1,000 years; then Yucca Mountain itself must do the job. At least superficially the mountain seems capable. It consists of a more or less solid chunk of tuff, or compacted ash, from volcanic eruptions more than 10 million years ago. The mountain is remote, about a dozen miles from the nearest settlement, and dry. Only six inches of water falls on it a year, and the water table lies 2,500 feet below its crest. The repository would be built in the so-called unsaturated zone, about 1,000 feet above the water table. These features allay concerns that water will penetrate the repository and carry radionuclides toward nearby populations.

Still, there are worrisome features. Some were identified years ago, such as the minor faults, a few of them seismically active, that lace the region. Other problems have just come to light. Bruce M. Crowe of the Los Alamos National Laboratory determined last year that a cinder cone 15 miles south of Yucca Mountain has erupted many times, perhaps as recently as 5,000 years ago. He says another eruption of the cinder cone could indirectly damage the repository by disrupting groundwater in the region.

A potentially more serious problem has been identified by one of the DOE's own geologists in Nevada, Jerry S. Szymanski. In a report made public this year, he suggests that tectonic pressure under Yucca Mountain could push the water table higher and also force water considerably above the table through faults. Calcite-silica deposits found in faults in the region could be the residue of such geothermal upwellings, according to Szymanski. Other Government geologists dispute his findings, saying the deposits resulted from water that seeped into the faults from the surface. Szymanski maintains that unless the issue is resolved "it would be reckless to go ahead." Gertz is confident more research will dispel such concerns. "We're 99 percent sure—well, make that 95—that Yucca Mountain will meet the regulatory requirements," he says. What happens if it does not? "We tell Congress and they decide what to do next," he replies. —*John Horgan*

"Slickums"

Sea-launched cruise missiles elude arms controllers

Experts call them "slickums." The suggestion of something slippery—and sinister—is appropriate. There may be no more ideal spearhead for a surprise attack on the U.S., whose capital and other major military and industrial centers are near open waters, than sea-launched cruise missiles (SLCM's). Yet, ironically, there are no weapons that the U.S. is more determined to exclude from a strategic-arms-reduction-talks (START) agreement with the Soviet Union, which President Reagan hopes to achieve before his term expires.

For more than a decade the U.S.S.R., which has lagged behind the U.S. in cruise-missile technology, has sought bilateral controls on these stubby-winged "flying torpedoes." The Soviets' persistence has recently paid off, in part. The intermediate nuclear forces (INF) treaty, signed by Reagan and Secretary Mikhail S. Gorbachev in December and now awaiting Senate

TOYOTA CAMRY V6

V FOR VROOM.
153 horsepower, four cams, 24 valves make for fast company. The new V6-equipped Camry Sedans or Wagons own the passing lane to a new world of driving pleasure.

NEW CAMRY V6 POWER.

PASSES ALL EXPECTATIONS.

UNSURPASSED COMFORT.
Behind the power: plush, satisfying comfort for five in a stylish, contemporary interior environment. Optional automatic transmission available.

The new Toyota Camry V6 will blur recent memories of ordinary sedans . . . especially if they're in the right-hand lane. A 25% boost in horsepower speeds up the process of putting fun behind four doors . . . putting traffic, hills and miles behind you. Down the road, you'll find Camry's resale value* as impressive as its power.

You might expect such good fortune to cost a small one, yet it's the *lowest-priced V6 import car* going.** And it arrives with all the reliability, comfort and sleekness you'd expect from a car bearing the Camry name—it just arrives more swiftly these days.

Camry Deluxe V6 Sedan. Not just some passing fancy.

Get More From Life . . . Buckle Up!

*The 1986 Toyota Camry 4-Door Deluxe Sedan retained 95.6% of its value, based on retail values listed in the March-April, 1988 *Kelley Blue Book Auto Market Report* for used car values, Western Edition. Actual resale value may vary, depending on optional equipment, geographic location, condition of vehicle and mileage.

**Based on manufacturers' suggested retail prices, *Kelley Blue Book New Car Price Manual* 4th Edition, 1988.

TOYOTA QUALITY
WHO COULD ASK FOR ANYTHING MORE!

ratification, banned ground-launched cruise missiles outright. U.S. and Soviet START negotiators have also agreed to limit air-launched cruise missiles, which are carried on bombers.

But the U.S. has kept SLCM's off the negotiating table. The main reason is that the Navy sees great potential in long-range SLCM's armed with nonnuclear, conventional warheads. With the modern Tomahawk SLCM, which navigates by comparing radar or optical scans of the earth with a map in its memory, American ships and submarines can attack targets more than 1,500 miles away with deadly accuracy; by 1990 the Navy plans to deploy some 4,000 Tomahawks, only one-fifth of them nuclear. U.S. officials will not consider limiting conventional SLCM's, and they maintain that an agreement limiting only nuclear SLCM's would be extremely difficult to verify. The problem is that nuclear warheads can be substituted for conventional warheads without altering the external appearance of the missile.

Schemes for verifying SLCM limits have been proposed. A typical scheme calls for outfitting nuclear SLCM's with irreproducible tags and conventional SLCM's with tamperproof seals. Each side would install inspectors at the other's assembly plants to ensure that seals and tags were properly emplaced; inspectors would also periodically search ships, submarines and storage sites for unsealed or untagged SLCM's. Plans requiring fewer or less intrusive inspections have also been advanced. Peter D. Zimmerman and Alton Frye of the Council on Foreign Relations suggest sealing SLCM's with electronic locks that, if tampered with, would automatically disable the missiles. This method, they maintain, would deter conventional-to-nuclear conversion while eliminating the need to inspect ships and submarines. Soviet officials have proposed keeping count of nuclear SLCM's on ships and submarines remotely, by means of helicopters or planes equipped with radiation sensors. The Soviets have offered to conduct an experiment with the U.S. to test the concept.

The U.S. has spurned the Soviet offer, arguing that the plan could be defeated simply by shielding the nuclear warheads. All the verification concepts proposed so far are flawed, according to Administration officials. None of the schemes, they contend, would prevent the Soviets from building nuclear SLCM's at a secret facility and deploying them in a crisis. Peter V. McFadden, a Department of State spokesman, says the Administration is still "open" to verification proposals, but he acknowledges that "the question of SLCM's is certainly something we wouldn't mind dropping." Because SLCM's fly at subsonic speeds, he adds, they pose less of a first-strike threat—and so are less important to control—than ballistic missiles.

Yet in several respects SLCM's may be more suited for a surprise attack than ballistic missiles, according to Milo D. Nordyke, who directs verification research at the Lawrence Livermore National Laboratory. Since their jet engines emit relatively little heat, cruise missiles are more likely to escape detection by infrared sensors on early-warning satellites. Because they are so small and fly close to the ground, they can slip more easily through radar defenses. From ranges within several hundred miles, SLCM's can even strike more quickly than ballistic missiles, which must follow a higher trajectory. By the mid-1990's the Navy may deploy SLCM's that are supersonic and "stealthy" (difficult to detect with radar) and consequently still more threatening. The Soviets are expected to do the same. Although U.S. SLCM's have heretofore exceeded their Soviet counterparts in both number and sophistication, the gap has been closing steadily. The U.S.S.R. is now developing SLCM's whose range and accuracy rivals the Tomahawk's, and it could deploy more than 1,500 nuclear SLCM's by 1995, according to the Arms Control Association.

The next generation of the ground-hugging missiles will be "stealthy" and will fly at supersonic speeds

TOMAHAWK CRUISE MISSILE launched from a submerged U.S. submarine more than 400 miles away explodes above an aircraft on San Clemente Island, Calif.

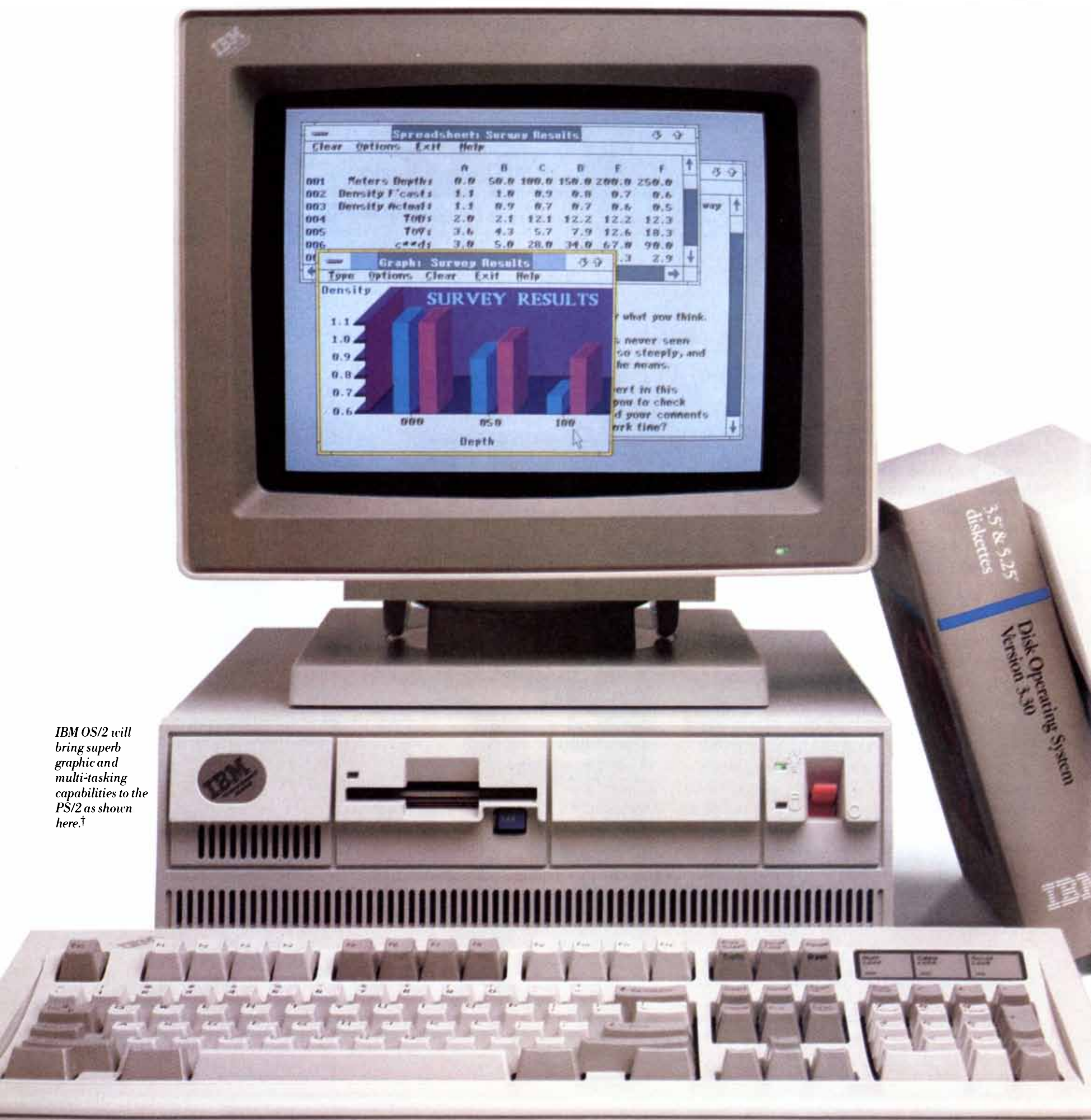
Nordyke agrees with Administration officials who argue that verifying limits on SLCM's would be difficult, but he disagrees with those who conclude that SLCM's need not be controlled. Pointing out that the weapons represent a "much greater threat to the U.S.," surrounded as it is by water, than to the more landlocked Soviet Union, he suggests that all SLCM's—conventional as well as nuclear—be banned. A total ban would be quite easy to verify; in order to violate it the U.S.S.R. would have to manufacture, test, store and deploy the missiles in secret. The question the U.S. faces, Nordyke says, is simple: Does the Navy's wish to keep conventional SLCM's outweigh the long-term benefits to U.S. security of eliminating nuclear SLCM's? —*J.H.*

Secret Science

Academia and industry fight expansion of secrecy laws

Is the Freedom of Information Act (FOIA) the best tool in the kit of the high-tech spy? The Administration apparently thinks so. Recent-

The best way to leap through next Monday



IBM OS/2 will bring superb graphic and multi-tasking capabilities to the PS/2 as shown here.†

ly it has sought two new restrictions on the act, citing concerns about economic competitiveness as well as national security.

The Administration is afraid that even unclassified technical information might be militarily or commercially valuable when pieced together, particularly when computer-search techniques are used. Officials have therefore expanded classification criteria and over the past few years have often invoked export regulations to keep some unclassified information from foreign eyes.

Such measures are chilling professional communication, a study suggests. Some technical societies have informally banned foreigners from their meetings to avoid problems with export regulations. According to a recent study of Government information controls by John Shattuck and Muriel Morisey Spence of Harvard University, "the evidence is mounting that the end result is likely to be more damaging than the evil it is intended to cure."

In February, in proposing a Superconductivity Competitiveness Act aimed at exploiting high-temperature superconductors, President Reagan asked for an exemption to the FOIA for any information produced in Government laboratories that could reasonably be expected to harm the economic competitiveness of the U.S. A month later a parade of witnesses from academia and industry argued against the measure before a Senate subcommittee. Robert L. Park of the American Physical Society pointed out that rapid communication had enabled American researchers to quickly duplicate Japanese strides in high-temperature superconductivity earlier this year. He noted that U.S. scientists would have been held back if Japan had a Superconductivity Competitiveness Act requiring bureaucrats to review data for commercial value. Park said an informal survey of research directors of major companies such as AT&T, General Motors, IBM and Westinghouse indicated that they opposed the proposed exemption: "Without exception they felt it would do more harm than good."

Evidence that the FOIA is widely exploited by foreign agents to obtain sensitive technical information seems sparse. William R. Graham, the president's science adviser, told the subcommittee that workers in Government laboratories have on occasion had to turn their notebooks over to foreign companies because of FOIA requests, but he could not detail such instances. The subcommittee chairman, Senator Patrick Leahy, castigated the Administration for claiming that Japan had obtained data on the space shuttle through an FOIA request that saved it "hundreds of millions of dollars and years of research." Leahy said his inquiries showed there never was such a request.

The uniform condemnation has probably relegated the FOIA exemption in the superconductivity proposals to legislative limbo. A second broad provision, however, recently proposed in legislation for the National Aeronautics and Space Administration, would allow the agency to withhold any data now subject to control by export regulations; at present only the Department of Defense has such authority. Similar legislation failed last year, but the Administration is persevering. *—T.M.B.*

PHYSICAL SCIENCES

Ice Storm

New observations are held to support a controversial theory

Two years ago Louis A. Frank, a respected physicist at the University of Iowa, ignited a fierce scientific dispute. He and his colleagues John B. Sigwarth and John D. Craven maintained that millions of invisible cometlike objects consisting mainly of ice hit the earth's upper atmosphere and vaporize there every year. Frank and his colleagues faced withering criticism. Investigators were therefore astounded when at the end of March a physicist at the Jet Propulsion Laboratory in Pasadena announced in a press release that he had obtained telescope images of objects that seemed as numerous as the comets Frank had proposed.

Frank's own evidence consisted of satellite-based observations of ultraviolet light emitted by the earth's atmosphere. The observations indicate that every few seconds, somewhere on the globe, a dark spot hundreds of miles in diameter suddenly appears against the background ultraviolet emission—"like a fly on a TV screen," Frank says—and then dissolves within a minute or two. Frank interpreted the spots as sudden discharges of water vapor from the breakup of one of his supposed objects: the released water vapor would absorb the ultraviolet light being emitted below it. If Frank is right, the unseen objects could have brought in enough water vapor over the aeons to fill the oceans—and stunningly alter solar-system science.

Impossible, most investigators said. If icy comets were hitting the atmosphere every few seconds at speeds of 10 kilometers per second, as Frank supposed, there should be many more craters on the moon, where there is no atmosphere to stop the comets. Frank replied that his comets were too light and fluffy (a few hundred tons each, about 10 meters in diameter) to make craters. Why, critics asked, were the objects not visible to telescopes or radar? Frank had a ready reply: they were coated with coal black dust that made them too dark to see and did not reflect radar beams.

Then Thomas M. Donahue of the University of Michigan measured the amount of atomic hydrogen in interplanetary space. He found less than one would expect to escape from the objects Frank had proposed. He therefore proposed a different type of object, and he said there were 10 million times fewer of them than those described by Frank. Frank countered that the dust enveloping his comets seals in the hydrogen. Some people thought that Frank's successive refinements of the theory, in the absence of more hard data, were fatal to its credibility.

Late last year Clayne Yeates of JPL decided to look for Frank's objects directly, working with the sensitive "Spacewatch" telescope at the Kitt Peak National Observatory in Arizona. The device has a wide field of view and is equipped with an electronic charge-coupled detector rather than film. Yeates panned the telescope against the sky, seeking dark comets in the orbit Frank had proposed for them.

His detector recorded "streaks" that apparently represented objects corresponding closely in brightness, position and quantity to Frank's objects. He and his colleague Tom Gehrels of the University of Arizona are confident they have ruled out known phenomena as explanations. Gehrels notes that, as is the case for Frank's dark spots, more of the streaks than usual are seen at times of meteor showers. Gehrels now thinks the objects might be dry debris from extinct comets, and that such debris could account for Frank's dark spots as well as ice balls could. Further tests were under way in late April.

The observations made by Yeates and Gehrels have attracted widespread interest. Most commentators praise Yeates for making the observations, although some criticize him for having announced the results before they had been subjected to peer review. An unforeseen interpretation might yet mean Yeates and Gehrels have found nothing new, but both say they are convinced. *—T.M.B.*

Roll Over, Wolfgang?

New experiments seek violations of the Pauli exclusion principle

The Pauli exclusion principle, named for its author, the cantankerous Austrian physicist Wolfgang Pauli, is a keystone of modern physics. Indeed, without it physics, if not matter, would collapse. Physicists consider the principle to be airtight. But now two theorists at the University of Maryland at College Park have formulated a relativistic quantum field theory that could poke a small but detectable hole in Pauli's principle.

The exclusion principle is invoked to explain why electrons in an atom occupy a succession of "orbits" at progressively higher energy levels instead of tumbling en masse to the ground state. Pauli proved that two particles of a class called fermions (such as electrons, protons and neutrons) cannot occupy the same orbit if they have identical quantum numbers. Electrons have a quantum number called spin, which can be either up or down. Thus each orbit can accommodate up to two electrons as long as their spins point in opposite directions.

Last year two Russian physicists, A. Yu. Ignatiev and V. A. Kuzmin, published a simple model that allows two identical fermions to fill one state. Intrigued, Maryland's Oscar W. Greenberg and Rabindra N. Mohapatra decided to attempt a more complete theory based on that model. They published their results last year in *Physical Review Letters.* The new theory allows rare "paronic" states to exist, in which two identical fermions can occupy a state simultaneously, violating the exclusion principle.

If paronic states exist, it may be possible to observe them. In a paronic atom, for example, one of the orbits would be filled by two electrons whose spins are parallel. The parallel-spin electrons would interact differently from antiparallel-spin electrons, and so the energy level of a paronic orbit would not be the same as that of a normal orbit. This difference should show up as tiny aberrations in the radiation that is emitted when electrons jump between atomic orbits.

Daniel E. Kelleher of the National Bureau of Standards will be searching for the simplest of such atoms, paronic helium. The spin makes the electron behave like a bar magnet, giving it a magnetic moment. The two electrons of a helium atom have antiparallel spins and so the magnetic moments cancel. In paronic helium, however, the spins are parallel and the magnetic moments add up. Kelleher will use a magnet to pull such atoms out of liquid helium. Laser spectroscopy would then reveal whether any of the selected atoms exhibit the ground-state energy of paronic helium, which differs by one part in 15,000 from that of ordinary helium. Robert L. Park, Erik J. Ramberg and Richard L. Talaga at Maryland propose two separate searches in other elements.

It will be some years before results are in. Meanwhile Greenberg and Mohapatra are examining the consequences of their theory. In addition to the exclusion principle, the new theory violates the CPT theorem, a pillar of quantum field theory. The situation might be salvaged, Greenberg says, by introducing the extra space dimensions that string theorists say curl around each particle. Greenberg speculates that CPT and the exclusion principle may be violated only in ordinary spacetime and may be conserved in the spacetime that includes these hidden dimensions. If the idea pans out,

Can a car quiet the wind? In a sense, that is precisely what the new Honda Accord LXi is designed to do.

The shape is aerodynamic. Clean all around. Windshield, windows and door handles are flush. Headlights are retractable.

The stance is low, over a newly refined double wishbone suspension system. Front and rear stabilizer bars are now larger. Stronger. So handling is more responsive.

Overall, the design provides an effective sound barrier against wind and road noise. And while you may be traveling on earth at something less than mach speed, you'll enjoy the simple comfort of quiet.

The experts at *Motor Trend* concur.

"Cruising the high-banked oval at top speed, passengers...carried on hushed


ound of speed.

he says, the experiments could provide the long-awaited test of string theory. —*June Kinoshita*

Quantum Holonomy

Phase shifts track a quantum system in "state space"

To understand the latest surprise served up by the quantum world take a pencil, lay it on the north pole of a globe and point it in the direction of any line of longitude that radiates from the pole. Move the pencil down along the line to the equator and, keeping it perpendicular to the equator, move it along the equator to another longitude. Move the pencil back to the north pole along the new longitude and you will find that although the pencil has returned to its starting spot, it no longer points along the original line of longitude.

This is an example of a purely geometric effect, known as holonomy, resulting from the fact that the pencil was forced to trace out a circuit on the surface of a sphere while remaining parallel to the meridians. The holonomy caused by such "parallel transport" around a circuit on a curved surface is not limited to tangible objects such as globes and pencils. The results of three different experiments recently reported in the same issue of *Physical Review Letters* show that geometric holonomy can exist even for abstract constructs in the microscopic realm of quantum physics.

In the quantum realm the state of a physical system is best regarded as a wave whose characteristics are determined by parameters, or physical quantities, that may affect the system. In the case of a photon such a parameter might be, say, its polarization (the direction of its associated electric field) or its intrinsic spin. Another important component of quantum waves is their phase: the positions of each wave's crests and troughs in relation to one another.

As was first pointed out in 1983 by Michael V. Berry of the University of Bristol, as a result of geometric holonomy a quantum system can exhibit different phases in its initial and final wave representations even though the system begins and ends with the same parameter values. The key to understanding how this can happen is to visualize all possible states of the system as points on the surface of a sphere in "state space." Because the initial and final states of the system are represented by the same point on the sphere, the intermediate states lie on a closed curve on the sphere.

The phase of the wave representing the system can then be envisioned as undergoing parallel transport (like the pencil on the globe) as the system goes from state to state around the curve, completing a circuit. As a consequence, when the system returns to its starting point it no longer has its original phase. In fact, the magnitude of the change in phase reveals the general form of the curve, because it is proportional to the area on the state-space sphere enclosed by the curve.

Rajendra Bhandari and Joseph Samuel of the Raman Research Institute in India succeeded in measuring such a phase shift in an interference pattern produced by a laser beam that was split and recombined. The shift was caused by varying the polarization state of the photons in one of the split beams while ensuring that the beam's photons began and ended in the same polarization state as those in the other split beam. A team headed by Raymond Y. Chiao of the University of California at Berkeley and Howard Nathel of the Lawrence Livermore National Laboratory tried altering the direction of the photons' momentum rather than their polarization. The team observed the predicted phase shifts even though the photons traced out a circuit in a different state space.

The third experiment, also done by a group at Berkeley, detected a quantum holonomy not in laser photons but in the radio signals emitted by the spins of atomic nuclei. Applying a variation of nuclear-magnetic-resonance interferometry, Dieter Suter, Karl T. Mueller and Alexander Pines worked with molecules consisting of atoms whose nuclear spins are aligned. They subjected the atoms to magnetic pulses that caused the nuclei to change their spin state and then return to their original state. As expected, the radio signals later emitted by the nuclei were phase-shifted. —*Gregory Greenwell*

BIOLOGICAL SCIENCES

Earlier Americans

The Americas may have been populated 30,000 years ago

When did human beings first come to the Americas? The conventional answer is that they came across a land bridge from Siberia only a little more than 10,000 years ago. Findings at several sites in North and South America, however, suggest that the first settlers may have come long before then. For example, a site called Monte Verde in Chile has produced radiocarbon dates of about 13,000 years ago. From a deeper layer at Monte Verde excavators have now retrieved material yielding dates of more than 30,000 years before the present.

The new dates at Monte Verde come from a pair of clay-lined depressions found in a layer of gray sand about two meters below the modern surface. According to the excavation's leader, Tom D. Dillehay of the University of Kentucky, the two clay-lined pits were probably hearths. From each of them enough charcoal or charred wood was retrieved to provide a radiocarbon date. One date was about 33,000 years before the present; the other was between 33,000 and 40,000.

Strewn near the two hearths (and a third pit that may be the remains of a scattered fire) were an assortment of 26 stones that had been modified for human use. Some had been flaked into recognizable tools, others were little more than flakes or battered cobbles. A microscopic analysis of the edges of the stones is now under way, Dillehay said; preliminary results indicate that some of the stones were prob-

ably employed for processing plants.

Dillehay added that one reason archaeologists have been slow to accept the concept of human presence in the Americas well before 10,000 years ago is that they expect elaborate and conspicuous sites replete with the remains of big-game kills and many stone tools of well-recognized types—sites like those of the Clovis culture that appeared in many areas of the Americas about 11,500 years ago. "Perhaps instead of elaborate sites we should be looking for traces of a more opportunistic lifeway: small groups of people hunting and collecting a few plants, making a few tools, then leaving," he said.

Although the debate about when the New World was settled is far from over, two sites have now yielded radiocarbon dates of more than 30,000 years. They are Monte Verde and Boqueirão da Pedra Furada in Brazil. Summing up his latest results, Dillehay said: "I find it very interesting that two independent studies in South America have now provided radiocarbon dates slightly greater than 30,000 years before the present."

"Given the skeptical atmosphere that prevails on this subject within the profession of archaeology," he continued, "I would have been more skeptical had there been only one such site, even if it were mine. Having two sites doesn't prove anything, but it certainly makes a much stronger case for an occupation of the Americas long before 15,000 years ago." *—John Benditt*

Compound Compound Eye

A crab eye does it with mirrors—and lenses too

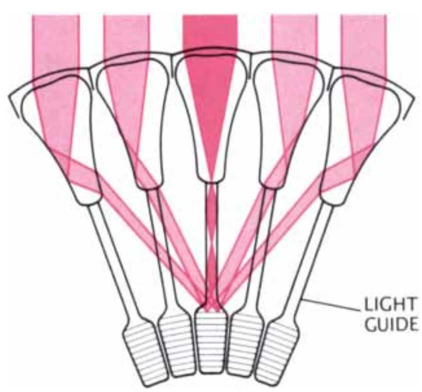
Many insects, crustaceans and other arthropods see through compound eyes. In some species the individual facets—each of them a separate optical unit—work independently, like an array of tiny telescopes aimed at different points. In the so-called superposition eyes of certain nocturnal moths and crustaceans, however, cooperation increases the eye's sensitivity: many facets combine light from the same point in the visual world to form a single image on the retina.

It now seems that many crabs and hermit crabs have compound eyes that work both ways. The price of this flexibility, writes Dan-Erik Nilsson of the University of Lund, who describes the eye in *Nature,* is a mechanism of "impressive complexity," made up of "a remarkable combination of ordinary lenses, cylindrical lenses, parabolic mirrors and light guides."

Two kinds of superposition eyes were known when Nilsson began his work. The compound eyes of shrimps and crayfishes consist of elements that are square in cross section and have flat, reflective sides, so that, like a billiard ball ricocheting out of a corner, light from a given point is deflected by each element toward the same focus on the retina. In the eyes of moths and krill light is redirected not by reflection but by refraction within each element's crystalline substance.

Examining some preserved sections of the eye from a species of crab, Nilsson noted certain features typical of superposition eyes. Yet the imaging system turned out to be entirely new. In this crab and many others, light entering each element of the compound eye is first refracted by a lens at the cornea, then reflected by the side of the element and finally refracted again by the crystalline material before it proceeds to the retina.

As in the shrimp and crayfish eyes, the reflecting border of each element serves to direct the light toward the common focus. It also has an additional function. The corneal lens causes parallel rays from a distant source to converge. As the light strikes the reflecting border the convergence is corrected, by virtue of the inside surface's convex, parabolic profile. Seen from above, however, the reflecting surface is not parabolic and convex but circular and concave. The circular section has a focusing effect that is counteracted by refraction in the element's core, which acts as a cylindrical lens.



OPTICAL ELEMENTS of a crab eye are shown in profile. Light entering each element is first refracted by a lens at the cornea and then reflected by the inner surface toward a common focus on the retina. In the element pointing directly at the light source the beam* (red) *is conducted to the retina by a light guide.

What does the crab gain by all this complexity? The answer may lie in the eye's ability to adapt to changes in lighting. In the dark many elements pool their light, but in daylight pigment floods the clear space between the optical array and the retina, blocking the converging beams. Yet the eye still functions, at lowered sensitivity: each element is still sensitive to light in its own direction. Such light misses the reflecting border and hence is not deflected into the clear zone. Instead the corneal lens focuses it onto the base of the element. There one final component, a thin light guide, carries the light across the pigmented zone to the retina. *—Tim Appenzeller*

Love on the Fly

The courtship songs of Hawaiian fruit flies

Frogs, crickets—and even cicadas—are well known for their singing. Now it seems that fruit flies too can carry a romantic tune. When a male fruit fly meets a female, he responds with a courtship song. That in itself is not so unusual; what really sets fruit flies apart from other animals is their diverse repertoire.

The discovery is reported in *Science* by Ronald R. Hoy of Cornell University, Anneli Hoikkala of the University of Oulu in Finland and Kenneth Kaneshiro of the University of Hawaii.

These investigators recorded the songs of three species of *Drosophila* and analyzed them on an oscilloscope. They found that the songs could be classified into four distinct acoustic types: click trains like those produced by cicadas, complex pulse trains like those of crickets, simple pulse trains and simple tone songs. Moreover, the way in which these sounds are produced is highly varied. Click trains, made by the males of *D. fasciculisetae,* consist of repeated short bursts of high-frequency pulses (six kilohertz on the average) produced by vibrations at the base of the wing. Similar high-frequency patterns have been recorded from cicadas and katydids but had not been known to occur in flies.

The complex pulse trains recorded from *D. cyrtoloma* consist of from two to seven cyclic pulses followed by a long trill. They are rhythmically complex, made by the male as he beats his wings in two-part rhythm. Simple pulse trains and simple tone songs are both made by males of *D. silvestris.* A simple pulse train is like a purr, consisting of a train of cyclic sound pul-

ses produced by vibrations of the abdomen. A simple tone song is produced by rapid wing vibration while the male is standing with his head under the wings of the female.

Hoy and his team discovered these four song categories by studying just 20 species, all of them belonging to a subset of fruit flies called the picture-winged drosophilids that is represented in Hawaii by a total of 106 species. The fact that these flies are prevalent in Hawaii is not insignificant. Nowhere in the world are fruit flies more taxonomically diverse than in that island archipelago, which hosts a whopping 500 species. (Only 1,000 to 2,000 species exist worldwide.)

How can such extraordinary levels of diversity be explained? Volcanic activity created the Hawaiian archipelago some five to six million years ago; since that time few kinds of animals have successfully traversed the 2,000 miles of open ocean separating the islands from North America, the nearest continental landmass. Those animals that did survive the journey (mostly winged forms) found and filled countless unoccupied niches, diverging and evolving into taxonomically distinct species. All the 500 fruit fly species now present in the Hawaiian archipelago, for example, are thought to be the descendants of one or two females. —*Laurie Burnham*

For the Birds

Genetic fingerprinting may help to rescue imperiled wildlife

A powerful new technique in biotechnology could play a key role in two ecological dramas. One involves the whooping crane: there are now only 135 of these elegant birds gracing North America's wetlands, more than at their nadir of 18 in 1938 but not enough to survive without help. The other is that of the California condor, possibly the continent's rarest bird: the species is now down to only 27 individuals, all (after a controversial court decision) in captivity. Both kinds of bird are victims of human disruption of their habitats; both will escape extinction only if their populations can be bolstered with birds bred in captivity. And there is the rub.

If closely related birds mate, as is likely in a small population, chances increase that the offspring will inherit many of the same genes from each parent. Such individuals are genetically disadvantaged. The reason is that harmful genes, which are usually masked by normal ones, are likely to be present in double dosage. The result is that fewer chicks hatch, and those that do are less likely to survive. It is hard to avoid such matings in a population taken from the wild because it is difficult to know which individuals are related.

Now biotechnology provides a way to measure relatedness between prospective parents. The technique is genetic "fingerprinting," which was pioneered in human beings by Alec J. Jeffreys of the University of Leicester in England. Jeffreys discovered that everyone has within his or her genetic material "hypervariable" DNA differing so much from one person to another that each individual's pattern is unique. When the DNA is digested with enzymes and separated on a gel, the pattern is visualized as a series of bands. Because each band has been inherited from a parent, shared bands serve to demonstrate ancestry. The method is reliable enough for genetic fingerprinting to have been accepted recently as evidence in criminal cases; it has also been used to determine paternity in zoo primates.

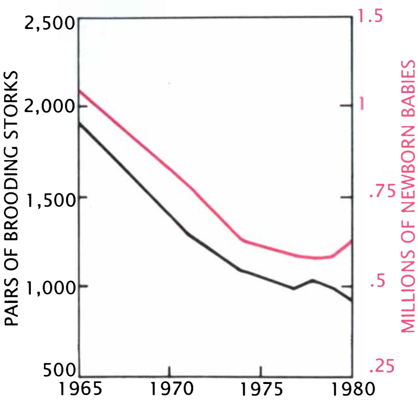
Jonathan Longmire of the Los Alamos National Laboratory is now applying a genetic-fingerprinting technique developed by Gilbert Vassart of the Free University of Brussels to examine the relatedness of 41 whooping cranes at the Patuxent Wildlife Research Center in Laurel, Md. Oliver A. Ryder of the Center for Reproduction of Endangered Species at the San Diego Zoo is using the same technique to assist in breeding captive condors. The analyses, which can be done from blood samples, are not yet complete, but both Longmire and Ryder say the initial tests have been encouraging.

Yet the story is not finished. Whooping crane eggs laid in captivity are taken to foster parents of a related species in the wild to be hatched; the idea is to allow the few wild whoopers to concentrate on looking after their own eggs. George F. Gee, who heads the whooping crane project at Patuxent, notes that doubts persist about whether birds reared by foster parents will breed successfully. Stephen J. O'Brien, another investigator who has worked with genetic fingerprinting, cautions that there are dangers of interpretation in species where little is known about natural variation. Longmire holds that even if the technique is not completely accurate, the data should help to avoid pairings between closely related animals and so help both species to regain a perch in the natural environment. —*T.M.B.*

Vector Analysis

Lack of carriers blamed for a falling birthrate

In a letter to *Nature,* Helmut Sies of the University of Dusseldorf presents two curves suggesting an explanation "that every child knows makes sense" for a long decline in West Germany's birthrate:



MEDICINE

Bethesda Brain Drain

Health bureaucrats try to stem an exodus of top NIH workers

How much can a topflight biomedical investigator expect to be paid? Remuneration is causing problems at the National Institutes of Health campus in Bethesda, Md., the largest biomedical research center in the world. Hampered by Government pay scales that compare poorly with those in universities or industry, the NIH is finding it increasingly difficult to retain senior research staff.

The problem was put in the spotlight late last year. An internal Office of Management and Budget paper revealed that the OMB was considering, as part of its tireless quest for smaller government, whether to turn the operation of the Bethesda campus over to the private sector, to be run like an independent university. In partial justification the OMB cited the restlessness of leading investigators. Following the predictable outcry, the OMB is now distancing itself from the privatization idea. Yet concern about the hemorrhage of top staff persists.

At stake is the preeminence of one of the world's most valuable intellectual resources. Fundamental scientific achievements by NIH investigators are legion: in all, the NIH supports nearly 20,000 scientists in some 1,300 institutions, and more than 3,000 work for the NIH directly.

The senior investigators at Bethesda are the heart of the enterprise, and the figures bear out the concern that the NIH is loosing them at an alarming rate: there are now 27 percent fewer top-ranking civilian investigators at the NIH than there were in 1978. Most of the stars who have left recently for jobs elsewhere at least doubled their salary; many are also attracted by more laboratory space and better equipment. Minds were concentrated on the problem when word spread that Robert C. Gallo, perhaps the leading AIDS researcher in the U.S., was considering another venue.

Basic salaries for the highest-ranking civilians at the NIH are about $85,000 per year; Public Health Service officers can earn somewhat more. Outside work is allowed to supplement basic pay by as much as $50,000 per year, although few take full advantage of this license. A department chairman in a medical school can, in contrast, earn from $200,000 to $300,000 depending on specialty.

In response to this competition a Senate committee in April proposed to increase spending for NIH facilities and approved a plan to establish an elite cadre of senior researchers under the rubric Senior Biomedical Scientific Service. In addition to higher salaries, those chosen would have more flexible retirement benefits. The plan, which would provide for investigators with a rank equivalent to full professor, is informally supported at the NIH but still faces legislative hurdles. Other options for the NIH are being weighed in a major study by the Institute of Medicine. Enabling the NIH to confer degrees is still a possibility, as is granting the director of the NIH the authority to fix staffing levels. *—T.M.B.*

Nature, Nurture and Death

A study of adoptees suggests there is no escaping heredity

There's a morbid old joke among population geneticists about the heritability of death: it seems that people whose parents die have a pretty high risk of dying themselves. A familial influence on the length of life has, at least, been established. Yet it is extremely difficult to assess how much of that influence can be attributed to shared genes and how much to a shared environment.

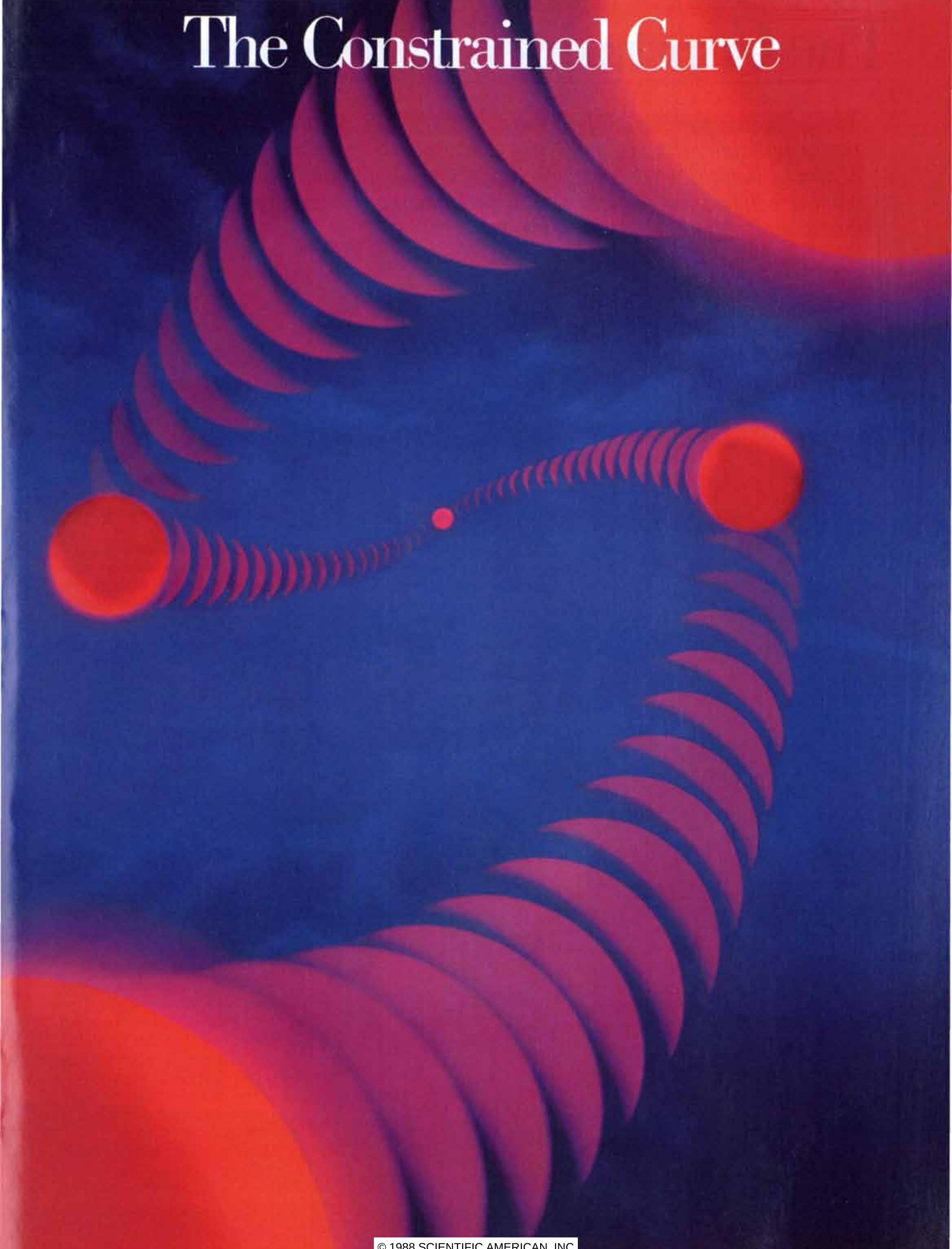
Almost 10 years ago a Danish clinician named Thorkild I. A. Sørensen decided to examine the histories of adopted children and their parents to find out more about the determinants of premature death—death before the age of 50. Sørensen wanted to know whether the premature death of an adoptee could be correlated with the premature death of either an adoptive or a biological parent. His results, which he and his colleagues at the Hvidovre University Hospital in Copenhagen published recently in the *New England Journal of Medicine,* comprise a remarkable body of evidence on the separate influences of heredity and environment on deaths from cardiovascular disease, infections and cancer.

Sørensen selected almost 1,000 children born between 1924 and 1926 from the Danish Adoption Registry, which records information on both the biological and the adoptive parents at the time of adoption. Through local population registries he traced the fate of the children and of their 4,000 parents up to 1982, gleaning ages and causes of death from death certificates. His efforts revealed a "strong genetic background" for premature death in adults.

If one of an adoptee's biological parents died prematurely of natural causes, that adoptee's risk of premature death from natural causes was double the risk of an adoptee whose biological parents had not died prematurely. An adoptee's chances of dying prematurely from cardiovascular disease were four times as high if one of his or her biological parents had died prematurely from that disease. And an adoptee's risk of dying from an infectious disease was five times higher if a biological parent had fallen victim to an infection. In contrast, the fate of adoptive parents had no correlation with premature death in adoptees, except in the case of cancer. Sørensen found that the death of an adoptive parent at an early age from cancer increased by five times an adoptee's risk of dying prematurely of cancer. No such connection emerged between biological parents and adoptees.

Sørensen's data have not surprised investigators studying the hereditary aspects of disease; a genetic predisposition to heart disease has long been suspected, and susceptibility of the immune system is also thought to have a genetic component. The effects of smoking, diet and other environmental triggers on cancer are notorious. Yet circumstances in which hereditary and environmental influences can be distinguished and quantified are scarce. The details of such circum-

The Constrained Curve



The Constrained Curve

The geometric path traced by a robot arm is independent of time. Now a mathematician at the General Motors Research Laboratories has devised a simple, innovative way to relate the path to time so that the machine can track the path and meet specific performance objectives without exceeding its physical operating limits.

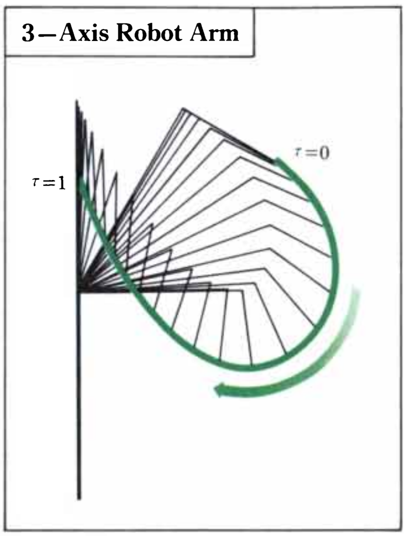


Figure 1: Schematic diagram of a 3-axis robot tracing a path in 3-space.

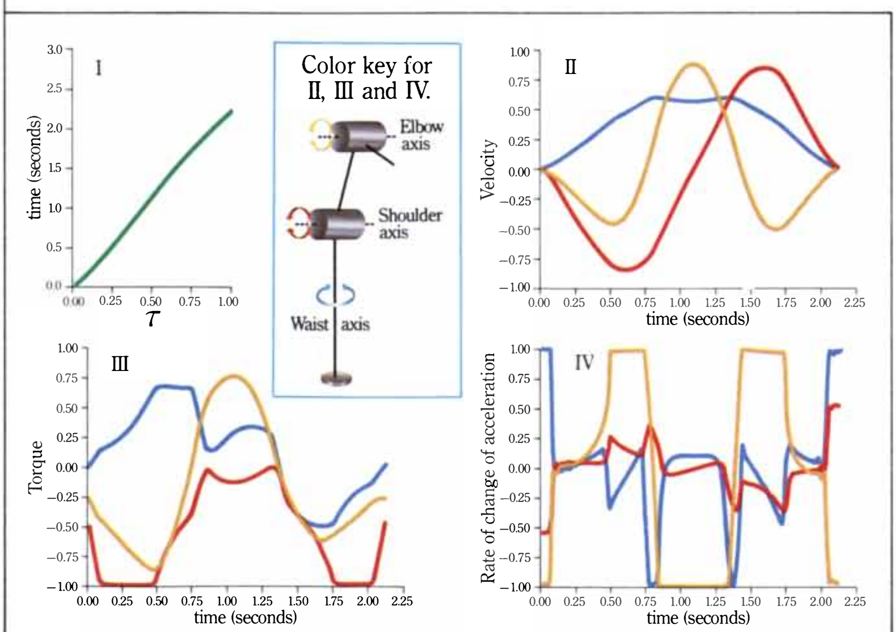
Figure 2: Results for Figure 1 path. I: Plot of the change of variables, $t=h(\tau)$. II, III, and IV: Normalized velocity, torque, and rate of change of acceleration for the waist, shoulder, and elbow (for any variable, a value of ±1 indicates operation at a limit).

Industrial robot arms are very good at repeating a well defined motion with a high degree of accuracy. A robot with a welding tool, a paint sprayer, or a grasping device at its tip can weld in the right spot, spray a precise pattern, or locate a part in a given place time after time.

This untiring precision makes robots valuable in a quality-oriented manufacturing process such as the assembly of an automobile. That's why General Motors has installed so many robotic manipulators in its plants, and why GM is intent on developing technology and software to use these machines to their best advantage.

When a robot is to apply sealant to a windshield opening, or move a part from one point to another, its tip is positioned at points along a fixed geometric path, always maintaining the orientation needed to perform the task.

Mathematically, tip position along the path can be described as a function of a one-dimensional position parameter τ that ranges from 0 to 1 as the path evolves from beginning to end. Actually, for a robot having three joints, Figure 1 for example, tip position is determined by a set of three functions of τ, one for each joint of the arm. Each separate joint function relates a specific angle of rotation, θ, about that joint axis to a given value of τ.



To get the robot to perform a task, however, its computer controller must associate each point on the path with some value of time—in effect telling the robot to be in position A at a certain time, position B at another time, and so on, throughout the path.

Establishing an appropriate correspondence between time and the path position parameter is an important prerequisite to actually controlling the robot to follow the path.

Dr. Samuel Marin, a mathematician at General Motors Research Laboratories, has devised an effective and efficient means of computing the required correspondence. His work addresses productivity concerns. Dr. Marin's objective is to make cycle time (the time it takes the robot to trace the path from beginning to end) as small as possible, yet to respect at all times the physical operating limits of the robot.

Dr. Marin noted that by seeking a correspondence that gives time explicitly in terms of the path position parameter, $t=h(\tau)$, the problem's character changes. It appears not so closely associated with control theory, where the problem has also been studied, but more like a problem of nonlinear optimization.

Setting $g(\tau)=h'(\tau)$, the derivative of h with respect to τ, allowed Dr. Marin to pose the minimum time problem in the following way: minimize $\int_0^1 g(\tau)\, d\tau$, subject to some constraints dictated by the physical operating limits of the robot mechanism. These limits on the robot—limits on velocity, acceleration or torque, and on rate of change of acceleration (Fig. 2)—can all be formulated as differential inequality constraints and are all expressible in terms of the unknown function $g(\tau)$, as: $g(\tau) \geq G(\tau, g, g', g'')\ \tau \in [0,1]$.

If the problem could be discretized, making it in some sense finite, it could be put on a computer and solved numerically. So Dr. Marin replaced the unknown function with a piecewise cubic approximation.

This allows the search for the unknown function to be confined to a class of functions that are completely characterized by a finite number of coefficients in a B-spline series.

He similarly discretized the constraints, replacing the infinite set of constraints with a finite dimensional subset that could be dealt with numerically.

He completed the formulation of the discrete problem by incorporating a grid-refinement strategy. Now the problem's dimension could be gradually increased to better approximate the continuous case.

What resulted was a classic nonlinear optimization problem, a finite dimensional problem in which it remained only to find the coefficients of the B-splines while satisfying the constraints.

A monotonicity property of this problem coupled with properties of the approximation method suggests that the simple technique of cyclic coordinate descent might best provide a solution.

"While not so effective in other applications, a cyclic coordinate descent-based algorithm appears to be exactly what is needed in this class of problems," notes Dr. Marin. "With modifications introduced to ensure that the iterates are strictly feasible, this method has consistently and rapidly solved the problem."

Working closely with mathematicians at Rensselaer Polytechnic Institute, Dr. Marin is confirming this method's utility. In comparisons so far with several widely used, general-purpose optimization codes, the special method consistently shows itself to be superior.

"My work in path parametrization is just part of the story here at GM," emphasizes Dr. Marin. "Many aspects of this problem's formulation are rooted in deeper concerns about how robots can be made to move faster and more accurately. These concerns originated in the work of Dr. Robert Goor, my colleague in the Mathematics Department, and have motivated several significant advances in robot control and trajectory planning.

"Until all the pieces are put together in a production system, it's difficult to gauge the full value of this work. However it will help reduce our manufacturing costs and will enhance our product quality."

General Motors

MARK OF EXCELLENCE

THE MAN BEHIND THE WORK

Dr. Sam Marin is a Senior Staff Research Scientist in the Mathematics Department of the General Motors Research Laboratories. He is also the Manager of the Department's Mathematical Analysis and Computation Section.

Dr. Marin received his undergraduate degree in mathematics from St. Vincent College in Latrobe, Pennsylvania, and holds both an M.S. and a Ph. D. in that discipline from Carnegie-Mellon University. Between graduate degrees, Sam was an officer in the U.S. Navy, teaching mathematics at the Naval Nuclear Power School.

Since joining General Motors in 1978, Dr. Marin has pursued interests in numerical analysis and approximation. He has published research relating these areas to a variety of applications, including robotics, geometric curve design, and acoustics.

Sam is a member of the Society for Industrial and Applied Mathematics. He lives in Rochester Hills, Michigan, with his wife and two children.

stances are often considered confidential in the U.S. In Denmark, Sørensen plans to expand his data base in order to explore the interaction between genetic and environmental factors. His next study will survey 24 years of adoption records and 14,000 adopted Danes. *—Karen Wright*

℞: Art

Drawing or sculpting can help traumatized Vietnam veterans

Although he sells his paintings and sculpture and has been commissioned to create a Vietnam war memorial, Steven R. Piscitelli, who lives in Amherst, Mass., does not consider himself a professional artist. He calls himself "a patient doing therapy." Creating art, he says, literally helps him to stay alive.

Piscitelli, who was wounded three times while serving as a marine in Vietnam in 1969 and 1970, suffers from post-traumatic stress disorder (PTSD). According to the Veterans Administration, about 375,000 Vietnam veterans—12 percent of all Americans who served in the war—have the disorder. Victims continually relive traumatic events in nightmares and flashbacks. Alternately depressed and panic-stricken, they often seek refuge in alcohol, drugs or suicide.

Piscitelli is one of a small number of veterans who have been treated for PTSD with art therapy, in which patients express themselves through drawings, sculpture and other visual media. In 1983, after he had tried to commit suicide, his parents took him to the V.A. Medical Center in Northampton, Mass. There he enrolled in an art-therapy program. At first he simply doodled; then he produced a flood of drawings and sculpture of "all these things that were haunting me": a soldier cradling a dead baby, another soldier whose hand has been blown off. "It was easy to draw a look of terror on a soldier's face," he recalls, "because it was right there in front of me." His sense of relief was immediate, he says. "Those images were frozen on paper, and so they left me."

Piscitelli's art therapist in 1983 and 1984 was Deborah R. Golub, now an assistant professor of education at Wright State University in Dayton, Ohio. She has worked with many victims of trauma, including Cambodian refugees and indigent children in Brazil. Although Piscitelli's talent and ongoing commitment to art is unusual, his positive response to art therapy is not, Golub says. She calls the therapy an "extraordinarily powerful" method both for helping the therapist understand the disorder and for helping the patient control his memories.

Elizabeth A. Brett, an associate professor of psychiatry at the Yale University School of Medicine who has done extensive research on veterans with PTSD, agrees. She points out that some veterans recall traumas in such vivid detail that they temporarily lose contact with reality. "In art therapy you are tapping in directly to that sensory experience," she remarks. "That's one reason why it can be very powerful."

David Read Johnson, a psychologist at the V.A. Medical Center in West Haven, Conn., and editor of the *International Journal of Arts in Psychotherapy,* says art therapy can be particularly beneficial for veterans just entering treatment. The artwork may simply reveal that the patient is in fact suffering from PTSD, which often goes undiagnosed. "Most vets don't want to talk about their experiences at first," Johnson explains. "Because a picture is static and outside you, it is safer." Once veterans have learned to express themselves through art, he notes, they are often better prepared for other potentially therapeutic forms of expression, including drama, poetry and group "rap sessions."

The West Haven center is one of only a few places that offer "creative" therapies to veterans, according to Johnson. Although there are about 2,500 art therapists in the U.S., he says, most are female and enter the field to help women traumatized by rape and other forms of physical abuse; the inability of the V.A. to pay adequate salaries to art therapists, who generally have at least a master's degree, poses another problem. Johnson thinks the growing recognition of the value of art therapy for veterans could counteract these problems.

CLAY SCULPTURE by Steven R. Piscitelli recalls injuries he sustained when he stepped on a land mine in Vietnam.

Even the most enthusiastic proponents of art therapy emphasize that it cannot "cure" victims of PTSD. Golub observes that "you can just help them live with the pain." Piscitelli feels he has come a long way since he first met Golub, when he had "no memory of anything" but combat. His art reflects his progress: he now sculpts ballerinas as well as wounded soldiers. Nevertheless, he still must counteract his psychic disorder by creating at least one "poisonous" work every day. "I listen to classical music," he says, "I love ballet and art, and yet I still can't go to sleep at night because the enemy is after me." *—J.H.*

RICHARD P. FEYNMAN
1918–1988

Millions of his fellow citizens saw Richard Feynman just once, but in an absolutely characteristic moment. There he was, a theoretical physicist unknown outside the world of science, visible on all the versions of the nightly news. He owed this attention to his appointment to the commission set up to study the tragic end of the space shuttle *Challenger*. Before your eyes Feynman chilled a small neoprene O ring in ice water to turn it hard and inflexible; thus one of the proximate causes of the disaster became a homely experience sifted out of the engineering complexities. Such was his lifelong delight: grasping the world by particular example, even if drawn from the most abstract of generalizations. He had no peer in this style, whether with O rings, the contemporary response to Isaac Newton's prism experiments, Mayan almanacs or the state vectors of fermions and bosons.

His varied talents extended beyond physics. He was also raconteur, artist, musician, dancer, gambler, puzzler, cipherer, locksmith, sometime chemist. He rolled with the good times at the carnival in Rio and in the casinos of Las Vegas. Most of the anecdotes that cluster around him are true!

I met Richard first in March of the hard wartime year of 1943, when he stopped at the University of Chicago on his way out to the laboratory just being established on the mesas of Los

Alamos. Perhaps his ingenuity and expertise would be of use to the group of young theorists lakeside, also engaged in the secret enterprises of the Manhattan Project. He was not yet 25, his highly original Ph.D. thesis was still unpublished, yet he brought with him an open reputation for unmatched quickness at the hard integrals that arise in mathematical physics. We all came to meet this brash champion of analysis, who for the first time was entering a circle of physicists wider than his long-impressed fellow students and teachers.

Feynman was patently not struck in the prewar mold of most young academics. He had the flowing, expressive postures of a dancer, the quick speech we thought of as Broadway, the pat phrases of the hustler and the conversational energy of a finger snapper. But he did not disappoint us; he explained on the spot how to gain a quick result that had evaded one of our clever calculators for a month. Such a display was a shallow way to judge a superb mind; still, it made a lasting impression. To this day physicists of the postwar generation recommend their best students with the mock-modest concession that he or she "is no Feynman, but...."

One of Feynman's last publications has just appeared, the text of a lecture given at the University of Cambridge in 1986 as a memorial to the great English physicist P. A. M. Dirac. It is another typical performance: it begins with a creditable drawing of Dirac that Feynman himself sketched in 1965, and it includes a photograph of the ebullient Feynman engaging the attentive but warily reserved Englishman. The lecture is a tour de force of exposition, seeking to derive a couple of the most important results of quantum field theory not "in the spirit of Dirac with lots of symbols and operators" but by explicit arguments that flow from those summary zigzag sketches known everywhere as Feynman diagrams. True to his way, Feynman explains that he will set out some "very simple examples...because if you do you will understand the generalities at once—that's the way *I* understand things anyhow."

Adept at formalism, inventor of more than one mathematical device of sweep and power, Richard Feynman nonetheless fought for an understanding that was explicit, concrete, without the constraints of jargon, usage, decorum and precedent. In that way he was like a poet, at home in metaphor and image, cheerfully exhibiting the relation between a dance figure and the rotation of a two-value amplitude. The Feynman diagrams began as a genuine aid to intricate formalism, since they clearly index the bewildering alternatives that beset any calculation of what happens among interacting particles. Once Feynman tried to convey the difficulties of such a calculation to someone utterly strange to such tasks: "You know how it is with daylight saving time? Well, physics has a dozen kinds of daylight saving."

A few examples of the diagrams are painted like pictographs on the family van that stands parked outside his suburban house near Caltech, where he taught and worked for so many years. The diagrams are a form of shorthand analysis, but they are far from a geometrization of the events. Their fidelity relies on careful and highly original mathematical rules that accompany each straight or wiggly line, usually generating at every vertex an exercise in matrix algebra; between intersections they imply a clever nesting of integrals that serve to sum over the spacetime excursions that intervene.

Certainly Feynman diagrams have become as useful to the quantum field theorist as circuit diagrams ever were to the electronics designer. Like circuit diagrams they are rather less than maps, rather more than logical outlines. The diagrams allow a transparent ordering that makes clear the intricate calculations that would have seemed inhumanly beyond masters of the same theory. The diagrams helped Feynman to gain the understanding that in his and other hands allowed quantum electrodynamics to be made into the most accurate of physical theories, even though it is plagued by deep inconsistencies, logical icebergs that can be skirted but not removed. Like quantum field theory itself, Feynman diagrams transcend any single set of forces and particles. They guide the computation of the intercourse of quarks and gluons, just as they have guided the computation of the electromagnetic processes for which Feynman devised them in the early postwar years.

That was by no means the end of his accomplishments in physics. Besides the insights concerning quantum electrodynamics for which he shared a Nobel prize, he made a strong mark at the right time on the nature of the weak interactions, on the conceptual wave functions that describe liquid helium and on the foundations of quantum mechanics. It was his encouragement and his fresh interpretations of electron-scattering experiments that signaled to all that free-moving subunits, his "partons," were deep within proton and neutron. That opened one main experimental path toward quarks.

When the feeling was abroad that the teaching of physics to beginners was growing stilted, Richard created a series of lectures that grew into a multivolume text. It is tough, but both nourishing and full of flavor. After 25 years it is a guide for teachers, and for the best of beginning students.

For him learning was never passive either. He was no omnivorous reader. Yet one evening long ago he ran across for the first time Prescott's famous accounts of the conquests of Mexico and Peru. Fascinated, he borrowed the heavy volume, to read it steadily all night and day. He bubbled with questions. The interest never left him; 30 years later he worked for a while, with the early success of a gifted amateur, at deciphering Mayan documents.

Frustrated with anatomical terms on one of his excursions into the brain, he asked the librarian for "a map of the cat." At the time of his death he had left half a dozen tracks behind in topics from molecular biology to computer algorithms to the psychology of hallucination.

The arts of the theater are inherently twofold. The actor on the stage pretends to be who he is not, by artful empathy and the words of another. That was not Richard's way. His theater—and it is impossible to evoke him without the word "theatrical"—was on the other side. Richard's was the stage where dancers, wire walkers and magicians daringly perform. What they do is striking, and not dissembled or illusory. It is real, whole, expressing mastery of some challenge, trivial or urgent, posed by nature and by human perceptions. On that stage he performed in four real dimensions.

In Far Rockaway days, his remarkable father once explained to the curious boy that the name of a wood thrush or some other bird was certainly interesting, part of human beings and their rich languages. Yet almost nothing of the wood thrush itself lies within its many names. What you must do to know more of the thrush is not to name it but to listen to it, watch it, think of what you notice. No powerful theorist in our time has so conspicuously kept a keen eye on the great pied bird of nature. The example Richard Feynman set us each day is no longer renewed. We miss him deeply, and we all learned from him while he was here. —*An old friend*


England
China
Switzerland
SWISS MADE
Japan
11:18
United States
WALTHAM
Italy
The Netherlands
France
Germany

SCIENTIFIC AMERICAN June 1988 Volume 258 Number 6

U.S. Economic Growth

It can be enhanced by improving workers' skills, supporting research and development and encouraging investment in capital that applies technological innovations

by Ralph Landau

Since 1982 the U.S. economy has enjoyed an extended period of expansion. Yet there are abundant signs that all is not well: they include a huge Federal deficit, an adverse balance of payments and status as the world's leading debtor. The most telling sign of all is a rate of economic growth lower than that of other industrial nations: since 1979 the annual growth rate of the U.S.'s real gross domestic product has averaged about 2.2 percent while Japan's annual G.D.P. growth rate has averaged 3.8 percent. Although this may not seem like much of a difference, it is actually a matter of grave concern. Only 2.3 percentage points separated the annual average G.D.P. growth rate of the U.S. from that of Great Britain between 1870 and 1913. Indeed, given that the U.S. population increased more rapidly than the population of Great Britain during that period, the difference in per capita G.D.P. growth rate amounted to only one percentage point. Yet that difference was enough to propel the U.S. past the leading industrial power of the 19th century while making it possible for the standard of living to nearly double with each generation. Such is the power of compounding economic growth over long periods of time: a few tenths of a percentage point, which may not appear significant in the short term, reflect an enormous economic and social achievement if sustained over a few decades.

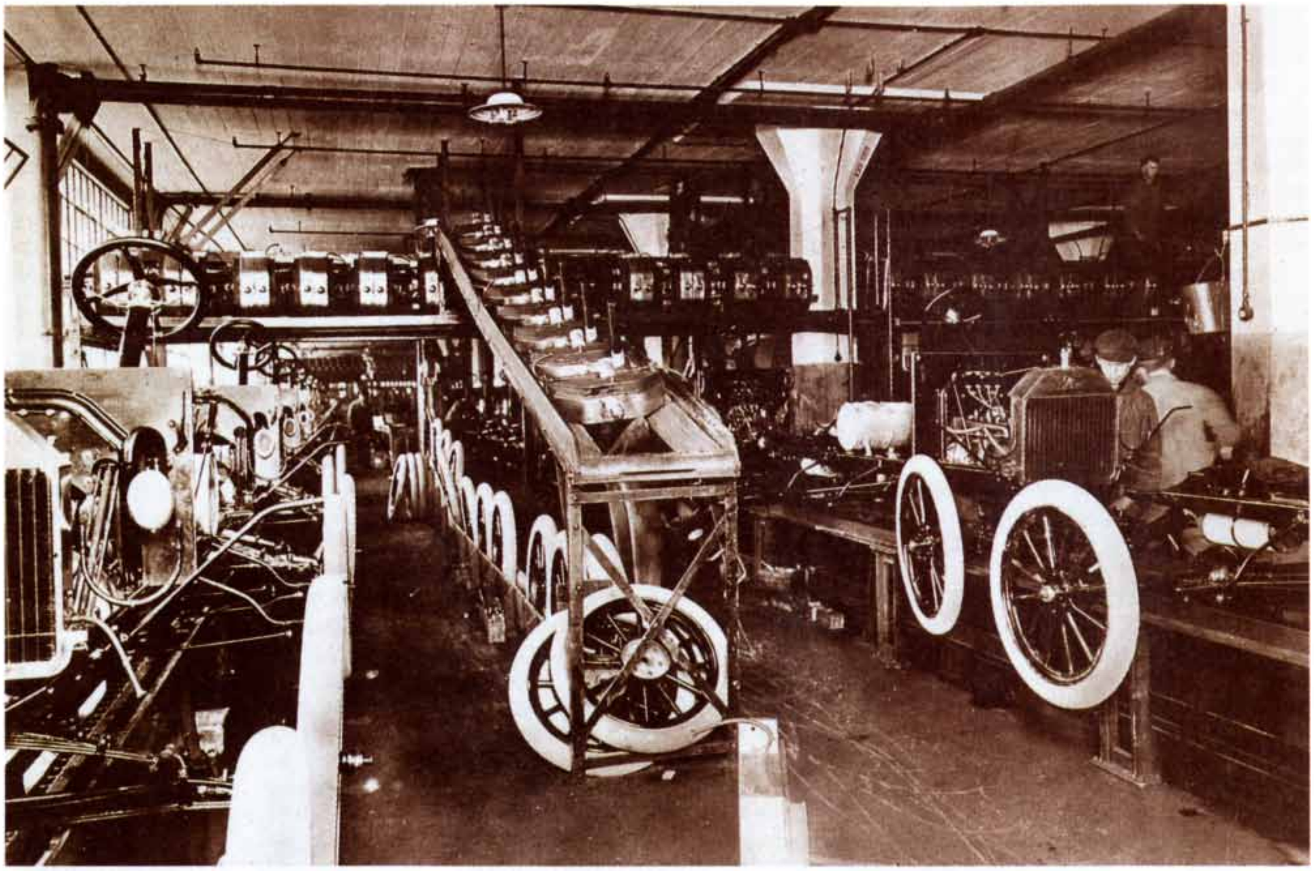
RALPH LANDAU heads Listowel Incorporated and is consulting professor of economics at Stanford University and a fellow of the faculty at the Kennedy School of Government of Harvard University. He codirects programs in technology and economic growth at both institutions. Landau, who holds an Sc.D. degree from the Massachusetts Institute of Technology, brings firsthand experience to the problems of fostering technological innovation in the U.S.: in 1946 he cofounded a chemical-engineering firm, which he headed for 36 years. Since 1981 he has been vice-president of the National Academy of Engineering, and in 1985 he received a National Medal of Technology. Landau would like to acknowledge the contributions of Michael Boskin of Stanford, Dale Jorgenson of Harvard and particularly Nathan Rosenberg of Stanford in the preparation of the article.

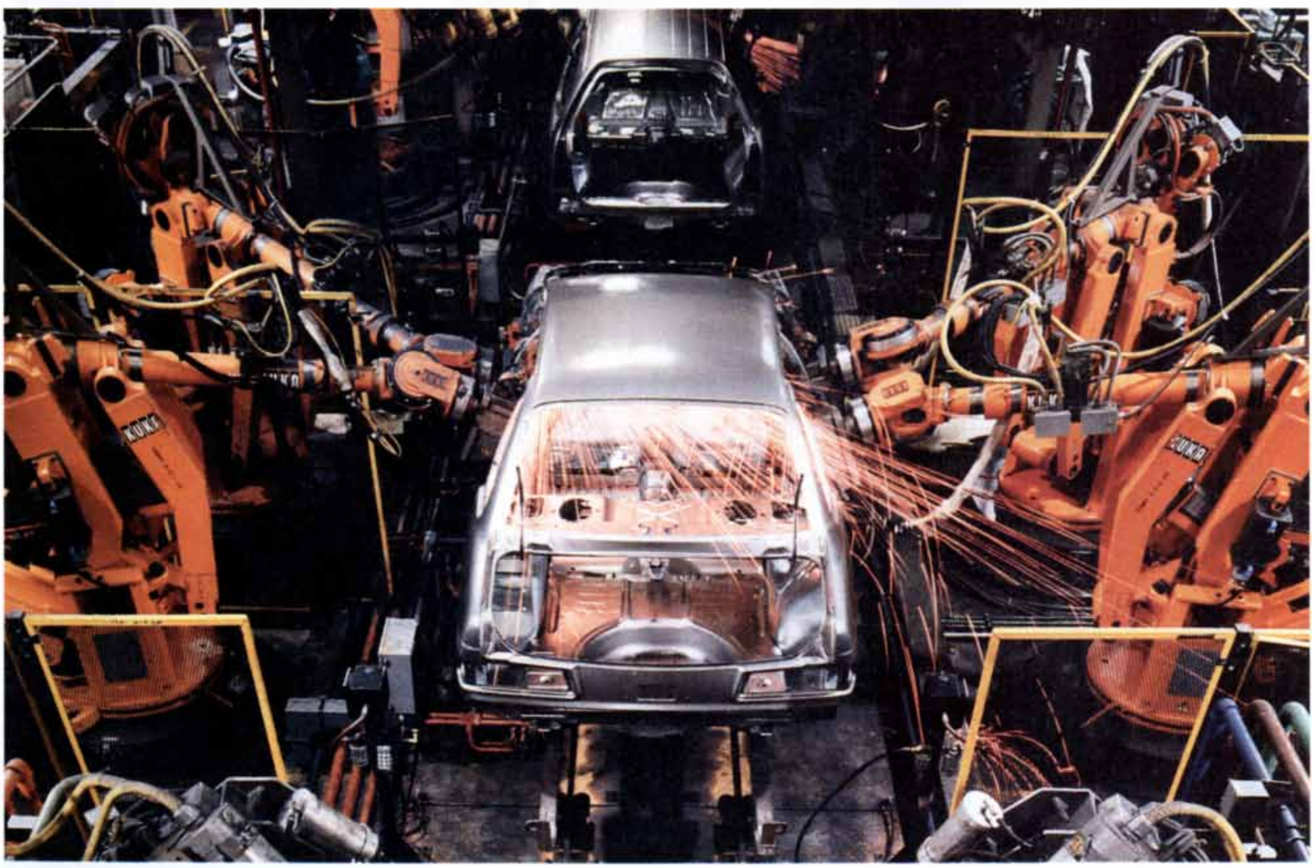
If the rate of U.S. economic growth is not increased soon, the ever greater employment opportunities and ever higher levels of material well-being to which most Americans have grown accustomed will no longer be possible. Is there a way for the U.S. to maintain a high growth rate and thereby ensure for itself both the benefits of growth and a more prominent position of global economic leadership?

My colleagues at the Program in Technology and Economic Growth at Stanford University and the Program in Technology and Economic Policy at Harvard University and I believe the answer is positive. Our reasons for thinking so are based on our examination of the issues from the perspectives of economic theory and our experience in business. Studying the intricate tapestry of the world economy reveals a number of interwoven threads that affect a country's rate of economic growth. One major determinant is the advance of technological knowledge. Equally important is investment in the capital required to apply that knowledge in the manufacture of products or the provision of services. Therefore the development of new technology as well as its diffusion throughout society is influenced by such factors as government policies affecting the money supply, national budgets and taxes. Although a government cannot force innovation, it can nonetheless foster growth by providing the proper economic climate for the private sector to apply the fruits of technological innovation.

In this connection we believe not enough attention has been paid to the role of capital investment in stimulating technological change; specifically the U.S. Government has not provided as advantageous an economic environment for capital formation as the governments of competing countries have. Indeed, in the past few decades its policies have worked against long-term investments in favor of short-term ones. If U.S. competitiveness and economic growth are to be improved, alternative policies that encourage long-range economic planning in the private sector must be favored.

A country's economy can supply more goods and services either by employing more factors of production, such as labor and capital (physical plant and equipment), or by increasing the productivity of such





TECHNOLOGICAL CHANGE is clearly evident between the manufacturing techniques applied in 1914 (*top*) and those applied today (*bottom*) at assembly lines of the Ford Motor Company. Because new capital (plant and equipment) incorporates technological change and thereby spurs it onward, adequate investment in capital is crucial for continued economic growth.

factors: the amount of output per unit of resource input. (The total output of a country's economy is often measured in terms of the country's G.D.P.: the value of all goods and services produced within its territory. G.D.P. differs somewhat from another common measure of economic output, the gross national product, in that it omits international transactions.) Economists for many decades considered the use of additional inputs as the main source of economic growth. Yet the first serious attempts to gauge the importance of additional inputs in economic growth—done in the 1950's—contradicted this view.

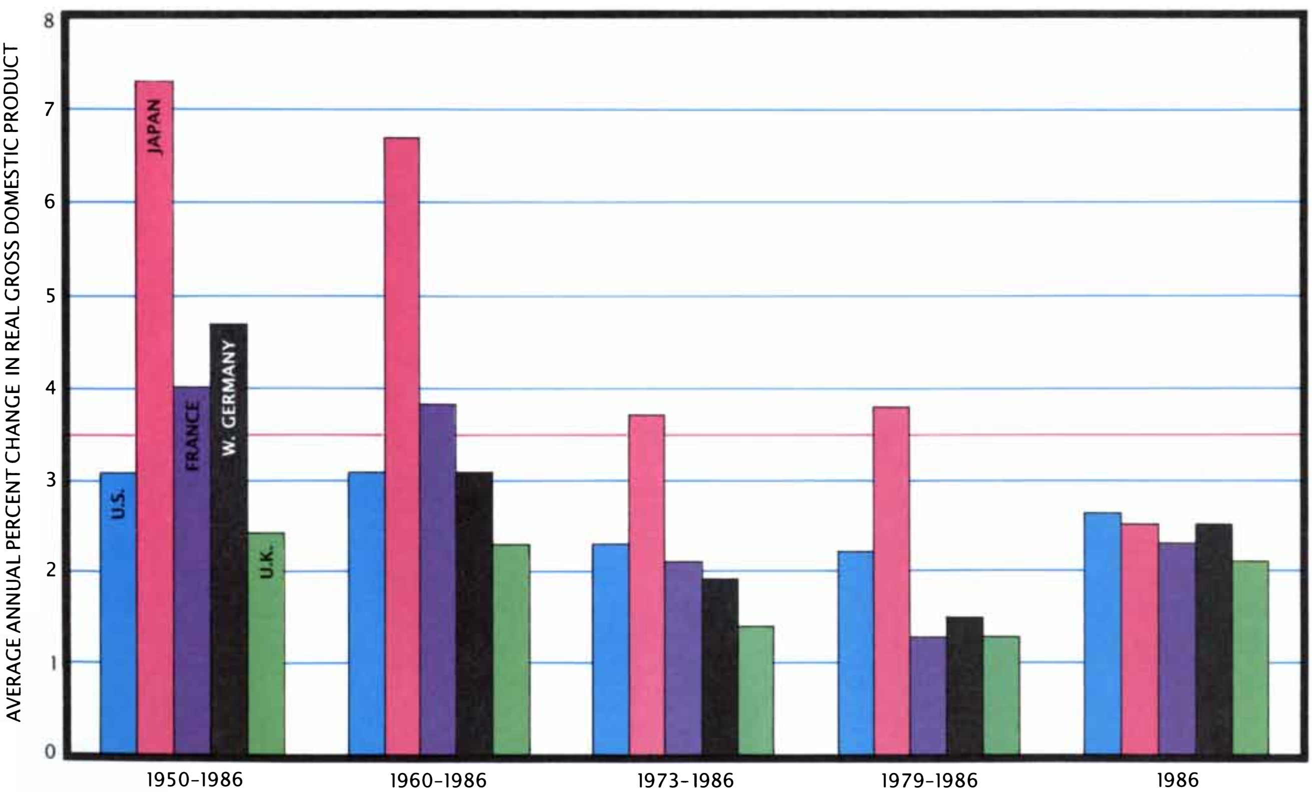
Robert M. Solow of the Massachusetts Institute of Technology, among others, formulated the basic relation: An economy's growth is simply the sum of the growths in the inputs of capital and labor, each weighted by a coefficient reflecting its average contribution to the value of products, as well as a third variable that represents the increases in factor productivity (the combined productivity of capital and labor). At first most economists were satisfied to label the third variable "technological change" and to relate it loosely to scientific and engineering research and development. Nevertheless, it clearly includes other important elements: more efficient resource allocation and economies of scale as well as many social, educational and organizational factors that serve to improve the quality of labor and management.

In the original formulation of the relation the three variables are assumed to be independent, so that increasing the rate of growth of, say, capital does not affect the rate at which productivity increases. The contribution of increases in productivity to economic growth is therefore the difference between the rate of growth of output and the weighted sum of the rates of growth of capital and labor. In this way Solow and others estimated that as much as 85 percent of U.S. economic growth per capita as recorded in historical data seems to be attributable to increases in productivity or technological change. (It was for this work as well as other contributions to the study of economic growth that Solow was honored last year with the Nobel prize for economics.)

Moses Abramovitz of Stanford had earlier applied a similar technique and had also found that only about 15 percent of the growth could be traced to the use of more inputs. He was rather circumspect in interpreting the large 85 percent residual, however, calling it "a measure of our ignorance." What seemed to emerge very clearly from those early analyses was that the country's long-term economic growth had until then been overwhelmingly a matter of using capital and labor more efficiently rather than simply increasing those inputs.

Later calculations with data from more recent years presented the matter in a somewhat different light. Whereas the factor productivity of the U.S. economy had increased by an average of 1.19 percent per year between 1913 and 1950, between 1973 and 1984 the factor productivity actually *declined* by about .27 percent per year. Moreover, in contrast to what Solow, Abramovitz and others had deduced from the earlier data, what growth there was in the U.S. G.D.P during the late 1970's was brought about almost entirely by increases in capital and labor (particularly the latter as the "baby boom" peaked).

Explanations for this striking collapse in factor productivity have varied. A distinctly promising explanation has emerged from recent comparative studies of individual sectors of

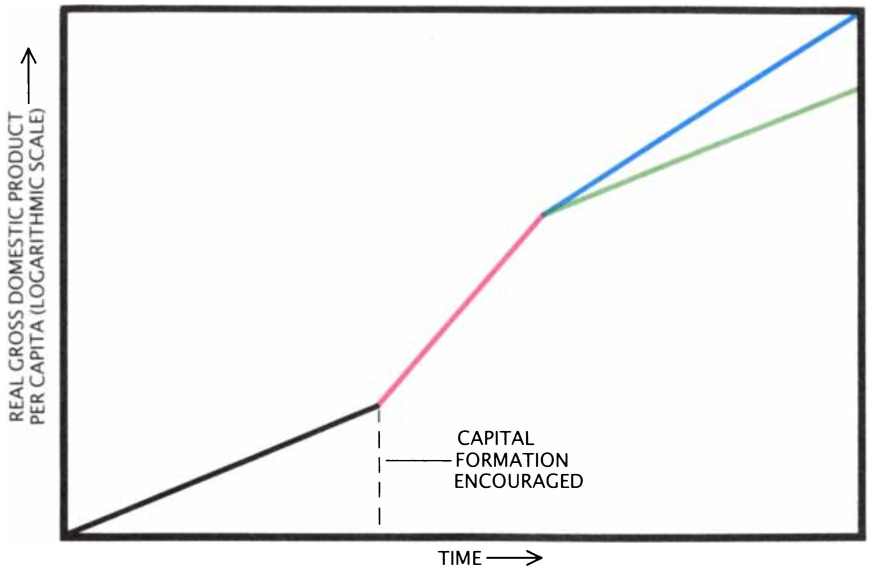


INTERNATIONAL COMPARISON of average annual rates of economic growth shows that the U.S. historically has had lower rates than Japan. The author argues that the U.S. should aim for an average growth rate of 3.5 percent per year (*red line*) in order to ensure long-term economic prosperity. (Gross domestic product is the value of all goods and services produced within a country's borders. It differs slightly from gross national product in that it excludes net international transactions.)

the Japanese and U.S. economies by Dale W. Jorgenson of Harvard. His work suggests that the productivity of labor is significantly affected by the per-worker rate of capital investment: the amount of money spent in building up the country's capital stock. Jorgenson has also shown that the Japanese factor-productivity growth rates (which between 1973 and 1984 averaged .43 percent per year) were heavily influenced by the high rate of capital investment in a number of industrial sectors. In fact, the rate of Japanese investment was at least twice as high as the rate in many U.S. industries.

Based on recently available data and a more sophisticated methodology that prices capital in relation to its age, Jorgenson states that since World War II capital investment has been responsible for about 40 percent of the growth of the U.S.'s G.D.P., whereas factor productivity accounts for about 30 percent. The remainder can be traced to increases in the quality and quantity of labor. Angus Maddison of the University of Groningen in the Netherlands has recently reviewed a number of studies on economic growth done in the past 30 years and reaches similar conclusions.

Other data also suggest a close correlation between national rates of capital investment and increases in productivity and economic growth. West Germany and France, for example, which have investment rates roughly twice the U.S. rate, also enjoy about twice the growth rate in productivity. (It should be pointed out that measurements of productivity growth usually do not take account of improvements in the quality of goods and services, which are also important.)



RATE OF ECONOMIC GROWTH is constant (*black line*) as long as capital formation (the construction of new factories and production equipment), labor-force improvements (the training of workers) and technological change (the development of new inventions) take place at a constant rate. When capital formation is encouraged by changes in a nation's economic policy, the growth rate increases (*red line*), since the nation acquires a greater capacity to supply goods and services. If there are no interactions among the rate of capital formation, the quality of the labor force and the pace of technological change, the economy returns to its original rate of growth in the long term (*green line*). But if increasing the rate of capital formation accelerates the rate of labor-force improvements and stimulates technological innovation, there may be a longer-term increase in the rate of growth (*blue line*).

How does investment in capital stimulate productivity to cause economic growth? The answer is found in the way technological change is incorporated into capital. Except for a small part devoted to basic science, research and development is seldom undertaken unless its results are expected to be applied in new facilities and superior operating modes that can either increase productivity, reduce costs or raise the quality of goods or services. Hence capital investment embodies technological change and also spurs it, since the availability of superior technology is a major incentive to invest. But a superior technology can only arise from research and development. Capital investment therefore serves as a catalyst that can set off a self-generating "virtuous circle" connecting technological change with research and development, eventually resulting in productivity or quality improvements in goods and services.

Similarly, improvements in labor quality through increases in knowledge and training can be seen as another form of investment (in human "capital") that is both a requirement of and an inducement to technological change. In other words, technological change is incorporated in each of the basic factors of production. This relationship is true to a much greater extent than was ever thought before, although the precise extent is difficult to quantify.

As Michael Boskin of Stanford and others have pointed out, these interactions between technological change, capital and labor quality imply that raising the rate of capital investment can actually increase the rate of economic growth over a longer period of time. This is particularly true in exploiting the results of "breakthrough" research, which requires large capital investments but can create entirely new industries. Unfortunately many managers, government officials and economists still do not understand these critical relationships in growth theory. They often maintain that increasing the rate of capital investment will not substantially accelerate technological change and improvements in labor quality.

Solow has recently joined the ranks of those who criticize this thinking. He has stated he fears that his theory of growth, which downplayed the importance of capital, might have been carried too far, contributing to "a severe underinvestment" in physical capital. Since "you can't take an old plant and teach it very new tricks," he has called for increased spending on behalf of the nation's research-and-development centers and factories—the sources of technological change.

Earlier distinctions between the roles of technological change, capital and labor in economic growth therefore need to be modified in favor of a view that sees them as intertwined parts of the same process. Technological change has been central to U.S. economic growth, both directly (in which case it can be said to account for perhaps 30 percent of economic growth) and through its positive effect on other factors of production (which can be said to account for perhaps another 40 or 50 percent of economic growth). Although scholarly research has not yet provided conclusive evidence, my colleagues and I believe it is in this broad sense that technological innovation is the key to viable strategies for future economic growth: it can raise the factor productivity of the economy at an accelerated rate.

The interactions among technological change, capital investment and labor quality suggest some strategies for U.S. companies to pursue in order to sustain economic growth. Companies in every sector of the economy must adjust to a faster pace of change, insist on continual training and remove obsolete policies that impose unnecessary constraints on technological innovation.

These needs are most acutely felt in the manufacturing sector. It is this sector that currently carries out about 95 percent of private research and development. It provides the major part (about two-thirds) of U.S. merchandise exports. (Indeed, both the bulk of the private research and development and the major part of the exporting is done by only a handful of large manufacturing companies, which also tend to be the ones that invest most in new capital.) The manufacturing sector also buys a large part of the output of the service sector, which in turn helps to improve manufacturing productivity. Finally, manufacturing produces material essential for national security.

Because manufacturing has such a key role in the U.S. economy, no strategy for economic growth that neglects this sector is likely to succeed. One way the sector can attain a capability for quick response to changes in the market is by relying more on the evolving technique of flexible manufacturing: the manufacture of different products by the same capital equipment. Of course, in order for such manufacturing to exploit its advantages fully, product design must be tightly coupled to market needs and be capable of rapid reductions in production costs. As the Japanese have recognized, consumers also demand constantly improving quality and reliability. This means that a company's manufacturing plants need to be in close communication with engineering and marketing staffs. Because they are principal agents of technological change, engineers must understand more than science and engineering; they need to cultivate economic judgment and learn how to use it in the face of technical and commercial risks. Greater worker involvement is also essential; this is a key feature of the successful Japanese manufacturing strategy.

Flexible manufacturing plants integrated on a national scale would allow rapid introduction of new and better products according to market conditions as well as the introduction of new processes that come with growing knowledge and experience. Such improvements in both products and processes need to be made at an increasingly rapid rate to maintain competitiveness. If it is prudently done, the capital investment can be largely recovered before imitation by competitors becomes a threat. U.S. companies that have achieved successes in this way include Allen-Bradley, Caterpillar, Chrysler, John Deere, General Electric, Hewlett-Packard, International Business Machines and Westinghouse.

Even so, U.S. companies that manufacture discrete mechanical goods, such as automobiles, appliances and consumer electronics, have had problems in attempting such integration of design, production and marketing,

FLEXIBLE MANUFACTURING approach applied in the production of lighting panelboards at the Salisbury, N.C., plant of the General Electric Company has made it possible to cut costs by 30 percent and to deliver orders in days instead of weeks.

partly because the companies have been reluctant to adopt a "systems" approach, which includes feedback loops among the company's various departments. The chemical and refining industries of the U.S., in contrast, are in a more receptive frame of mind, owing in large measure to the earlier development of the systems approach of chemical engineering.

Industries that rely heavily on computers, automation and telecommunications are also in a particularly favorable position. They suit American talents, such as the extraordinary ability to establish the entrepreneurial companies that create most new jobs. Although it would seem that commodity industries (such as the petrochemical and steelmaking industries) would be at a disadvantage, some of them can still remain competitive as long as they have a significant advantage in terms of scale, resources or technology. For many commodities, however, the transfer of their manufacture to lower-wage countries is probably inevitable. No one economic strategy fits every company.

The rapid integration of computer-based technologies and transformation of the work force can be exploited by the services sector as well. Even where the bulk of the value added to a product is in marketing, sales and distribution, large amounts of capital are still required, as is immediate feedback to manufacturing. Both sectors can thereby be in a position to add higher value to the products or services they supply. If this turns out to be the case, even high wages would constitute only a small part of the total cost of production, enabling U.S. companies to compete with companies in countries with low wage rates. The key point is that sectors in which the value added per worker is high are the ones that have built up substantial tangible capital (such as physical equipment and machinery) and nontangible capital (such as worker training, developed markets and technical knowledge) per worker.

What has been described so far has to be carried out by the private sector in order to remain competitive in both domestic and international markets. What happens at this microeconomic level, however, ultimately determines long-term growth rates, and it can be affected by such "supply side" factors as Government policies regarding taxes, trade, environmental regulation and labor. Yet the private sector must also operate within a macroeconomic environment established primarily by the Government's fiscal and monetary policies. These policies, which control the "demand side" of the economy, are generally shaped by political exigencies and by reactions to the many short-term and cyclical problems that beset the economy. As a result these policies have changed frequently.

Such a volatile macroeconomic environment can be inhospitable to major investments in research and development and can therefore hinder prolonged economic growth. Science and technology require distant planning horizons; they are bound to be undervalued if returns on investment in research and development are measured within a short time frame.

An economic climate that puts a premium on quick returns has in fact prevailed in recent years—a result, in large part, of the inflation that took place in the 1970's. Memories of the double-digit percentage increases in prices at that time still linger today, as can be seen in the high real interest rate on long-term bonds. As a consequence managers skilled in financial matters have often been favored over engineers in promotions and salaries, and short-term research-and-development projects have been favored over longer-term work that might produce technological breakthroughs.

Similarly, recent economic "stabilization" policies will adversely affect growth for many years to come. Perhaps the most critical has been the proliferation of debt. Government deficits, which have grown to be in the range of $200 billion per year, are financed to a large extent by borrowing from national savings, reducing the amount of money available for capital investment. At the same time private debt in all sectors has also grown. Total public and private debt in 1986 was more than 200 percent of G.D.P., up from 163 percent in 1975. These heavy debt burdens have resulted in a huge "inflation" of credit that is inherently dangerous, because as soon as there is an economic downturn, servicing the debt squeezes out investment in capital.

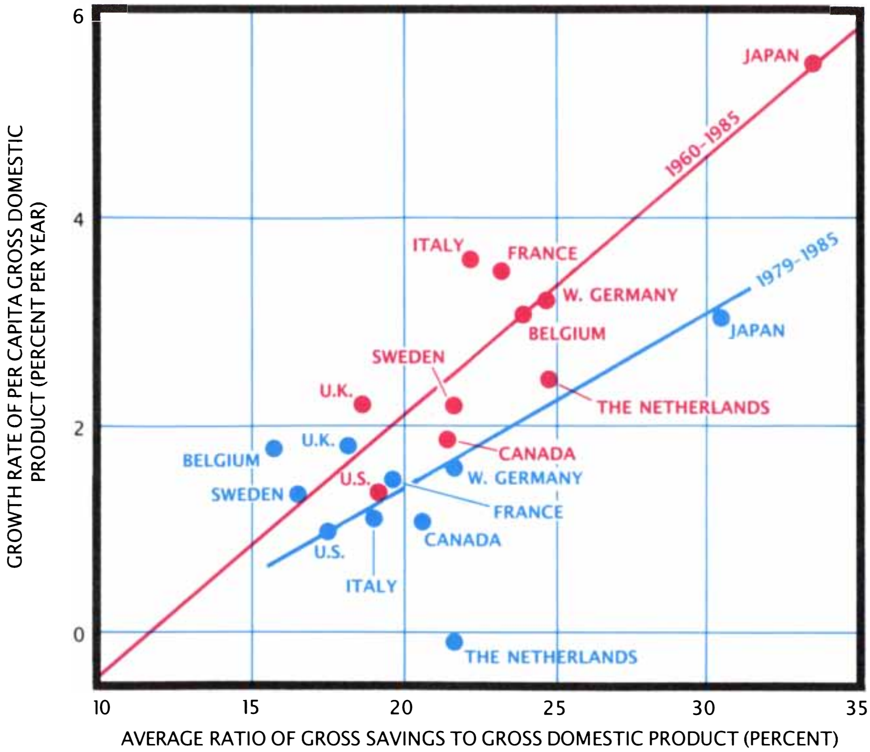
As a result a debt-laden company is less likely to accept the risks involved in conducting long-range research and development or in investing in new capital. In fact, the discouragement of long-term capital investment coupled with interest rates that have declined somewhat from their peaks in the early 1980's has prompted investors to move into common stocks, fueling a booming but volatile stock market. In addition investors have turned their attention to higher-yielding, riskier financial instruments, such as the so-called junk bonds.

Some of the debt—particularly that caused by the sale of junk bonds—has gone into mergers, acquisitions, takeovers, restructurings and leveraged buyouts. In effect debt has been substituted for equity. This tactic is encouraged by tax laws that allow the interest on debt but not dividends to be deducted from corporate income; the U.S. Department of the Treasury, in effect, has become the corporate raider's ally. Because borrowed money thus becomes cheaper, takeovers and leveraged buyouts have become more attractive than productive investments, which require a more judicious mixture of debt and more "expensive" equity.

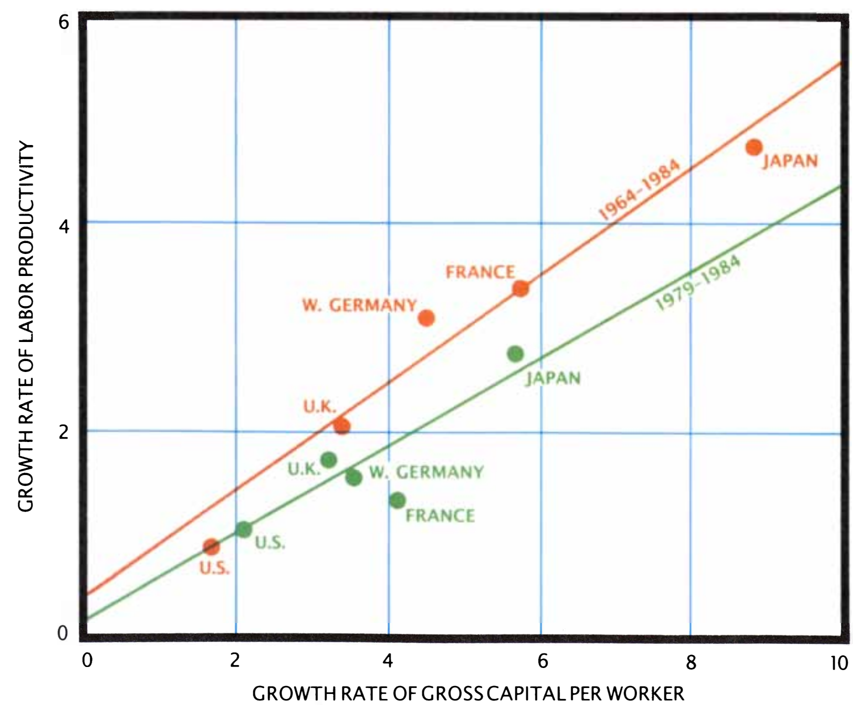
To be sure, an affected company sometimes becomes more efficient: restructuring has resulted in some gains in productivity and cost control in American manufacturing. More often, however, only the existing shareholders benefit (and only in the short run) and almost everyone else suffers from the constrictions in jobs, money, markets and future-oriented research and development. Moreover, the company (if not dismembered) still has to operate with extraordinarily high debt-equity ratios, a condition that is not conducive to long-term growth.

A prudent macroeconomic strategy for long-term growth would include a tight fiscal policy (aimed at reducing government deficits), a looser monetary policy (aimed at lowering interest rates and stabilizing the value of the U.S. dollar) and a tax system that encourages investment instead of consumption. (Changes in the U.S. regulatory and legal systems to cut down on wasteful lawsuits would also help.) Probably the most important prerequisite is that capital in all its forms, including the capital required for improved education, training and research and development, should be abundant and cheap; it should also incorporate the most efficient technology available. Above all, this means pursuit of policies that encourage high rates of savings, since it is from the savings pool that a nation's money for capital investment is drawn.

Boskin has reported that the savings rate of Japan is from two to nearly three times that of the U.S., depending on how it is measured. A high savings rate allows the Japanese to make heavy long-term domestic and international investments while still fi-



NATIONAL SAVINGS RATES correlate with per capita growth rates, since a country's aggregate savings account represents a pool of money on which companies can draw for capital formation. As can be seen, the U.S. had one of the lowest savings rates among industrial nations both from 1960 through 1985 (*red*) and from 1979 through 1985 (*blue*). The flatter slope of the line through the more recent data indicates the fact that growth rates in general have declined throughout the world since 1979.



GROWTH RATE OF CAPITAL per worker determines how quickly the productivity of labor rises. Because the U.S. has not increased its investment in capital per worker as fast as Japan has, labor productivity has grown at a lower rate in the U.S. than in Japan over the periods from 1964 through 1984 (*orange*) and from 1979 through 1984 (*green*). The steeper slope of the line for the data from 1964 through 1984 is a result of the fact that Japan and other countries were then "catching up" to the U.S. by constructing more new plants that incorporated the latest technology.

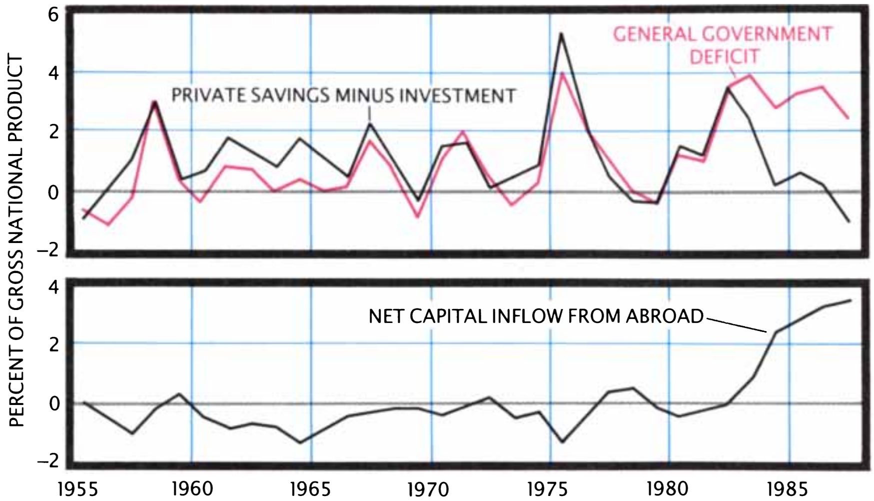
nancing government spending. It also explains the fact that the effective cost of Japanese capital has been from a half to a third as great as that of American capital. (Other reasons for the higher cost of American capital include the higher interest rates demanded by cautious investors, who have been hurt by past economic policies of the U.S., and the "second" tax on corporate income entailed by taxing dividends.) Because the future returns on an investment in Japan can be discounted at a half or even a third of the U.S. rate and still cover the cost of the investment, Japanese corporate chiefs have a much longer horizon for decision making. Consequently Japanese can better afford to finance research and development; they can also incur substantial initial losses while establishing markets abroad (as they did with videocassette recorders). Such activities generally require much longer time horizons than investment in plant and equipment.

Most economists agree that a suitably progressive consumption or excise tax—beginning perhaps with a tax increase on gasoline, tobacco and alcohol—would favor a higher savings and investment rate, provided it did not open the door for new spending programs that increase consumption. The U.S. could therefore cut the cost of capital substantially by relying more on a consumption-tax system and less on its income-tax system.

In spite of some positive features, the 1986 Tax Reform Act is an example of what not to do: it will probably increase the cost of funds for investment, considering the loss of the investment tax credits, the changes in depreciation allowances and the elimination of the capital-gains tax differential. It is curious that although the 1986 act implicitly recognized the desirability of incentives for more research and development, it ignored incentives for the necessary downstream capital investment. A further example of the antisavings aspects of the 1986 act can also be found in the way it has made Individual Retirement Accounts substantially less attractive. David A. Wise of Harvard has recently argued that there was relatively little substitution of IRA's for other forms of savings—they represented a promising new source of net savings.

Compared with such macroeconomic policies, other contributions of government may be less dramatic, but they are nonetheless important in smoothing the path to greater long-term growth. These include support of education in general, funding of research and development, maintenance of the infrastructure necessary for commerce (such as the national highway system) and the provision of safety nets for those excluded from the rewards of economic growth.

Of course, there may be valid reasons to enact macroeconomic policies at home that are at odds with the goal of continued growth. The Federal policies in the past 20 years, for example, have tended to place the priority of job formation above that of capital formation. Managements of U.S. companies were in effect induced to substitute cheaper labor for capital investment, and in fact many jobs were created—far more than in Europe or Japan. Between 1955 and 1986, 47.4 million new jobs were created in the U.S., 16.6 were created in Japan and a total of only 5.7 in the European Economic Community. Indeed, unemployment in Europe is currently in the double-digit range, whereas in the U.S. it has dropped to about 5.5 percent. Because high unemployment rates are always a potential threat to domestic tranquillity, the extraordinary success of the U.S. in creating jobs may well justify in hindsight some of its past economic policies. Nevertheless, such policies have exacted a price in terms of lower productivity growth and a more obsolete capital stock.



GOVERNMENT DEFICITS are usually financed by borrowing from private savings after a country's investment needs have been satisfied. Nevertheless, since 1982 the combined budget of Federal, state and local governments in the U.S. has run sizable deficits, in spite of the fact that there have been few excess funds in savings. The shortfall in domestic savings has been made up by drawing on foreign savings.

All that has been discussed so far is complicated considerably by the inescapable fact that the U.S. economy must operate in the context of a truly global market not only for goods and services but also for technology and money. Goods and services are today bought and sold on a vast scale. Because there is a greater openness to trade, many countries are competing vigorously with the U.S. in the international market—often producing products that are better in quality and lower in price.

New technology is spreading faster than ever before among countries, including many with wage rates a great deal lower than those of the U.S. A century ago transfer of technology from one country to another took place over a number of years; today it may be as short as several months. The causes of this rapid spread include investment abroad (particularly by multinational companies that are moving closer to their markets), international licensing of technology, education of many scientists and engineers in the universities of the West, rapid publication of research and the increasing ease of international travel and communications.

A consequence of the rapid worldwide diffusion of technology is that the country that comes up with a technological innovation is not necessarily the first to commercialize it. Color television, for example, was an American invention superbly perfected and marketed by Japan. In addition other nations have capital-investment rates that are much higher than those in the U.S. Consequently new production technology developed in the U.S. can often be embodied in capital stock sooner in foreign countries than in the U.S. itself. In this way economic competitors of the U.S. can increase the productivity of their workers more rapidly.

The result is that industrial technology has now taken root in many places throughout the world; no longer does the world depend on the manufacturing ability and technical progress of just a few industrialized nations. The sources of supply are nearly global: Malaysian electronic products, Korean automobiles, Swiss pharmaceuticals and Chinese fabrics flood the markets of the U.S. and other nations. It is a historic change, and a permanent one. There is no realistic chance that this situation can be unilaterally changed by the U.S. or any other country, although individual companies have become more conscious of the risks of licensing valuable technology. Governments are also taking more steps to protect the intellectual property of their nationals—a particularly important matter in this day and age, since there are many exciting technologies awaiting exploitation.

Transnational flows of money are now also able to penetrate almost any market in the world instantaneously, exposing a country's policies to strong external market forces. Indeed, trading in financial assets far exceeds trade flows in goods and services, perhaps by as much as a factor of 20. It would be no exaggeration to say that since about 1979 currency-exchange rates have become an important driving force of the U.S. economy. This situation reflects the fact that money flows in a direction opposite to that of the flow of goods and services. The reason is that whenever the demand for U.S. dollars becomes stronger (as it did in the early 1980's, when high interest rates prevailed in the U.S.), its value in relation to the other currencies increases. Consequently U.S. exports (whose prices are tied to the dollar) become overpriced and imports flood in from many countries. The result is that foreign suppliers gain large domestic-market shares and weaken many U.S. companies, forcing some even into bankruptcy.

This has been the situation until recently, causing the merchandise trade balance (the difference between exports and imports of goods) to plunge to a record $171-billion deficit in 1987. The so-called current-account balance, which in addition takes into consideration the values of such things as payment for international services, repatriated earnings, royalties from abroad and interest on foreign debt, was slightly less: it shows a deficit of $161 billion. As long as exchange rates continue to float, the global financial system will exert a major influence on national economic policies.

Another effect of the tremendous liquidity of borrowed funds in international financial markets has been the startling increase in American debt to other countries. The 1987 Economic Report of the President states that at least 50 percent of net investment in the U.S. came from abroad, mostly because U.S. companies cannot count on domestic savings for investment capital. (Whereas net investment in the U.S. in 1987 was 5.3 percent of the country's G.D.P., net domestic savings was under 2 percent.) At this rate of inflow it will take only a few years for the cost of servicing the foreign debt and the payment of dividends on assets in the U.S. owned by foreigners to become a significant drag on U.S. economic growth. Yet as long as the U.S. savings rate remains unable to cover productive investment in the U.S., the pool of available investment capital will have to be supplemented from abroad (which means higher interest rates). In short, the U.S. has been consuming too much, producing too little and borrowing from abroad to maintain its standard of living.

The negative current account balance of the U.S. (which amounts to about 3.5 percent of the country's G.D.P.) reflects this fundamental imbalance between domestic investment and domestic savings. To remedy the situation either U.S. savings must increase or investment in the U.S. must decrease, which would be catastrophic. Hence the only real choice is to increase savings in a systematic way, by ensuring that the Government's spending does not overwhelm Americans' private thrift. (Although the decline in the value of the dollar since 1985 has helped, it certainly will not be enough.) The goal should be to enact policies that regulate consumption so that in a given year it can grow no faster than about 1 percent less than the rate of growth in ouput, thereby making it possible for net savings rates to increase an average of 1 percent per year more. Over a period of between eight and 10 years the net savings rate would move into the double-digit range typical of the U.S.'s international competitors.

Of course, given the size of the U.S. economy, a change in domestic consumption or savings can affect many nations throughout the world, and this has to be taken into consideration when reducing the trade and budget deficits. Indeed, in this sense the country's recent overconsumption and underproduction has actually benefited the world, since it allowed many debt-ridden, less-developed countries to sell their goods and services in the U.S. Nevertheless, it is time for the U.S. to change its priorities and to lay the basis for better long-term growth in a world economy.

The irrevocable changes in the global market I have described pose immense new difficulties particularly for U.S. managers, who gained their experience in the first few decades after the war, when the U.S. dominated world markets and faced little if any competition for domestic markets. International trade has shifted from a situation in which countries could count on long-term stable comparative advantages to a dynamic state in which comparative advantages are constantly being altered. Hence since the 1970's managers everywhere in the industrial world have been compelled to spend a great deal of their time dealing with the consequences of sharp fluctuations in worldwide energy prices, currency-exchange rates and inflation rates. The barrage of external influences has made it difficult even for good U.S. managers to balance short-term exigencies against the longer-term strategic planning that enables a firm to grow and prosper.

Given the current economic situation, there is certainly some room in a general strategy to improve U.S. growth for selective trade actions to meet egregious examples of foreign protectionism. Yet these measures should be applied to open up other nations' markets rather than to close those of the U.S. In fact, the U.S. already has a formidable protectionist arsenal of its own in the form of subsidies, trigger prices, quotas and tariffs. Gary Hufbauer of Georgetown University estimates that protection of some kind covered about 25 percent of U.S. imports by 1986 (it was 8 percent in 1975). Although these measures are considered "temporary," their effect on long-term growth may be much more significant. American companies would lose touch with global technological and economic developments, and the pressure to excel would be reduced.

The true significance of international competitiveness lies in the recognition that it is not an end in itself but a means to raise long-term growth rates so that jobs can be created and the standard of living can be improved. In order to accomplish these goals the U.S. has to restore its long-term real average growth rate of G.D.P. to at least its historic average (between 1870 and 1984) of 3.39 percent per year. Nevertheless, a growth rate of 3.5 percent should be the target for the time being, because it may well take as much as 1 percent of this growth to pay the debts of past profligacy. Taking into account the potential 1 percent penalty for the servicing of foreign capital and a 1 percent increase in the work force in the next decade (down from a peak of 2.9 percent per year between 1976 and 1980), the per capita growth in real G.D.P. would be about 1.5 percent per year.

Although achieving a real per capita G.D.P. growth rate of 1.5 percent will be virtually impossible without some Government cooperation, the Government must limit itself primarily to establishing the right macroeconomic climate for long-term economic growth. Pervasive Government intervention in domestic markets is likely to be counterproductive. Many in Government and elsewhere are already using the concept of "competitiveness" as a code word for old-fashioned Government manipulation of industry. Yet the Government would be overwhelmed in trying to direct the private sector through the rapidly changing conditions imposed by the pace of technological innovation.

Only if we increase investment in both capital and technology in all sectors of the U.S. economy (particularly manufacturing) and improve the quality of labor at all levels can the American standard of living rise at an acceptable rate. In the present highly competitive world market the U.S. has some historically demonstrated advantages, but it must take the longer view and pursue those seemingly trivial increases of a few tenths of a percentage point in growth rate each year.

FURTHER READING

THE POSITIVE SUM STRATEGY: HARNESSING TECHNOLOGY FOR ECONOMIC GROWTH. Edited by Ralph Landau and Nathan Rosenberg. National Academy Press, 1986.

TECHNOLOGY AND ECONOMIC POLICY. Edited by Ralph Landau and Dale Jorgenson. Ballinger Publishing Company, 1986.

PRODUCTIVITY AND U.S. ECONOMIC GROWTH. Dale Jorgenson, Frank M. Gollop and Barbara M. Fraumeni. Harvard University Press, 1987.

GROWTH AND SLOWDOWN IN ADVANCED CAPITALIST ECONOMIES: TECHNIQUES OF QUANTITATIVE ASSESSMENT. Angus Maddison in *The Journal of Economic Literature,* Vol. 25, No. 2, pages 649–698; June, 1987.

Introducing the all new Chrysler New Yorker. It gives you the one thing you always wanted in a luxury car.

Everything.

Elegance. Luxury. Front-wheel drive technology. Electronic fuel-injected V-6 power. And Crystal Key, the ultimate owner care program. Everything.

- Advanced front-wheel drive
- Powerful new V-6 engine
- Electronic fuel injection
- Four-wheel disc anti-lock braking system*
- Self-leveling suspension
- Crystal Clear paint
- Adjustable front and rear seat headrests
- Power rack-and-pinion steering
- Automatic temperature air conditioning
- Power six-way driver's seat
- Mark Cross leather seating
- Rear-seat stereo headphone controls
- Electronic instrument panel
- Electronic speed control
- On-board travel computer
- Crystal Key owner care

THE CRYSTAL KEY PROGRAM. BETTER OWNER CARE THAN EVEN ROLLS ROYCE OR MERCEDES.

- **5-year/50,000-mile Basic Warranty**† on entire car. You take care of normal maintenance, adjustments and wear items, we take care of the rest...the works. No ifs, ands, buts or deductibles.
- **7-year/70,000-mile Protection Plan**† covers engine and powertrain. 7 years/ 100,000 miles on outer body rust-through.
- **24-hr. Customer Hotline** for any questions on warranty or service.

Chrysler 7/70
Division of Chrysler Motors

CHRYSLER. DRIVING TO BE THE BEST.

*Some items are optional or not available on certain models. †See dealer for copy of limited warranty. Deductible on powertrain after 5/50. Restrictions apply. Crystal Key not available on Turbo model. BUCKLE UP FOR SAFETY.

THE LIFESTYLE RESOURCE™

At The Lifestyle Resource we give you all the facts and details necessary to make an informed purchase. Your satisfaction is our primary concern. If your purchase doesn't meet your expectations, return it in original condition within 30 days for prompt refund.

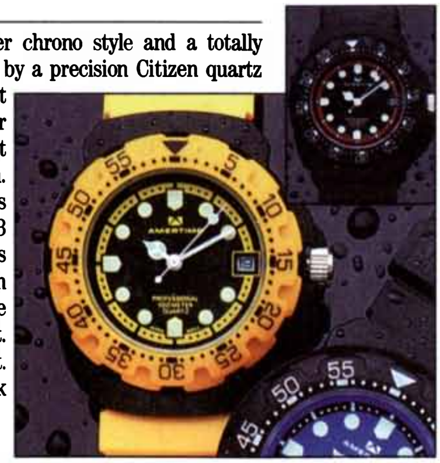
EXPERIENCE NATURAL STATES...THE VIDEO

At last—the real thing in relaxation videos. No more inane fishtanks or fireplaces: Miramar's "Natural States" cassette presents a 45 minute trip through the Pacific Northwest from California's Big Sur to Washington's Cascade Mountains. This is NOT a travelogue — not a word is spoken as the unique "steadicam" camera glides through rainforests, past verdant waterfalls, over and around mist-shrouded peaks. This unusual technique, mated to superb aerial shots, produces something very nearly like an out-of-body experience. The incredibly crisp footage is enhanced by an award winning score that was specially composed "note-to-frame," rather than just layered on in post-production. This is the first nature/music video to "go gold" in sales, and for good reason: its breathtaking beauty will bring tranquility to the worst case of stress and burnout. Just slip "Natural States" into your VCR; in 45 minutes you'll be human again! Hi-fi Stereo with Dolby Surround plays mono on regular machines. **$29.95** #2010.

WORRY-FREE WATERING

Water grows things best when it arrives at regular intervals, a fact that sometimes gets by old Mother Nature. Now you can give your lawn or garden the advantage of timely watering. The new Rainmatic 2000 *automatically turns water on and off* up to four times a day, seven days a week, a different schedule every day. Lets you set any length watering interval. Easily programmed to set up your own customized watering schedule or use built-in program. One-touch manual override. Great for sprinklers and drip/trickle systems. Operates all season on four alkaline D-cells (not included). Weather-protected electronic circuitry. LCD readout with digital clock. Locks on to faucet. Only **$49.95** #1080 with one-year warranty and instruction manual.

WHICH WATCH?

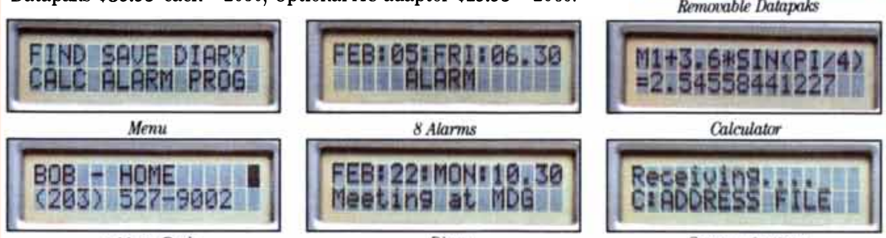


You choose. You get this year's most sought-after chrono style and a totally professional competition sports watch, powered by a precision Citizen quartz movement. Dive with it: it can go an incredible 660 feet deep. Monitor elapsed time for running, swimming, or cycling: the deep-knurled, uni-directional bezel can't accidentally reverse, and you can set it with gloves on. All metal case with scratch resistant mineral glass crystal seals out dust and water. Shock resistant 3 hand movement with date read out window. Luminous hands let the action go on even in the dark. You've seen watches just like this for $145. Now, enjoy the same exceptional quality and look for a fraction of the cost. Adjustable PVC band can be trimmed for perfect fit. Yellow with black, #1941; Black with red, #1951; Black with blue, #1961. **$69.95.**

A 3 INCH BY 6 INCH PERSONAL COMPUTER

Imagine being able to have every address, telephone number, appointment, every fact and figure you need at your fingertips. The Psion Organiser II makes it possible with a powerful PC that fits in the palm of your hand. *"The smallest serious computer ever..."* That's what *PC Magazine* said about the *Organiser II.* It represents the most significant downsizing in a computing instrument since the introduction of the laptop. Easy to use, the Organiser II accesses all functions from the "menu" which is automatically displayed when you switch on. With 32K of built-in software and powerful programming facilities, this 9-oz. computer replaces your address book, calendar, appointment diary, calculator, and all those little reminders you keep scribbling to yourself. PHONE, ADDRESS AND NOTE BOOK: Type in names, addresses, phone numbers, or any other personal or business data, press one key and it's stored. To find it again, enter any part of the information from the record stored and *in less than a second all of the information is displayed on the screen.* Lines can be up to 254 characters and automatically scroll horizontally so you can read the entire line. APPOINTMENT DIARY: The built-in diary program lets you record in half-hour increments, all appointments and reminders through the year 2000. Turns itself on and buzzes you up to an hour before an appointment. It also has 8 separate alarms and easily keeps track of your daily expenses. CALCULATOR: Calculations are displayed on the screen as they are entered and changed. 12 digit accuracy, ten memories, full mathematical and scientific functions. REMOVABLE DATAPAKS: *Just like a desk-top computer,* the Organiser II has the equivalent of two disk drives—tiny solid-state Datapaks. These perform the function of floppy disks and keep information safe even when removed. The *Model XP* provides *288K* of useable storage...the *largest* of any hand-held computer. While the *Model CM* gives you up to *136K* of useable storage. COMMUNICATIONS: Organiser II with an RS232 interface easily connects to printers, other computers and telecommunications devices. *Over 150,000 Organiser II's have been sold in Europe* in the past year and a substantial amount of software is available. Through an exclusive arrangement, *The Lifestyle Resource* is pleased to offer it to you in the U.S. Comes with a catalog of hardware and software, user-friendly manual, and one nine-volt battery. Model CM 8K Organiser II **$179.95** #2030; Model XP 32K Organiser II **$249.95** #2040; Additional 64K plug-in Datapaks **$89.95** each #2050; Optional AC adaptor **$19.95** #2060.

Removable Datapaks



Menu — *8 Alarms* — *Calculator*

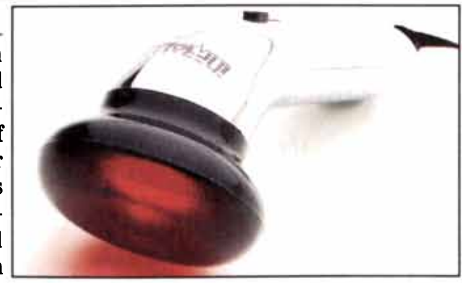
Address Book — *Diary* — *Communications*

THE LIFESTYLE RESOURCE™

THE 38,000 HOUR LIGHTBULB!

The problem with bulb savers is they reduce light output by almost 50%. But the patented, UL listed Power Mizer has solved that problem. These little buttons are activated by a state-of-the-art microcontrol switch that limits voltage to the bulb to 85%, thereby prolonging the life of the filament. Light bulbs using Power Mizer lose so little perceptible light that a foot-candle meter is needed to calculate the light loss. Under normal conditions, the Power Mizer will extend the life of a regular one-way 100W bulb from 750 hours to 38,000 hours or about four years if used 24 hours every day! Many major hotels, corporations and universities use these types of devices in hard-to-reach places. They pay for themselves in light bulbs, labor and energy savings in no time. Set of 12 Power Mizers **$29.95** #1990; 18 for **$39.95** #1995.

THE NEW INFRALUX



Pain emanating from inflamed joints and muscles can be excruciating. Just ask sufferers of bursitis, neuralgia, or arthritis. Infrared heat is the natural remedy preferred by doctors and therapists to dispel such nerve-racking pain. Recently redesigned, the Infralux infrared therapeutic heating unit for personal use now is contoured, angled for reaching back and neck. The heating surface is tripled and the replaceable 10-watt bulb gives 45% more heat. Hi/low settings. Plugs into household outlet. UL listed. Unit with case **$39.95** #1140.

AN AMERICAN CLASSIC

The *very best* of the Adirondack chairs are still made in the Adirondacks, in Willsboro, N.Y., on the shores of Lake Champlain. Willsboro's skilled woodworkers have done more than just re-create the original Adirondack chair - they've added refinements such as a fully contoured seat and 5″ wide arms for the ultimate in comfort. And the workmanship! The seat and backrest are hand-doweled and glued to the frame for strength, using solid hardwood pegs (not capped screws) in the true yankee manner. The rich rock maple is an extremely dense and attractively grained wood, all sanded and sealed - no sharp corners or splintering. Willsboro's artisans have incorporated a unique (and patented) tongue-in-groove folding mechanism that lets you fold the chair to a mere 10″ thickness for storage; sturdy hardwood locking pins secure it in the open position. When each piece is completed, it is treated with Willsboro's special zinc-based, rot-resistant clear protective coating; leave them to weather naturally or they can be painted or stained if you choose. There never has been an Adirondack chair like this...and we have a matching footrest, designed to create a chaise lounge, and an end table to go alongside - both constructed in the same fine tradition. The chair is 38″ x 14″ x 20″, **$129.95** #1880. The footrest is 14″ x 20″ x 22″, **$69.95** #1890. The end table is 16″ x 14″ x 24″, **$59.95** #1900.

THE FINEST WAX

Virginia cabinet makers developed a wax to beautify and preserve the fine wood furniture they crafted 200 years ago. Williamsville has reproduced this formula of beeswax and lemon oil, heat-blended with other natural oils in a mixture readily absorbed by wood. Instead of burying wood's natural beauty under layers of silicones found in commercial waxes, this treatment prevents wax build up while nourishing the wood. Has no harsh abrasives or silicones that can damage wood. Two 8 oz. bottles **$12.95** #1521.

THE LIFESTYLE RESOURCE™ CREDIT CARD ORDERS CALL TOLL-FREE 24 HOURS **800-872-5200**

DEPT. SFAF18; 921 EASTWIND DR. SUITE 114: WESTERVILLE, OH 43081

	ITEM NO.	QTY.	DESCRIPTION	ITEM PRICE	TOTAL PRICE
SEND TO (PLEASE PRINT)					
ADDRESS					
CITY STATE ZIP					
☐ CHECK OR MONEY ORDER ☐ MASTERCARD ☐ VISA ☐ AMEX SIGNATURE					
ACCT. # EXP. DATE			Shipping Charge covers UPS, handling and insurance for guaranteed delivery.	SUB TOTAL	
				SHIPPING (see table at left)	
				TOTAL	

Up to $20 .... $ 3.95	$50.01 to $60 .... $ 7.95
$20.01 to $30 .... $ 4.95	$60.01 to $70 .... $ 8.95
$30.01 to $40 .... $ 5.95	$70.01 to $100 .... $10.95
$40.01 to $50 .... $ 6.95	Over $100 .... $12.95

UPS 2ND DAY AIR

UPS Second Day available for an additional $7.50 per order.

ORDER WITH CONFIDENCE

- *Most orders ship within 48 hours of receipt.*
- *Credit Cards aren't billed until after shipment.*
- *No risk 30-day return privilege.*

THE MEDIA DEVELOPMENT GROUP

©1988 640 ADVERTISEMENT

"Snurps"

The name stands for small nuclear ribonucleoproteins, particles that help to remove meaningless "introns" from the messages issued by a cell's genes. Without them cellular activity would grind to a halt

by Joan Argetsinger Steitz

Eleven years ago I spent the better part of my sabbatical from Yale University in an ultimately fruitless pursuit. I was trying to develop antibodies that would bind to a curious class of protein molecules found in the nucleus of mammalian cells. No one knew exactly what the molecules did, but I suspected that they were important intermediates in the biochemical pathway from genes to proteins. I wanted to use antibodies to block their activity and thereby illuminate their function.

If success is measured in terms of achieving what one sets out to achieve, then I was woefully unsuccessful. I never did isolate antibodies to those mysterious proteins. But scarcely a year later my colleague Michael R. Lerner and I stumbled across an entire new genre of nuclear components that in fact turned out to be quite important mediators of gene expression. They are small complexes of RNA (a derivative of DNA) and a variety of proteins. We called them small nuclear ribonucleoproteins, or snRNP's (pronounced "snurps").

There are many different kinds of snRNP's, and functions have been assigned to only a few. Those few have already been shown to play a vital role in cell activity. They are the critical components of a sophisticated molecular assemblage called a spliceosome. As such they take part in the splicing of messenger RNA (mRNA), an essential intermediate through which the instructions contained in the genetic code reach the rest of the cell.

The splicing of mRNA is a delicate operation that must be carried out with the utmost reliability and precision. Perhaps it is not surprising, then, that the snRNP's in spliceosomes specialize: each performs a different task during the splicing procedure. The picture of snRNP's working in concert in the spliceosome suggests nothing if not a well-oiled machine.

JOAN ARGETSINGER STEITZ is professor of biophysics and biochemistry at Yale University, where nearly 10 years ago she and a graduate student discovered "snurps." For that discovery and her subsequent research she received numerous honors, including the National Medal of Science in 1986. Steitz has a B.S. from Antioch College and got a Ph.D. from Harvard University in 1967. She worked at the Medical Research Council Laboratory of Molecular Biology in Cambridge, England, before joining the faculty at Yale in 1970.

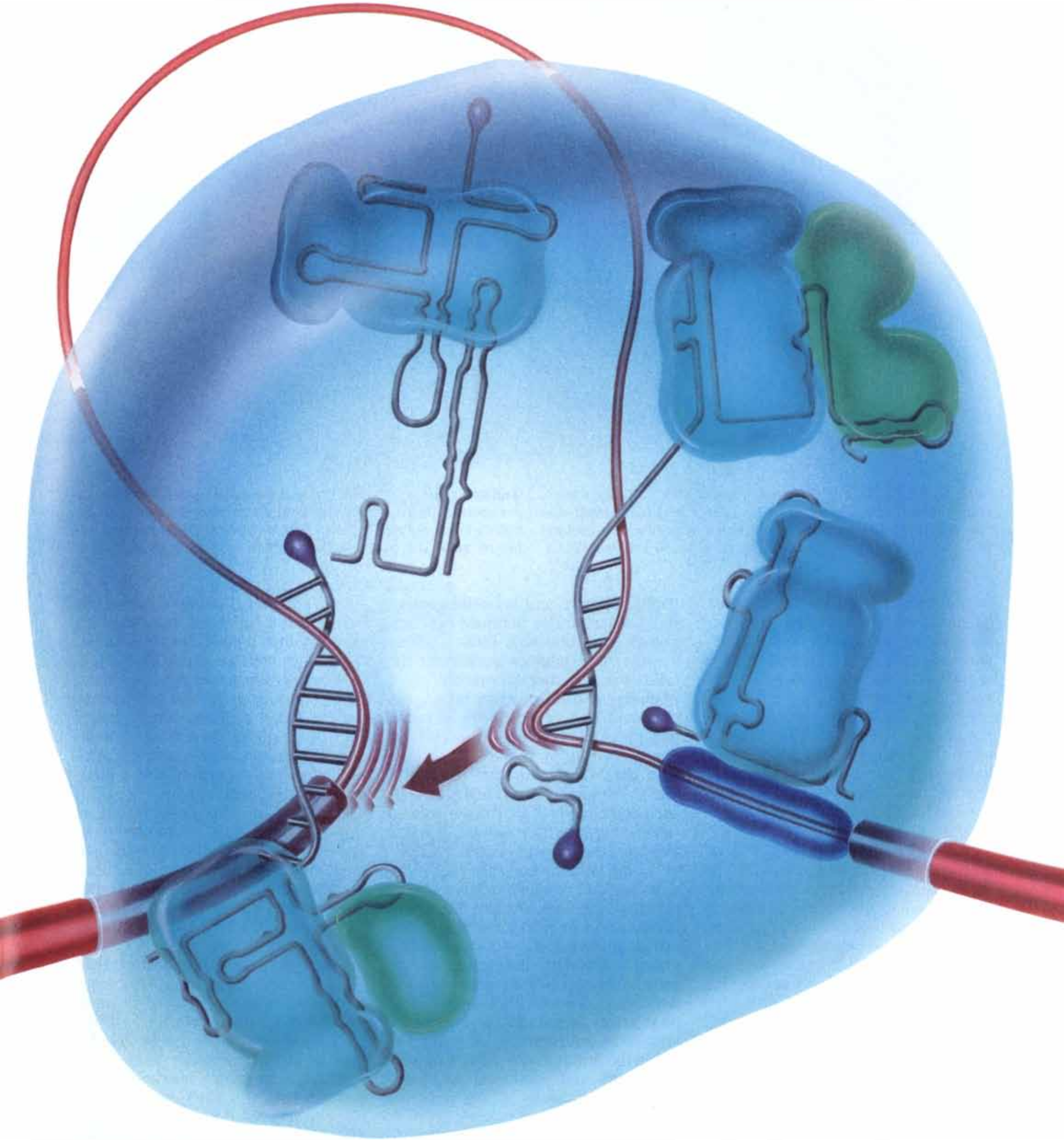
The significance of snRNP function can best be appreciated in the general context of gene expression: the process by which the genetic information in DNA directs the synthesis of proteins. In the cells of higher organisms the process begins in the nucleus, where the coding strand of double-strand DNA, which is made up of chemical units called nucleotides, is transcribed into single-strand mRNA. Eventually the mRNA transcript is exported to the cytoplasm, where its nucleotide sequence is translated into a sequence of amino acids. Amino acid chains fold to form proteins [*see illustration on page 58*].

Molecular biologists have known for a long time that two other kinds of RNA participate in gene expression. Like mRNA, they are transcribed from a DNA template, but they do not code for proteins: instead they help to translate mRNA. One kind, known as transfer RNA, or tRNA, forages for amino acids and escorts them to the mRNA template. The other kind, called ribosomal RNA (rRNA), is a major component of cellular organelles called ribosomes. Ribosomes are stable complexes of RNA and protein that mediate the interaction of mRNA, tRNA and amino acids during protein synthesis.

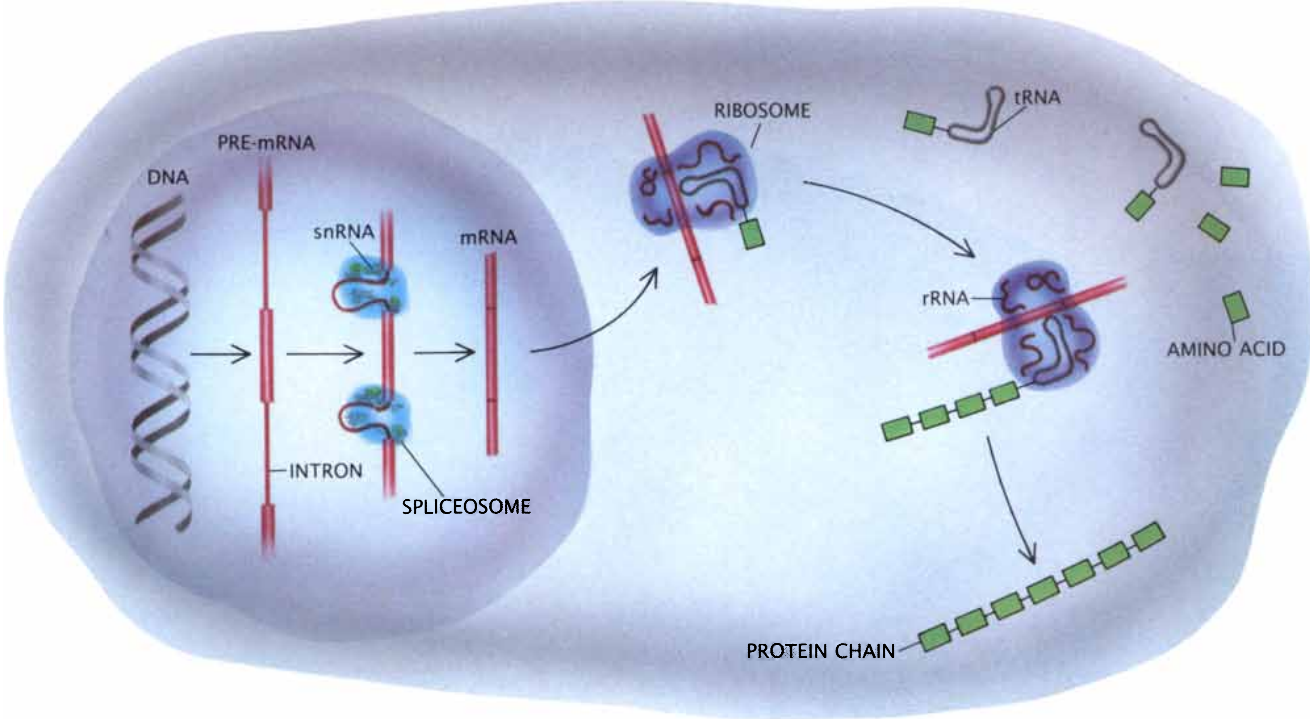
In higher organisms gene expression is complicated by the presence in DNA of noncoding regions called introns. Introns can be from 65 to 100,000 nucleotides long, and a single gene can harbor as many as 50 of them. Because DNA sequences are transcribed wholesale, with no discrimination between introns and the coding "exon" regions, the primary transcript is littered with segments of genetic nonsense. If protein synthesis were to begin on such a piece of "pre-mRNA," it would founder the moment it encountered an intron. Before mRNA transcripts are sent out into the cytoplasm, introns must therefore be removed and exons rejoined in the same order that they occupied in the DNA.

It is this crucial splicing procedure that snRNP's mediate. Their involvement in mRNA processing was established several years ago, and now the mechanism of their action is gradually coming to light. When I joined the faculty at Yale in 1970, however, the word "intron" had not been coined and notions of mRNA processing in the nucleus were entirely speculative. Initially my research was focused on the structure and function of RNA molecules in bacterial cells, but in the mid-1970's, during a sabbatical in the laboratory of Klaus Weber at the Max Planck Institute for Biophysical Chemistry in Göttingen, I decided to investigate the RNA in mammalian cells.

At that time the behavior of mRNA in the cells of higher organisms appeared to be quite mysterious. Experiments in many laboratories had uncovered the puzzling fact that only a small fraction of the RNA transcribed in the nucleus ever makes its way into the cytoplasm. Rather, most of it becomes degraded within a few minutes of synthesis. Furthermore, unusually long strands of RNA, dotted with proteins like beads on a string, had been described in 1968 by O. P. Samarina, G. P. Georgiev and their co-workers at the Academy of Sciences in Moscow. It was known that the protein beads, which were called heterogeneous nuclear ribonucleoproteins (hnRNP's), form soon after precursor mRNA



INSIDE A "SPLICEOSOME" the four most abundant snRNP's cooperate to excise a noncoding segment called an intron from a messenger RNA precursor (*red*). Here the snRNP's assist with the first step in the splicing procedure, in which two ends of the intron are brought into proximity. SnRNP's interact with mRNA directly through chemical bonds (depicted in this illustration as rungs on a ladder) or indirectly by means of protein intermediaries such as the violet protein on the right.



GENE EXPRESSION involves four kinds of RNA. Initially a messenger RNA (mRNA) precursor is transcribed from DNA; the "pre-mRNA" contains introns on which spliceosomes assemble. The spliceosomes, made in part of small nuclear RNA's (snRNA's), cut out the introns and the mRNA strand enters the cytoplasm. There a ribosome (containing rRNA) translates the mRNA into protein. Transfer RNA's (tRNA's) collect amino acids to build the protein chain, which is eventually released.

is transcribed from the parent DNA.

These hnRNP's were the object of my hapless search for antibodies. I reasoned that the binding of the protein "beads" might somehow be critical in determining which parts of the pre-mRNA were targeted for degradation and which would be exported to the cytoplasm. If I could obtain antibodies that reacted with those proteins and blocked their function, I could perhaps discover what that function was.

Ordinarily a biologist "raises" antibodies to a particular antigen by injecting the antigen, in this case hnRNP, into animals. The animals' immune systems respond by manufacturing antibodies against the invader, and the antibodies can then be harvested from the animals' blood serum. For seven months I injected hnRNP from rat liver into rabbits and chickens, but the animals did not cooperate. Apparently their immune systems did not recognize rat hnRNP as foreign. I could therefore conclude that hnRNP's are similar across vertebrate species, but when I returned to Yale in late 1977, I still had no antibodies against hnRNP.

In the meantime other workers had made the monumental discovery of introns and RNA splicing in mammalian cells. First reported by Phillip A. Sharp and his colleagues at the Massachusetts Institute of Technology and Thomas R. Broker, Louise T. Chow and their co-workers at the Cold Spring Harbor Laboratory, these phenomena explained why mRNA precursors are so long and why so much pre-mRNA is degraded. It became apparent that many genes are more intron than exon, and that "caps" and "tails"—distinctive chemical groups at each end of an RNA molecule—remain intact throughout splicing. The biochemical mechanism of splicing, however, was still not known.

Soon after I returned from Germany, my new M.D.-Ph.D. student Michael R. Lerner and I noticed an article in the British journal *Nature* discussing aberrant antibodies that appear in the serum of patients with mixed connective-tissue disease. This disease and its relatives rheumatism and lupus are known as autoimmune diseases because the immune system of a patient goes awry and produces antibodies against components of the patient's own cells. One particularly prevalent antibody in mixed connective-tissue disease was called anti-RNP. The authors did not know the precise nature of the antigen to which anti-RNP binds, but they knew that it consists of RNA and protein and is localized in the cell nuclei of many mammalian species.

This antibody, I thought, just might be the one I had tried, but failed, to raise against hnRNP. Lerner immediately suggested we contact the rheumatology group across the street from our laboratory to get the appropriate patient blood samples. Within hours John A. Hardin of the School of Medicine had given us sera containing anti-RNP as well as sera containing a related antibody called anti-Sm.

Our very first experiments told us that our motivating assumption was wrong. Neither anti-RNP nor anti-Sm reacted with hnRNP. Instead the antibodies bound to smaller nuclear components that seemed to be abundant and widespread in mammalian cells. We might have written off this latest batch of antibodies as another dead end, but because RNA-protein complexes were just beginning to be implicated in RNA processing in bacteria, we decided to probe the molecular identity of the antigens.

It took more than a year and countless discouraging experiments to isolate the targets of anti-RNP and anti-Sm. We finally found success with cultures of human cells whose RNA is labeled with radioactivity. From such cells it is possible to extract the antigens by means of standard immunoprecipitation techniques. First the cells are broken and a small amount of patient serum is added. The antibodies in the serum bind to their targets in the cell extract, and the resulting antibody-antigen complexes are selectively precipitated with the addition of special bacterial cell-wall preparations that combine with the antibodies. Tagged with a radioactive label, the RNA in the precipitates can then be isolated and classified according to size. Hence even though the antibodies actually bind to proteins in the precipitated particles, the particles are ultimately identified by their RNA components. The particles we isolated with this procedure are, of course, what we now call snRNP's.

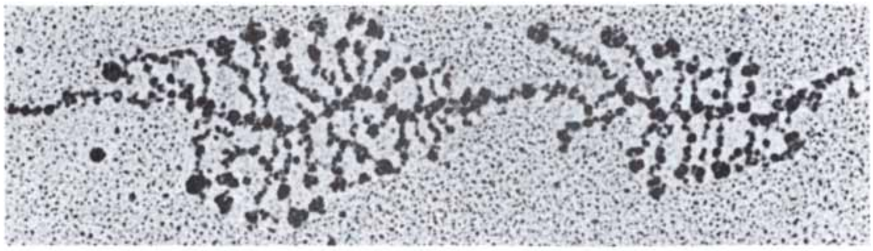
With this assay Lerner and I found five abundant species of snRNP RNA called U1, U2, U4, U5 and U6 (the *U* stands for uracil, a major chemical constituent of the snRNA). U4, U5 and U6 had never before been described, but U1 and U2 had already been characterized and their nucleotide sequences deduced by investigations in Harris Busch's laboratory at the Baylor College of Medicine in the early 1970's. Indeed, even before we had precipitated U1 we had examined its reported sequence and realized that the RNA adjacent to its cap could pair with the RNA near the splice sites of introns. The splice sites themselves have "consensus sequences": short sequences of nucleotides that are common to many different introns.

John Rogers and Randolph Wall of the University of California at Los Angeles had also noticed the striking complementarity between splice sites and the U1 sequence and proposed independently of us that U1 might be the agent that aligns the two ends of an intron during splicing. We now know this hypothesis was too simplistic, but it was on the right track.

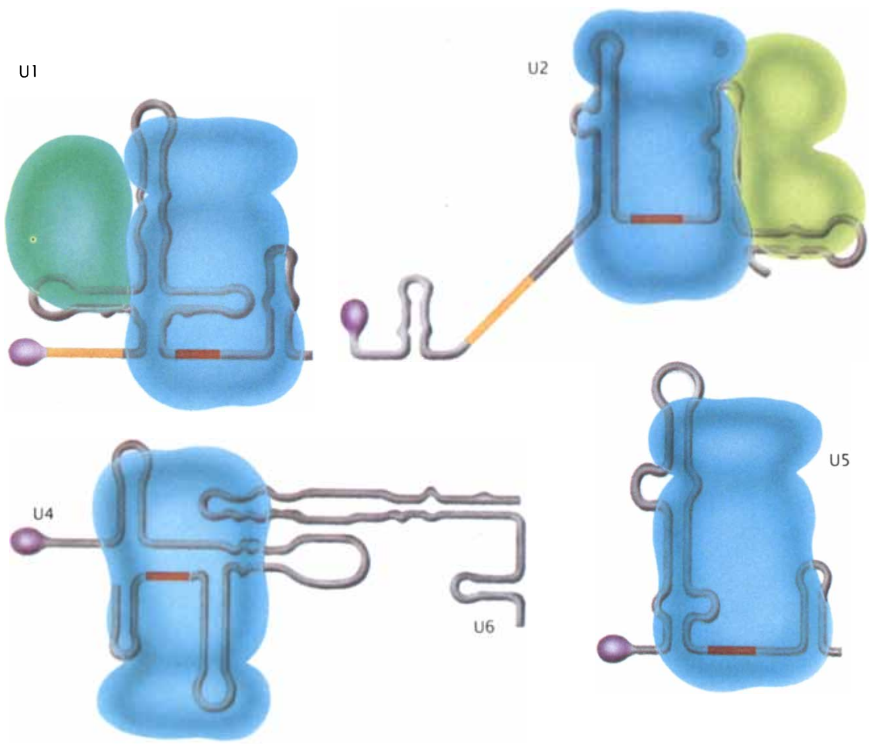
Direct evidence for the involvement of the U1 snRNP in splicing was contingent on the development of cell extracts that could carry out splicing in the test tube. That was achieved only slowly over the next few years, primarily in laboratories headed by Walter Keller at the German Cancer Research Center in Heidelberg, by Sharp at M.I.T., by Tom Maniatas at Harvard University and by John N. Abelson at the California Institute of Technology. Meanwhile we found out that the U1 and U2 snRNP's contain at least seven different proteins each; some of these are unique to one kind of particle and others are common to all. Later Carl Hashimoto, a graduate student in my laboratory, and Reinhard Lührmann's group at the Max Planck Institute for Molecular Genetics in Berlin recognized that U4 and U6 usually reside together in the same snRNP particle. The other snRNP's contain just one kind of RNA. Thus the five species of snRNP RNA occur in only four kinds of snRNP's.

With the snRNP's we purified for these analyses, another graduate student, Stephen M. Mount, attempted the first snRNP "protection" experiment [*see illustration on next page*]. Such experiments are designed to disclose which regions (if any) of the pre-mRNA a snRNP binds to. The procedure is straightforward: a solution of the pre-mRNA substrate is combined with a solution of the snRNP, antibodies are added to tag the snRNP, and an enzyme that degrades unbound mRNA is mixed with the solution. The region of mRNA to which the snRNP has bound is protected from degradation and can then be isolated and sequenced. The location of the protected fragment on the pre-mRNA strand can be determined by comparing its nucleotide sequence with that of the strand.

Mount constructed a model pre-mRNA substrate with one intron and flanking exons and then added U1 snRNP's. The U1 particles bound to and protected a sequence of about 17 nucleotides located at the "upstream" splice site in the intron (the site that is



SPLICEOSOMES IN ACTION look like large dots on the mRNA strands extending above and below the horizontal DNA. Smaller beads on the mRNA are the heterogeneous nuclear ribonucleoproteins (hnRNP's) that launched the author's studies. Yvonne N. Osheim of the University of Virginia School of Medicine made this micrograph.



FOUR KINDS OF snRNP'S and five species of small nuclear RNA have been implicated in mRNA splicing. The snRNA's (*gray*) are known as U1, U2, U4, U5 and U6. U4 and U6 reside together in a single snRNP particle. U6 is the only snRNA lacking the characteristic cap structure (*purple*). The snRNP's share a complex of proteins (*blue*); other, unique proteins (*green*) associate with U1 and U2 snRNA. The binding sites for the common proteins are shown in orange and sites for mRNA binding are in yellow.

closer to the cap of the mRNA strand than to the tail). The particles did not bind to the downstream site, and they did not cut or splice the pre-mRNA. The snRNP's lost their specificity and would not bind efficiently if they were not intact.

Mount's observations were verified by definitive experiments when splicing extracts became available in 1983. Collaborating with Sharp and his colleagues at M.I.T., we observed that antibodies specific for the U1 snRNP inhibited the in vitro splicing reaction. Keller and his colleague Angela Krämer, in collaboration with Lührmann's laboratory, decapitated the U1 RNA and thereby abolished splicing as well. The evidence was conclusive: U1 particles engage in splicing by binding to the upstream end of the intron, and the region of U1 RNA near the cap is essential to the process. Later, genetic experiments by my colleague Alan M. Weiner and his graduate student Yuan Zhuang would confirm our initial assumption that the nucleotides near the cap in U1 snRNA pair with the nucleotides at the upstream splice sites of introns.

One by one the roles of the other snRNP's came to light. In 1984 the elucidation of intermediate mRNA structures in splicing suggested another site at which a snRNP particle might bind. The studies revealed that splicing takes place in two steps [*see illustration on page 63*]. In the first step the upstream splice site is cleaved and coupled to a nucleotide near the downstream splice site, so that the intron ends up looking rather like a lariat. Then the two exons are joined and the intron is released in its lariat form, eventually to be degraded.

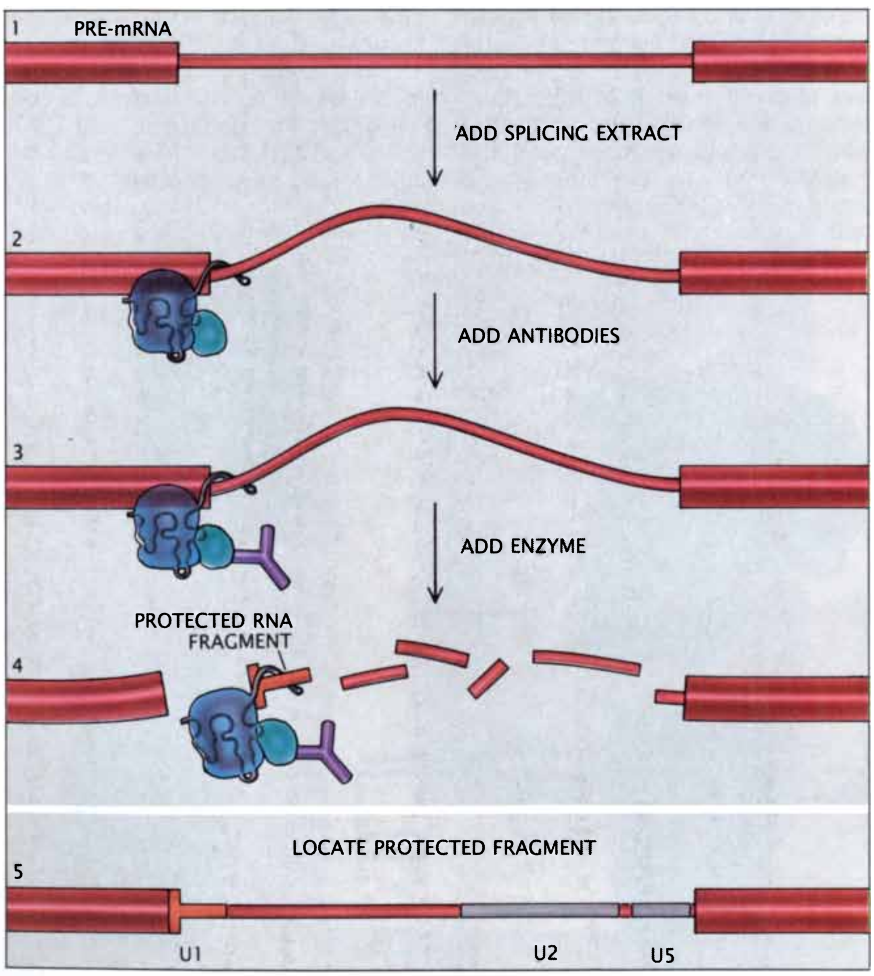
The intron branch point—the point at which the upstream end of the intron attaches to form the lariat—seemed a probable recognition site for a snRNP. My graduate student Benoit Chabot carried out protection experiments with rare patient antibodies specific for U2 snRNP's and found that the U2 particle binds to a 40-nucleotide intron fragment surrounding the branch point. Our hunch was right: the point does constitute a snRNP binding region, and the snRNP that binds is the U2 particle. At the same time Douglas L. Black in my laboratory and Adrian R. Krainer and Maniatis at Harvard showed that intact U2 snRNP's are required for splicing in the test tube.

Chabot also did protection experiments that identified the U5 snRNP as the particle binding to the downstream splice site. The binding, which involves a region about 15 nucleotides long, is probably mediated by a protein. Most recently Black, along with Susan M. Berget and Barbara L. Robberson of the Baylor College of Medicine, found that destroying U4 or U6 in extracts debilitates splicing. Hence the U4-U6 particle must also be necessary for splicing, but so far no one has been able to ascertain whether it interacts with the pre-mRNA strand or with other snRNP's during the reaction.

The lengths of the binding regions we did manage to identify agree nicely with the lower size limit observed for naturally occurring introns. These observations, as well as experiments in which intron segments are deleted, fix the minimum functional length of an intron at about 65 nucleotides. This figure roughly equals the sum of the sizes of the intron fragments to which U1, U2 and U5 bind in protection experiments. And so it appears that the geometry of snRNP binding dictates minimum intron length.

Other evidence has substantiated the emerging model of snRNP binding. The putative RNA-RNA interactions at U1 and U2 binding sites have been supported by recent genetic experiments manipulating the nucleotide sequences in both snRNA and the pre-mRNA. Although some of the studies examined yeast rather than mammalian cells, extrapolation seems justifiable. Not as much is known about the protein-RNA interaction involved in recognition of the downstream splice site. It also remains to be seen how many snRNP contacts are RNA-RNA interactions and how many act instead through proteins.

At the same time that we were amassing evidence of snRNP involvement in splicing, other data indicated the active splicing complexes are very large indeed. It turns out that snRNP's



PROTECTION EXPERIMENTS define the segments of mRNA to which a snRNP binds. First a pre-mRNA is added to a splicing extract (*1*), a cocktail of elements required for splicing in the test tube. SnRNP's in the extract bind to the pre-mRNA (*2*). Then antibodies that attach to a particular snRNP are added (*3*). Most of the mRNA is destroyed by an enzyme that degrades unbound mRNA (*4*), but the mRNA bound to a snRNP remains intact and is recovered by precipitating the antibody. Such experiments identified the upstream U1 and the downstream U2 and U5 binding sites (*5*).

New equipment using the latest technology in an easy-to-operate system will be provided by a new air traffic control system for the Federal Aviation Administration (FAA). In competition to develop the Advanced Automation System (AAS), Hughes Aircraft Company's version will include controller displays, advanced computers, communications networks, and software that can handle many tasks now completed manually. This dynamic new system will make possible direct, fuel-efficient flights, more efficient use of runways and airport "landing slots," fewer delays, and a safer overall system. To meet FAA requirements, the AAS is being designed for a downtime of less than three seconds per year.

Carried aboard a new satellite, positioned to detect storms threatening the East Coast, are two experiments. The Geostationary Operational Environmental Satellite (GOES) H, designed and built by Hughes for the National Oceanic and Atmospheric Administration, includes a space environment monitor (SEM) and an experimental receiver. The SEM assesses magnetic field strength and direction, solar x-ray fluctuations, and particles in its vicinity that make up solar wind and radiation belts around the Earth. The receiver will be used to aid in international search and rescue missions by monitoring radio distress signals from troubled ships or aircraft throughout most of North and South America. GOES H is in geosynchronous orbit above the Atlantic seaboard.

A single high-power microwave system will distribute 60 television channels in the San Francisco Bay area from one centrally located headend site. The 60-channel AML® high-power array, built by Hughes, will allow United Cable Television Company to interconnect San Francisco's East Bay, South Bay, and Peninsula areas. By using Hughes' AML microwave link, six communities will be served without the expense of hardline transportation trunks or duplicating headend equipment and building facilities. United will also cut operating costs by maintaining only one headend instead of six.

A new inspection tool can locate breaks and imperfections in optical fibers and visually inspect the continuity of individual fibers. The Hughes Glocater™ fiber-optic fault locator consists of an "enhanced" helium-neon laser, high-efficiency fiber coupler, and power supply. Designed for use during installation, re-arrangement, maintenance, and other procedures, the locator is attached to fibers being tested. In addition to its use in system acceptance testing of cables, jumpers and pigtails, the Glocater can serve as a tool for fiber-optic training centers. It can also be used on both single-mode or multimode fiber of any wavelength to verify continuity and provide positive end-to-end identification.

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

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

HUGHES
AIRCRAFT COMPANY

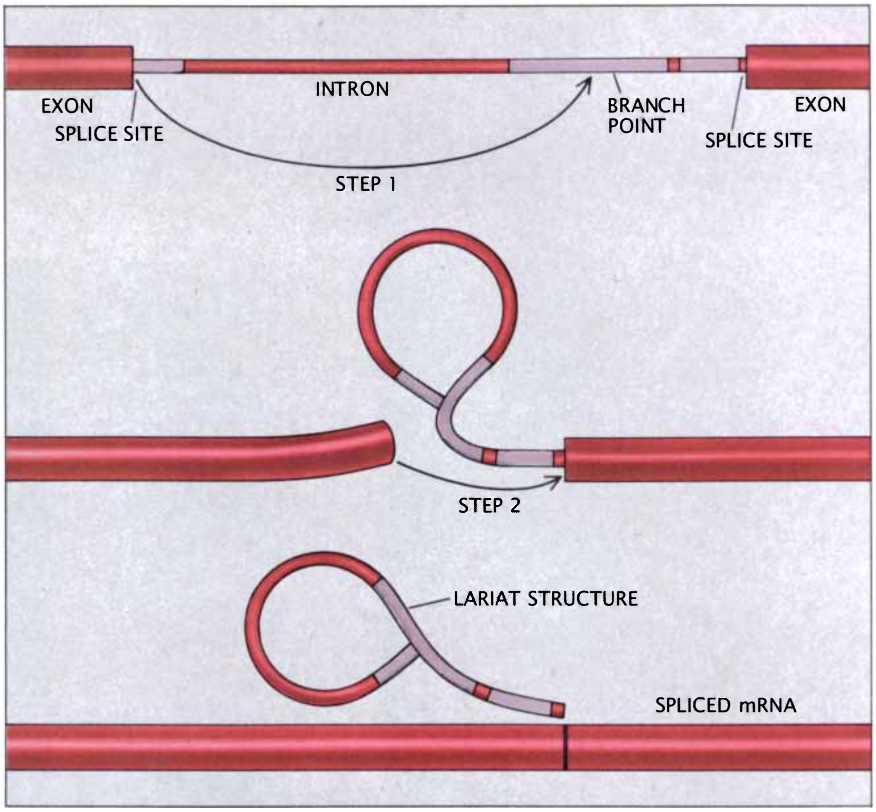
Subsidiary of GM Hughes Electronics

and their pre-mRNA substrate probably account for no more than a third to half of the mass of a spliceosome. The rest of the mass may represent protein factors required to assemble and disassemble the spliceosome complex; there is evidence that some of these proteins are the same as those in the hnRNP beads that provoked our initial investigations of snRNP's.

Even now very little is known about those beads, but spliceosomes are somewhat less mysterious. The size of spliceosomes answers at least one question regarding pre-mRNA splicing: it explains how a cell "knows" when mRNA is ready for export to the cytoplasm. Strands of mRNA that have many introns are retained in the nucleus until the last intron has been removed. Spliceosome complexes are large enough to prevent an mRNA transcript from slipping through the nuclear pores, and they do not disassemble until the splicing reaction is finished. The presence of even a single spliceosome on an mRNA strand could therefore keep the strand from escaping prematurely.

One of the most intriguing aspects of spliceosome function is that the entire assemblage, rather than any individual component, seems to be responsible for the catalysis of the splicing reaction. Hence damaging any one of the four known snRNP's cripples splicing, but none of the snRNP's can execute splicing on its own. Perhaps the snRNP's and other spliceosome components cooperate to arrange the intron in a configuration that encourages "self-splicing." In some organisms precursor mRNA forms lariat structures and in effect splices itself without the aid of protein or other RNA factors; it could be that spliceosomes merely provide missing intron sequences necessary to achieve correct alignment for such "spontaneous" splicing.

Many questions about spliceosome function still need to be answered. Splicing is subject to regulation: the same pre-mRNA can be spliced in different ways at different developmental stages or in different tissues. As yet no one knows whether or how snRNP's or other spliceosome factors could mediate such regulation. The pathway of spliceosome assembly and the total number of components involved are also unclear. The agent that holds the upstream exon until it is joined to its downstream partner has not been identified; neither has the mechanism that ensures the fidelity of splicing. Could it be that spliceosome operation includes "proofreading" steps?



LARIAT STRUCTURE forms in the course of intron splicing. Splicing is a two-step process mediated by snRNP's, which are not shown here. In the first step the upstream splice site is cut and attached to the U2 binding site, giving rise to the intron lariat. In the second step the exons are joined and the intron is released.

My discussion of snRNP's has been limited to the four most abundant particles and their five RNA species. In the course of recent investigations, at least 10 other snRNP's have been identified in mammalian cell nuclei. These particles are not nearly as prevalent as the spliceosome snRNP's. My graduate student Kimberly L. Mowry demonstrated that one of them, a particle containing the snRNA called U7, is involved in processing the downstream end of a subclass of mRNA transcripts. It seems likely that the other low-abundance snRNP's take part in mRNA processing events as well.

It is important to reiterate that all the snRNP's we have studied thus far have been isolated using antibodies from patients with rheumatic diseases; consequently even our most arcane efforts have had an impact on the understanding and treatment of these diseases. Furthermore, snRNP's are not the only small ribonucleoproteins. Similar protein-RNA complexes have been identified in the cytoplasm and in a special nuclear region called the nucleolus. These particles are known as scRNP's ("scyrps") and snoRNP's ("snorps") respectively. Although several of them have been characterized structurally, the function of only one has been established. Indeed, considering the relative variety of snRNP's and other small RNP's, it is probably safe to assume that subsequent research will disclose an even greater diversity of roles for these particles in living cells.

FURTHER READING

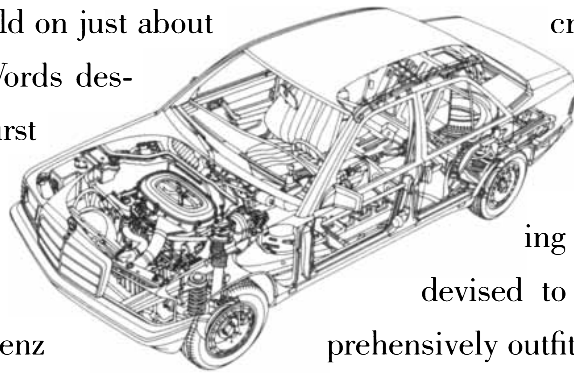
SPLICING OF MESSENGER RNA PRECURSORS. Richard A. Padgett, Paula J. Grabowski, Maria M. Konarska, Sharon Seiler and Phillip A. Sharp in *Annual Review of Biochemistry,* pages 1119-1150; 1986.

THE ROLE OF SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES IN PRE-MRNA SPLICING. Tom Maniatis and Robin Reed in *Nature,* Vol. 325, No. 6106, pages 673-678; February 19, 1987.

IDENTIFICATION OF THE HUMAN U7 SNRNP AS ONE OF SEVERAL FACTORS INVOLVED IN THE 3′ END MATURATION OF HISTONE PREMESSENGER RNA'S. Kimberly L. Mowry and Joan A. Steitz in *Science,* Vol. 238, No. 4834, pages 1682-1687; December 18, 1987.

SMALL NUCLEAR RIBONUCLEOPROTEIN PARTICLES: STRUCTURE AND FUNCTION. Edited by M. L. Birnstiel. Springer-Verlag, 1988.

THE MERCEDES-BENZ 190 CLASS: THE AUTOMOBILE EVERY SPORTS SEDAN WANTS TO BE WHEN IT GROWS UP.



It has been written that the Mercedes-Benz 190 Class "...will challenge some of the finest sporting sedans in the world on just about any road you choose."* Words destined to leap to life the first time you choose a likely road and storm it in a 190 Class sedan.

But Mercedes-Benz engineers have never believed in building one-dimensional automobiles. The 190 Class repays your investment—and separates itself from today's legions of single-minded sports sedans—by promising infinitely more for your money than occasional moments of driving nirvana.

Such as a permanent sense of automotive integrity. Proceeding outward from a steel vault of a body structure, through technology that system for system may be the most fully developed in the automotive world, to fit and finish that prompted *Car and Driver* to ask, "Why is it that Benzes fit together better than anything else in the world?"

Producing, in turn, an aura of driving civilization unique not only among sports sedans but among all automobiles. Matching such exploits as triple-digit test-track maximums with velvety running smoothness. Blending sports car cornering power with a gentle ride. So reducing friction between airstream and automobile at cruising speeds that wind noise is all but obsolete.

You inhabit a cabin styled not to indulge driving fantasies but ergonomically devised to serve driving needs. Comprehensively outfitted. Soothingly restful. And as its built-in SRS Supplemental Restraint System with driver's-side air bag attests, soothingly safety-minded.

And having joined the driving elite as a Mercedes-Benz owner, you join an ownership elite that affords pleasures no sports sedan can match. As in long-term ownership satisfaction so satisfying that more Mercedes-Benz owners would buy again than would owners of any other make. Which may help explain why Mercedes-Benz cars as a line—and regardless of age—have retained a higher percentage of original value than any other make sold in America.

The 190 Class: proof that your automotive judgment can be very grown up, while your automotive enthusiasm remains very young.

Engineered like no other car in the world

Road & Track

Particle Accelerators Test Cosmological Theory

Is there a limit to the number of families of elementary particles? Debris from the big-bang origin of the universe suggests there is, and accelerators are reaching the energies required to confirm the limit

by David N. Schramm and Gary Steigman

Over the past decade two subfields of science, cosmology and elementary-particle physics, have become married in a symbiotic relationship that has produced a number of exciting offspring. These offspring are beginning to yield insights on the creation of spacetime and matter at epochs as early as 10^{-43} to 10^{-35} second after the birth of the universe in the primordial explosion known as the big bang. Important clues to the nature of the big bang itself may even come from a theory currently under development, known as the ultimate theory of everything (T.O.E.). A T.O.E. would describe all the interactions among the fundamental particles in a single bold stroke.

The confluence of cosmology and particle physics is even changing the way science is done. Traditionally astronomy has been an observational rather than an experimental science, in which passive observations were made with telescopes and actively controlled experiments have been virtually unknown. Traditionally the tools of particle physics have been high-energy accelerators. Now that cosmology has begun to make predictions about elementary-particle physics, it has become conceivable that those cosmological predictions could be checked with carefully controlled accelerator experiments. It has taken more than 10 years for accelerators to reach the point where they can do the appropriate experiments, but the experiments are now in fact in progress. The preliminary results confirm the predictions from cosmology.

It appears therefore that cosmology has become a true science in the sense that ideas not only are developed but also are being tested in the laboratory on time scales that are shorter than a scientist's lifetime. This is a far cry from earlier eras in which cosmological theories proliferated and there was little way to confirm or refute any of them other than on their aesthetic appeal. Conversely, telescopes may eventually be employed to test ideas from fundamental physics, such as proposals for a T.O.E. Indeed, tests of theories involving interactions of particles with enormous energies could very well have only one available laboratory: the big bang itself.

DAVID N. SCHRAMM and GARY STEIGMAN have been pioneers in developing the interface between cosmology and particle physics. Schramm, who is Louis Block Professor of Physical Sciences at the University of Chicago, received his B.S. in 1967 at the Massachusetts Institute of Technology and his Ph.D. in physics in 1971 from the California Institute of Technology. He enjoys athletics, having climbed mountains throughout the world and written about his experiences for various outdoor magazines. Steigman is professor of physics and astronomy at Ohio State University. He got a B.S. in physics at the City College of the City University of New York in 1961 and a Ph.D. from New York University in 1968. In his spare time he likes to hike, ski, scuba dive and dance.

Among the most exciting offspring produced by the marriage of cosmology and particle physics is the first-begotten. The universe would not look the same if there were too many different fundamental types of elementary particles. In other words, from cosmology it is inferred that the number of fundamental particles must be small. This specific prediction emerged from our analysis of the nuclear reactions that occurred when the universe was roughly one second old. We took the bold step of converting cosmological quantities (such as average energy density) into quantities of interest in particle physics (such as the number of fundamental particles).

Our prediction remains a significant one, but it was particularly powerful at the time it was put forward because the prevailing attitude then was that whenever a particle accelerator attains higher energies, novel particles will be discovered. Theories from particle physics had set no significant limits on how many types of fundamental particles can exist. There appeared to be no end in sight. Now a theoretical prediction made from cosmological considerations contradicted that empirical deduction. As time passes the prediction has continued to hold up and even looks stronger. The number of elementary particles must be limited, otherwise the universe would be different from the one we know.

To explain how cosmological considerations set limits on the number of types of elementary particles, we must first give a brief overview of particle physics. Over the past half century experiments at particle accelerators have established that fundamental particles can be separated into two broad classes known as fermions and bosons (which are named for the Italian-American physicist Enrico Fermi and the Indian physicist S. N. Bose). Fermions are the particles that make up matter and bosons are the carriers of the forces between particles. Fermions in turn are divided into two subclasses: quarks and leptons. The word "quark" comes from a curious line in James Joyce's *Finnegans Wake,* "Three quarks for Muster Mark!" and "lepton" comes from the Greek *leptos,* meaning small particle. Quarks are the constituents of neutrons, protons and related particles called hadrons. Leptons, if they carry an electric charge like that of the electron, can orbit the nucleus to make up atoms; if they are uncharged, like the leptons called neutrinos, they can traverse the entire earth without interacting with anything. Each particle also has an antiparticle with the same mass and lifetime and opposite electrical properties.

The interactions among the various particles are governed by four fundamental forces, each of which is carried by a separate boson or set of bosons. The photon, or quantum of light, carries the electromagnetic force, which couples electric charges; the graviton carries the gravitational force, which couples masses; eight gluons carry the strong nuclear force, which couples quarks, and the intermediate vector bosons carry the weak nuclear force, which is responsible for certain nuclear decays. At present it appears that all interactions in the universe can be reduced to combinations of these four interactions.

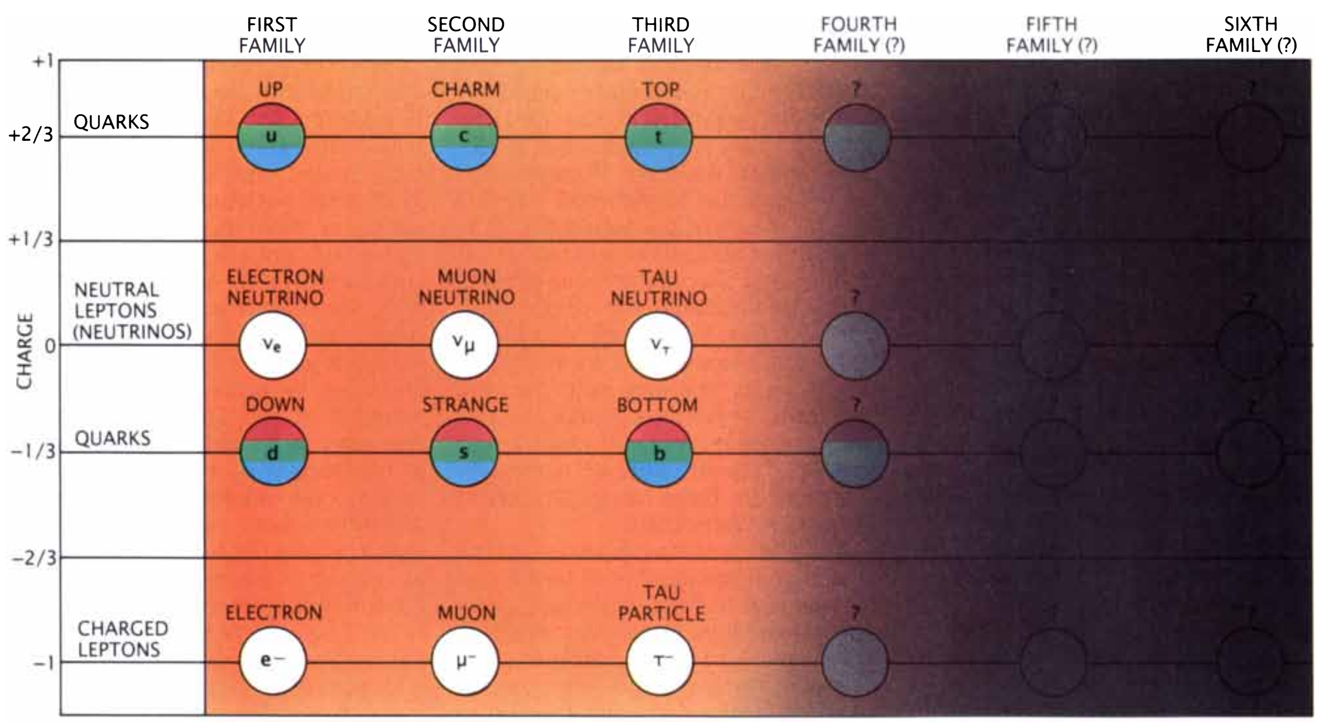
One of the most exciting developments in 20th-century physics has been the proof that at higher energies, or temperatures, the four forces begin to unify. In particular, experiments at CERN, the European laboratory for particle physics, have shown that the weak and the electromagnetic forces merge into a single electroweak force at energies greater than 100 billion electron volts (100 GeV). Such an energy corresponds to the temperature of the universe some 10^{-10} second after the big bang, which was more than four trillion times as great as room temperature. The CERN findings have raised hopes that the strong force will merge with the electroweak force at roughly 10^{15} GeV in some grand unified theory (G.U.T.) and that by 10^{19} GeV the force of gravity will join in to yield a T.O.E.

The energies needed to test G.U.T. and T.O.E. proposals are enormous compared with the energies accessible to existing particle accelerators. The world's largest accelerator, the Tevatron at the Fermi National Accelerator Laboratory, has a circumference of four kilometers and is just reaching energies of two trillion electron volts, or 2,000 GeV. An accelerator similar to the Tevatron but scaled up to energies at which G.U.T. proposals could be tested would stretch to the nearest stars, and a T.O.E. machine would reach to the galactic center. Both machines are beyond even the most optimistic science budgets. This simple fact has been one of the driving forces behind the attempts to utilize cosmological observations to test predictions of particle physics.

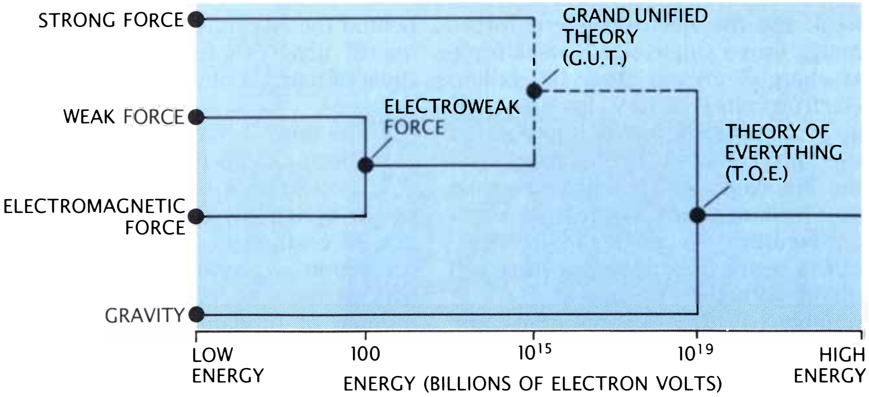
The flow of information has also begun to go the other way: the accelerators of particle physics are being employed to test a prediction of cosmology. The cosmological prediction we have been concerned with pertains to setting limits on the number of fundamental particles of matter.

It appears that there are 12 fundamental particles, as well as their corresponding antiparticles. Six of the fundamental particles are quarks, and they carry the whimsical names of "up," "down," "charm," "strange," "top" (or "truth") and "bottom" (or "beauty"). All the quarks have been discovered except the top quark; although theoretical arguments for the existence of that quark are strong, experimental evidence is at the moment absent. The other six fundamental particles are leptons: the electron, the muon, the tau particle, and three neutrinos associated with each of them, the electron neutrino, the muon neutrino and the tau neutrino.

The 12 particles are grouped in three families, each family consisting



FUNDAMENTAL CONSTITUENTS OF MATTER, called quarks and leptons, are grouped into families consisting of two types of each class of particles. The particles can be distinguished by their electric charge, among other properties. At present there are three families made up of 12 known quarks and leptons. All ordinary matter is composed of members of the first family. (The proton, for instance, consists of two "up" quarks and one "down" quark.) Theories from particle physics offer few predictions about how many families should exist; in principle the number could be infinite. Cosmological theories, however, suggest an upper limit of four families. A test of this limit is now being carried out at several particle accelerators.



FOUR FORCES OF NATURE account for all known interactions among elementary particles. The strong force couples quarks, the weak force is responsible for certain nuclear decays, the electromagnetic force couples electric charge and gravity couples masses. It is thought the four forces were once unified at the higher energies characteristic of the universe soon after the big bang, and indeed a theory that unifies the weak force and the electromagnetic force has already been verified. A grand unified theory (G.U.T.) would unify these forces with the strong force; a theory of everything (T.O.E.) would describe all four forces as aspects of a single force.

of four members. The first family is made up of the up and down quarks, the electron and the electron neutrino; the second family consists of the charm and strange quarks, the muon and the muon neutrino, and the third family of the top and bottom quarks, the tau particle and the tau neutrino. All ordinary matter is made of members of the first family. The proton, for instance, consists of two up quarks, each bearing 2/3 of a unit of positive electric charge, and a down quark, which carries a charge of 1/3 of a unit of negative electric charge. The neutron consists of two down quarks and an up quark. Every atom is simply a hard kernel of tightly bound protons and neutrons surrounded by a cloud of electrons.

Since all ordinary matter is made of members of the first family, one of the great mysteries of particle physics today is why the other families exist and how many of them there are. As the late I. I. Rabi said when the muon was discovered, "Who ordered that?" Based on the trend of more particles being found as the energy of particle accelerators is increased, one might expect families to continue to proliferate. Actually G.U.T. proposals have said little about the total number of families. For example, the first G.U.T. model to gain popularity in the mid to late 1970's, SU(5) (for special unitary group with five degrees of freedom), can have any number of families.

There is a good reason for having at least three families, however. M. Kobayashi of the Japanese proton accelerator, K.E.K., and T. Masakawa of the University of Tokyo have pointed out that the asymmetry between matter and antimatter observed in 1964 by Val L. Fitch of Princeton University and James W. Cronin of the University of Chicago is best understood if there are at least three families of elementary particles. The asymmetry may provide the explanation for the observed excess of matter over antimatter in the universe, thereby enabling matter to exist [see "A Flaw in a Universal Mirror," by Robert K. Adair; SCIENTIFIC AMERICAN, February]. Having at least three families is not quite useless.

Yet one would still like to know precisely how many families of quarks and leptons there are. If quarks and leptons are the fundamental building blocks of nature, one would like to be acquainted with all the components. If it appeared that the number of families were unlimited, one would question whether quarks and leptons are truly fundamental. Just as atoms are made up of protons, neutrons and electrons, so perhaps quarks and leptons are made up of still smaller entities [see "The Structure of Quarks and Leptons," by Haim Harari; SCIENTIFIC AMERICAN, April, 1983].

It is now clear that an answer to the question of the number of families of quarks and leptons could well come from cosmology. Cosmology suggests there must be a finite number of families and, further, limits the possible range to small values: only three or at most four families exist.

The prediction of the limit to the number of families is based on evidence garnered by observing the debris from the greatest accelerator experiment of all, the big bang. The big-bang model of the universe began as one of two rival theories of cosmology dominating discussion in the 1950's and early 1960's. The other theory was called the steady-state theory. Both were developed to account for Edwin P. Hubble's discovery in 1929 that the universe is expanding. The big-bang model holds that the universe was once hot and enormously dense and that as it has expanded it has cooled and become less dense. The steady-state theory holds that matter is continuously created, so that as the universe has expanded, its density has remained constant.

In the 1960's the big-bang model received several observational boosts, and by the 1970's it was the clear winner. The most publicized of these boosts was the Nobel Prize-winning discovery made by Arno A. Penzias and Robert W. Wilson of the Bell Telephone Laboratories. If the big-bang model is correct, at one time the universe would have been sufficiently dense and so hot that the matter in the universe would have generated a characteristic spectrum of thermal radiation. According to the steady-state theory, on the other hand, the density of the universe has always been what it is today, and so the universe never existed in a dense, hot state. Hence there should be no thermal radiation. Penzias and Wilson discovered a microwave background radiation that is consistent with the hot, dense scenario expected with the big bang.

The strongest support for the big-bang model comes from studies of primordial nucleosynthesis: the formation of the elements. Temperatures nearly 100 million times as great as room temperature are needed to forge many elements from protons and neutrons; such temperatures would have occurred about one second after the big bang. By measuring the relative abundances of elements, therefore, one can probe conditions as far back as one second after creation. In comparison, the microwave background radiation serves as a probe of the universe back only to 100,000 years after creation, when photons last scattered with matter at temperatures of some 3,000 degrees Kelvin (degrees Celsius above absolute zero), or about 10 times room temperature.

We shall delve into the details of big-bang nucleosynthesis below, since they not only help to establish the big bang but also lead to the connection with particle physics. But it is first worth noting that the theory of big-bang nucleosynthesis has predicted the abundances of several light elements and their isotopes including helium 3, helium 4, deuterium (the heavy isotope of hydrogen) and lithium 7. The predicted abundances span almost 10 orders of magnitude. Observations appear to verify all these predictions in quantitative detail.

The impressive agreement between the theoretical predictions of big-bang nucleosynthesis and the astronomically observed abundances of the light elements provides a bonus. Agreement between theory and observation occurs for a value of the density of protons and neutrons that is completely consistent with the density determined from studies of the dynamics of luminous matter in the universe. The predictions based on the evolution of the universe during the first 1,000 seconds after the big bang are consistent with observations made some 10 billion years later.

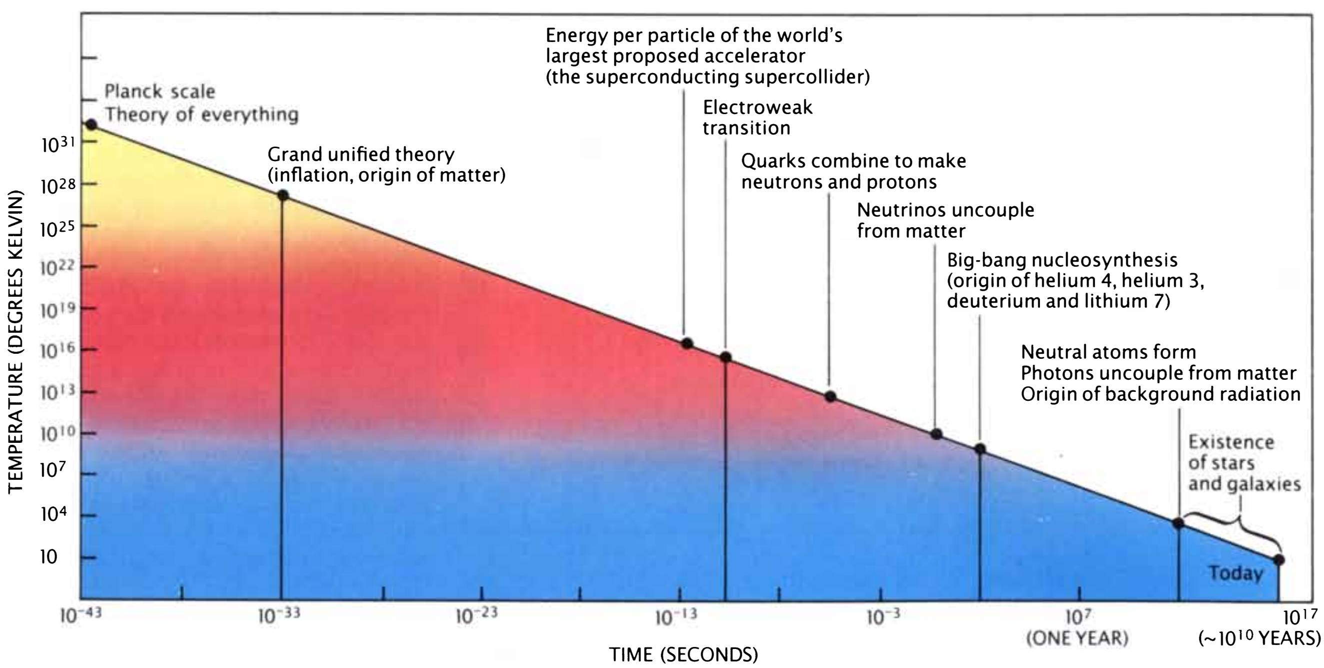
Physicists now seem to have a *quantitative* understanding of the behavior of the universe back to the time of big-bang nucleosynthesis. This detailed understanding has provided the confidence needed in attempting to push back to even earlier times appropriate for a G.U.T. or a T.O.E.

The power of the theory of big-bang nucleosynthesis derives from the fact that essentially all the input into the relevant equations is well known from laboratory experiments. In particular, the temperatures at which big-bang nucleosynthesis is thought to have occurred correspond to energies that are easily explored with relatively low-energy accelerators such as Van de Graaff generators. As a consequence the behavior of atomic nuclei under the conditions of big-bang nucleosynthesis is not a matter of guesswork; it is precisely known.

To calculate what happens, all one has to do is follow the evolution of a gas of neutrons and protons in an expanding and cooling universe. Since neutrons and protons are collectively referred to as nucleons, physicists call such a gas a nucleon gas. At temperatures much greater than 10 billion degrees K., which correspond to times well before the first second after the big bang, the protons and neutrons were in equilibrium and were present in equal numbers. These temperatures were too hot to allow protons and neutrons to fuse into more complex nuclei. Collisions with electrons and positrons (antielectrons) and neutrinos and antineutrinos changed neutrons into protons and protons into neutrons at approximately equal rates. The neutron is slightly heavier than the proton, and so neutrons change into protons more easily than protons change into neutrons. When the energies were very high, however, the mass difference had a negligible effect.

As the temperature of the universe fell to 10 billion degrees K., the mass difference became more significant, and the ratio of neutrons to protons dropped from one to less than a third. By the time the universe reached a billion degrees the ratio was slightly below a seventh. At that point the temperature was cool enough to allow protons and neutrons to begin to fuse into the simplest complex nucleus: deuterium, which consists of a single proton and neutron. Interactions of deuterium with other protons and neutrons produced tritium (a proton and two neutrons) and helium 3 (two protons and a neutron). These nuclei in turn interacted to produce helium 4 (two protons and two neutrons). Since helium 4 is much more tightly bound than any other light nucleus, the flow of reactions converted almost all the neutrons that existed at a billion degrees into helium 4. A small amount of beryllium 7 (four protons and three neutrons) and of lithium 7 (three protons and four neutrons) was produced when helium 4 interacted with helium 3 and tritium respectively. In short, the big-bang nucleosynthesis is believed to have generated helium 4 with traces of deuterium, helium 3 and lithium 7.

The flow essentially ceased at helium 4 because no stable nuclei are produced when a helium-4 nucleus interacts with a proton, a neutron or another helium-4 nucleus. The majority of the other elements were pro-



THERMAL HISTORY OF THE UNIVERSE, starting 10^{-43} second after the big bang and continuing to the present, shows that most of the helium 4, helium 3, deuterium (heavy hydrogen) and lithium 7 in the universe was synthesized approximately a minute after the big bang. Heavier elements were forged tens of millions to billions of years later in the interior of stars.

duced inside stars, which have densities sufficient to allow three helium-4 nuclei to combine to make carbon 12.

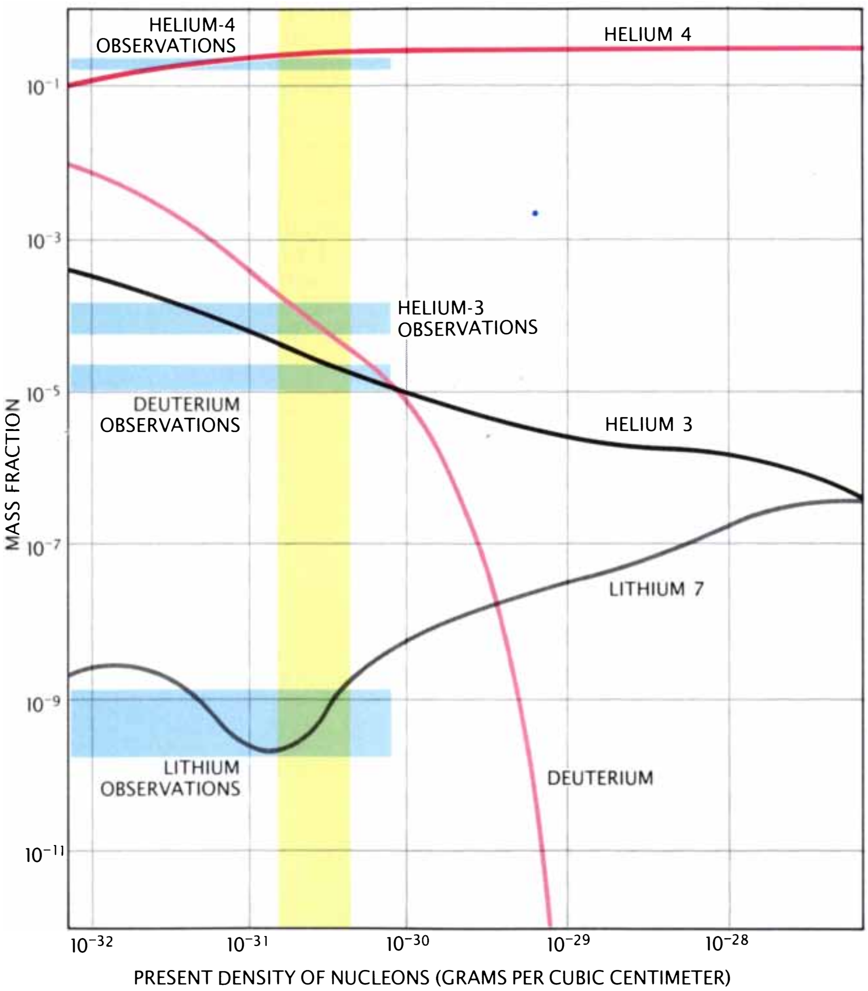
The abundances of the light elements predicted by the theory of big-bang nucleosynthesis, as we have mentioned, agree quite nicely with the observed abundances. According to the theory of big-bang nucleosynthesis, matter began to coalesce when the ratio of neutrons to protons was a seventh. Because virtually all neutrons were swept up into helium-4 nuclei (which contain equal numbers of protons and neutrons), the abundance of helium 4 should account for about a fourth of the total mass of ordinary matter. Actually the observed abundance of helium in galaxies, including our own, ranges from about 20 to 30 percent. The predicted abundances of deuterium, helium 3 and lithium 7, which range from less than one part in 10,000 to as little as one part in 10 billion, also match the abundances that are observed.

How does the theory of big-bang nucleosynthesis set limits on the allowed number of families of elementary particles? Quite simply, if the number of families exceeded three or four, the predicted abundance of helium 4 would exceed the observed abundance.

The reason such a statement can be made is that the predicted abundances of the light elements depend on only two variables: the density of nucleons and the density of radiation in the universe. Unfortunately the value of neither variable is precisely known. It turns out, however, that only a small range of values for each of the two variables produces abundances that are consistent with the observed abundances. By substituting the values of the observed abundances into the appropriate equations, one can determine what the density of nucleons and of radiation must be. Knowing those values leads to a number of interesting conclusions.

The density of a nucleon gas increases in proportion to the cube of the temperature, so that when the universe was twice as hot as it is today, it was eight times as dense in nucleons. By determining what the nucleon density must have been during nucleosynthesis to produce the abundances of deuterium, helium 3 and lithium 7 seen today, one can calculate the present nucleon density. It is between 2×10^{-31} and 5×10^{-31} gram per cubic centimeter. Such a range of values is consistent with the density of luminous material inferred from studies of the dynamics of galaxies and clusters of galaxies, but it is at least 10 times less than the estimated density of gravitational mass needed to close the universe, or halt the big-bang expansion. If the universe is closed, additional non-nucleonic matter is needed. The search for such matter, which would be dark, or invisible to telescopes, and made up of something other than nucleons, is now under way [see "Dark Matter in the Universe," by Lawrence M. Krauss; SCIENTIFIC AMERICAN, December, 1986].

An analogous line of reasoning for the radiation density provides the constraints on the number of families of elementary particles. The radiation density is important for big-bang nucleosynthesis because it controlled the rate of expansion of the universe during that time. The radiation density at any time is proportional to the number of types of radiation or, equivalently, the number of types of particles moving at nearly the speed of light. During big-bang nucleosynthesis it is thought there were nine types of relativistic particles: the photon (of course), the electron and the positron, the electron neutrino, the muon neutrino and the tau neutrino and their three antiparticles. Neutrinos are either massless or have such a small mass that they travel at nearly the speed of light; the electron and the positron have a mass that is small enough so that at the high energies achieved at the time of primordial nucleosynthesis they too would have traveled at close to the speed of light. The radiation density associated with the nine kinds of particles leads to conditions that would have been just right to produce the observed abundance of helium 4. (Interestingly, the



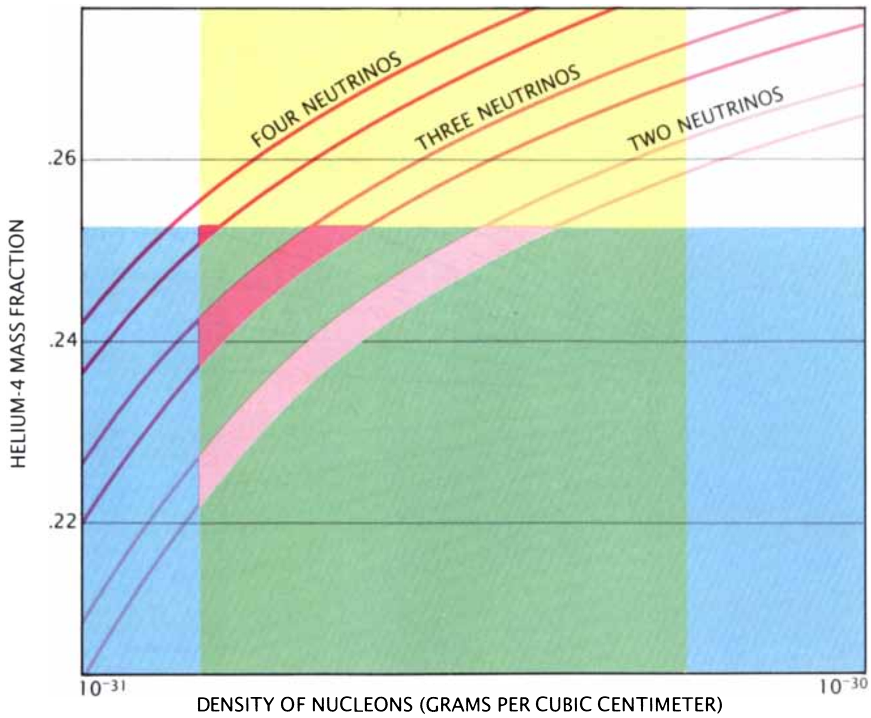
PREDICTED ABUNDANCES of helium 4, helium 3, deuterium and lithium 7 in the big-bang model of the universe (*curves*) closely agree with the observed abundances (*shaded horizontal bands*). The predicted abundances change as a function of the density of nucleons (protons and neutrons) at the time of the big bang; the shaded vertical band indicates the best cosmological estimate of that density today. The close agreement is one of the strongest arguments supporting the big-bang model.

helium-4 density is virtually independent of the nucleon density, a fact that Fred Hoyle and his colleague Roger Taylor first noted in the 1960's and that was later verified more rigorously by a number of investigators.)

In calculating the abundance of helium 4 we have allowed for photons, electrons and the three known neutrino species, as well as their antiparticles. If other families of fundamental particles exist, the calculation would have to be modified. The only new family members that would affect the calculation are neutrinos, since in any family beyond the first, they alone are light enough to travel close to (or at) the speed of light. Presumably each new family beyond the third would contribute one neutrino and a corresponding antineutrino.

If the gas from which the universe was made had contained additional neutrinos, its density of radiation would have been greater. As a consequence the cosmological expansion during the period of big-bang nucleosynthesis would have been more rapid. It so happens that the ratio of neutrons to protons is quite sensitive to the rate of cosmological expansion. A higher rate of expansion would have meant that neutrons would have had less time to change into protons, and so more neutrons would have been left: the ratio of neutrons to protons would have been greater. Because neutrons were quickly swept up into helium-4 nuclei, the abundance of helium 4 would be greater.

Helium 4 is the most abundant of the nuclei synthesized in the big bang, and so it is the element observers are able to measure with the greatest accuracy. Since helium 4 is produced in stars, however, it is important to estimate what part of the helium observed in astronomical objects is primordial—from the big bang—and what part was generated by stars after the big bang. In collaboration with John S. Gallagher of the Lowell Observatory, we have found that the additional amount of helium 4 produced by stars can be tracked by measuring the carbon content of objects; stars that make helium also make carbon, so that the helium abundance increases with the carbon abundance. This allows one to "subtract" the contribution of stars to the helium-4 abundance from the observed abundance in order to infer the true primordial abundance. We have determined that the highest allowed value of the primordial abundance of helium 4 is slightly less than 25 percent.

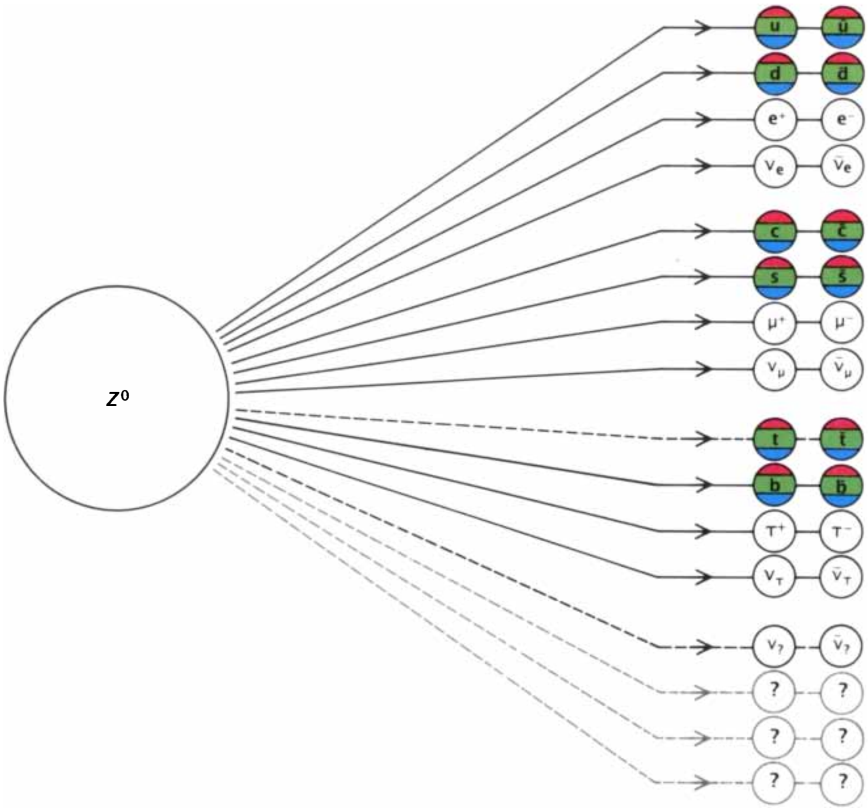


HELIUM-4 ABUNDANCE suggests there are at most four families of elementary particles. The three curves represent an enlargement of the part of the helium-4 curve lying within the shaded vertical band in the illustration on the opposite page; the narrow curve in that illustration resolves into the three broad curves. The bottom curve shows the helium-4 abundance predicted if there were two families of particles. The middle curve shows the abundance predicted for three families, and the top curve shows the abundance predicted if there are four families. The predicted abundances of helium 4 for two and three families of particles are well within the region defined by helium-4 observations and estimates of nucleon density (*green region*). A fourth family would produce an abundance that would be very close to the allowed extremes. Clearly there is no room for any more than four families.

Our calculations show that such a low abundance could have emerged from the big bang only if there is no more than one additional kind of neutrino and corresponding antineutrino. If more neutrinos had existed, the radiation density would have been so great that the amount of helium 4 produced during the big-bang nucleosynthesis would be greater than the observed abundance. In other words, the total number of families of elementary particles is three or at most four. Our finding suggests that all the fundamental families of elementary particles may already have been discovered. This basic argument was made by us more than 10 years ago in collaboration with James E. Gunn of Princeton University; subsequently the measurements of the helium-4 abundance and the estimate of its primordial value have improved significantly. What makes the argument particularly exciting and timely today is that accelerators are now beginning to check it.

Searching for new kinds of neutrinos has always been a difficult and tedious procedure. Traditionally the only means of discovering a neutrino has been to produce its associated charged lepton first. The drawback of the approach is that even though neutrinos are quite light or even massless, a great deal of energy is necessary to generate the associated leptons; the heavier the mass of the associated lepton is, the greater the energy of the accelerator must be to generate it. The tau particle has such a high mass, for instance, that Martin L. Perl and his collaborators at the Stanford Linear Accelerator Center (SLAC) needed several billion electron volts of energy, corresponding to temperatures exceeding 10^{13} degrees K., to find it. With such an approach one can always argue that the next lepton is just beyond the limits of the current accelerators.

The new way of searching for neutrinos grew out of the CERN experiments mentioned above, which have shown that the weak and the electromagnetic forces are actually aspects of a single electroweak force. In 1983 the CERN investigators, a team of hun-



TESTS OF FAMILY NUMBER are under way at various particle accelerators around the world. A particle called the Z^0 boson can decay into quarks and leptons, each particle paired with its antiparticle. (An antiparticle has the same mass as its corresponding particle but opposite electrical properties; it is often represented by the symbol for the corresponding particle with a horizontal bar over it.) The allowed decays are shown by solid lines. The more families there are, the more decay routes the Z^0 boson should have and the shorter its lifetime should be. Measurements of the lifetime should therefore indicate the number of families. Current estimates place the limit at five particle families. Future experiments should yield more precise results.

dreds of physicists led by Carlo Rubbia, successfully accomplished what they set out to do: they proved the existence of the intermediate vector bosons, the conjectured carriers of the weak nuclear force. Three such particles were found, the W^+, the W^- and the Z^0 bosons.

The discovery of the Z^0 boson was particularly relevant to our work. The Z^0 boson is electrically neutral, which means that it can decay to pairs of neutrinos and antineutrinos, since they are also electrically neutral. (The Z^0 boson can also decay to charged particle-antiparticle pairs such as electrons and positrons.) In other words, the Z^0 boson allows one to produce each kind of neutrino directly, without first producing the associated lepton. The lifetime of the Z^0 boson serves as a measure of the number of families of elementary particles, because the more families there are, the more options the particle has for its decay. Hence a greater number of families should mean a shorter lifetime for the Z^0 boson. Careful measurements of the lifetime of the Z^0 boson could therefore reveal the number of families of elementary particles.

To measure the decay properties of the Z^0 boson one must first have a machine with enough energy to produce that particle. Older accelerators, in which high-energy beams of protons or electrons struck stationary targets, spent most of their energy on motion, leaving relatively little energy available for producing particles. The novel approach taken with the CERN machine, utilizing an idea by Simon van der Meer, is to have protons and antiprotons collide head on so that most of the energy can be utilized in producing new particles.

Several accelerators at sites around the world now employ head-on collisions. The Tevatron at Fermilab collides protons with antiprotons; SLAC and the Deutsches Elektronen-Synchrotron (DESY) collide electrons with positrons. Although the energies of the last two are too low to produce real Z^0 bosons, through quantum-mechanical phenomena they can sometimes produce "virtual" particles that mimic the effects of the Z^0 boson.

Preliminary results from the machines indicate that there are at most five families of elementary particles. David B. Cline of the University of California at Los Angeles and the University of Wisconsin at Madison and a leader of the CERN neutrino-counting efforts has shown that the lifetime of the Z^0 boson is approximately what one would expect with just three families. Experimental uncertainties, however, allow for two additional kinds of neutrinos and hence two additional families. Theodore L. Lavine, a graduate student at Wisconsin, has combined data from SLAC and DESY and obtained a comparable limit on the total number of neutrinos of about five. For the first time accelerators are counting neutrino types and getting a small number, one that was predicted by cosmological theory, not by particle theory.

The next step promises to be even more exciting. As new accelerators are completed and begin producing more data with fewer uncertainties the cosmological limit of three or at most four families will be checked with extreme accuracy. The SLAC machine is being modified to generate copious numbers of Z^0 bosons; the new accelerator is called the Stanford Linear Collider (SLC). Another accelerator under construction at CERN called the Large Electron-Positron collider ring (LEP) will also produce large numbers of Z^0 bosons. The machines will probe the early universe with an effectiveness that no telescope will ever match.

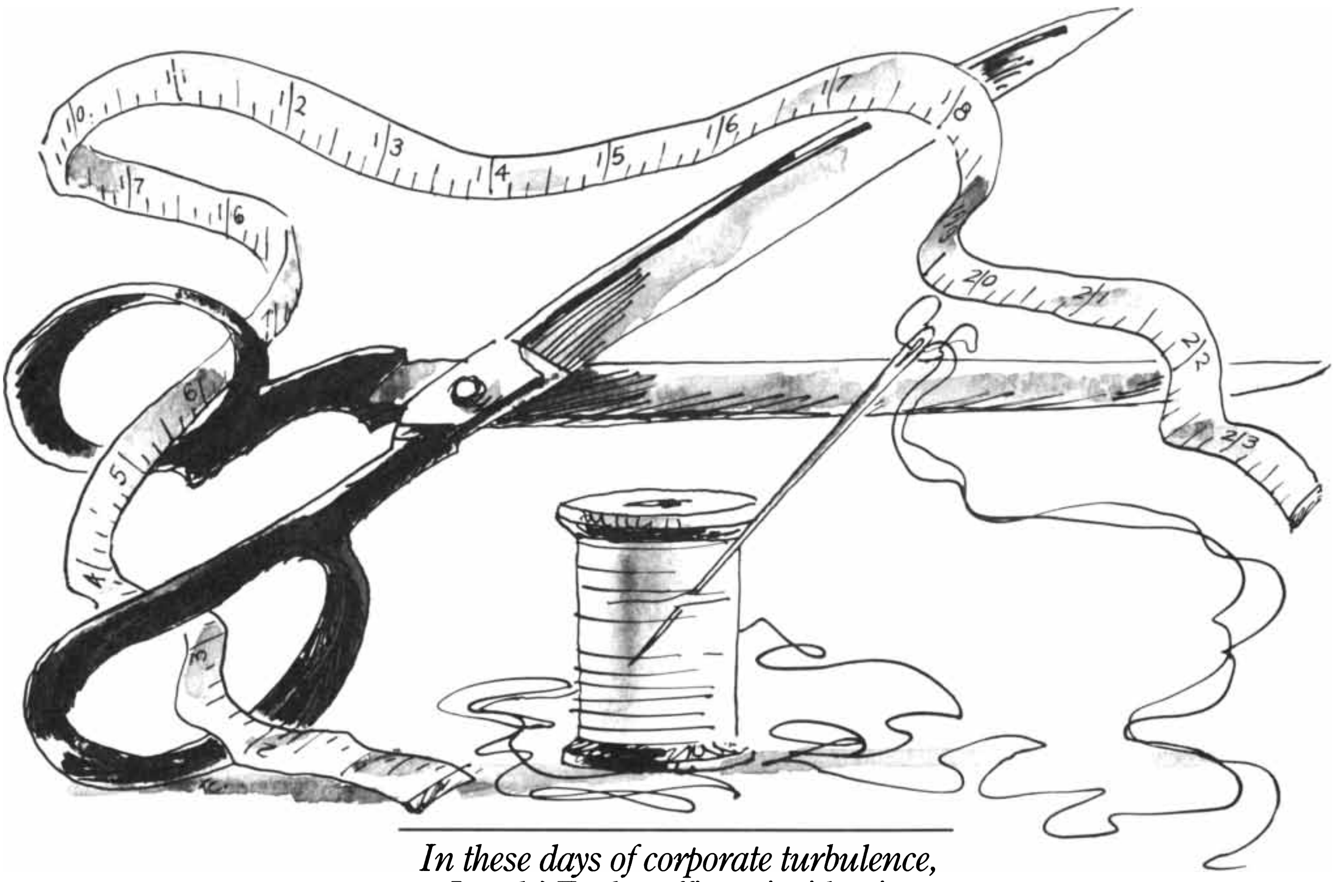
FURTHER READING

COSMOLOGICAL LIMITS TO THE NUMBER OF MASSIVE LEPTONS. Gary Steigman, David N. Schramm and James E. Gunn in *Physics Letters,* Vol. 66B, No. 2, pages 202-204; January 17, 1977.

THE EARLY UNIVERSE AND HIGH-ENERGY PHYSICS. David N. Schramm in *Physics Today,* Vol. 36, No. 4, pages 27-33; April, 1983.

BIG BANG NUCLEOSYNTHESIS: THEORIES AND OBSERVATIONS. Gary Steigman and Ann Merchant Boesgaard in *Annual Review of Astronomy and Astrophysics,* Vol. 23, pages 319-378; 1985.

NEUTRINO FAMILIES: THE EARLY UNIVERSE MEETS ELEMENTARY PARTICLE/ACCELERATOR PHYSICS. David B. Cline, David N. Schramm and Gary Steigman in *Comments on Nuclear and Particle Science,* Vol. 17, No. 3, pages 145-161; 1987.



In these days of corporate turbulence, Lands' End reaffirms its identity.

We're the "cut and sewn" people from Dodgeville, Wisconsin. Period. New paragraph.

Now that our stock is listed on the New York Stock Exchange (Symbol: LE), it seems a good time to make sure you understand we're still the same Lands' End you've always known—right down to the misplaced apostrophe in our name.

From the beginning,

...we've been hands-on merchants, and our cut-and-sewn heritage began when the sailing enthusiasts for whom we used to publish an equipment catalog came to us for a good, serviceable duffle bag. After a fruitless search for a quality item, we decided to make our own, and the Lands' End soft luggage we're so proud of today was born.

From there it was an easy step to clothing.

Fairly soon, we found ourselves searching out sources for rainwear, sweaters, parkas, and other good, warm, practical clothing. Even then, we kept our priorities straight. First quality, then price, and always, always service.

As we grew, we learned.

And as we learned, we grew. To this day, we emphasize a preference for natural fibers. For careful manufacture. For adding features to make a good item better. (Not cutting corners to make an item cheaper.)

At the same time, we fight for every crying dime of value, becoming direct merchants so we could bypass middlemen. And staying out of expensive quarters, trendy items, and lines of merchandise we had no gift or training to handle.

Finally, in terms of service,

...we offer the world's shortest, sweetest unconditional guarantee: GUARANTEED. PERIOD.® We ship most orders within 24 hours if an item is in stock. And as big as we get, we'll always remember to think small—serving our millions of customers, one at a time.

So, don't let anybody tell you different,

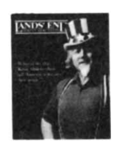
...we're still the "cut and sewn" people from Dodgeville, Wisconsin. Our toll-free number is still 1-800-356-4444. And when you call us, any hour of the day or night, one of us will answer.

Just like always.

LANDS' END® DIRECT MERCHANTS

©1988, Lands' End, Inc.

Please send free catalog.
Lands' End, Inc. Dept. Q-E2
Dodgeville, WI 53595



Name ______________________

Address ______________________

City ______________________

State __________ Zip __________

Or call Toll-free:

1-800-356-4444

Perovskites

Many of the new high-temperature superconductors belong to a family of ceramics called perovskites. The adaptable perovskite structure gives rise to materials that have a wide array of electrical properties

by Robert M. Hazen

Less than three years ago K. Alex Müller and J. Georg Bednorz of the IBM Zurich Research Laboratory discovered a material that was superconducting at higher temperatures than had ever before been achieved: it offered no resistance to the flow of electricity at temperatures as high as 28 degrees Kelvin (degrees Celsius above absolute zero). Since then other "high temperature" superconductors have been engineered that are efficient at temperatures of 100 degrees K. or more. Exactly why the materials are superconducting remains a mystery, but several groups, including my colleagues and me at the Geophysical Laboratory of the Carnegie Institution of Washington, have shown that the new materials have this in common: they are structurally flawed members of a crystallographic family known as perovskites.

Perovskites, which derive that name from the specific mineral known as perovskite, are ceramics (solid materials combining metallic elements with nonmetals, usually oxygen) that have a particular atomic arrangement. They are the earth's most abundant minerals and have long been of interest to geologists for the clues they hold to the planet's history. They fascinate from a technological point of view as well because, as a group, natural and synthetic perovskites exhibit an array of electrical properties. Whereas a given crystal structure is usually associated with a specific electrical property, perovskites run the gamut from insulators (nonconductors) to semiconductors, superionic conductors (in which whole ions, rather than just electrons, flow through the crystal), metal-like conductors and now high-temperature superconductors. Perovskites currently form the basis of a $20-billion-per-year electroceramics industry, a figure that may soon be eclipsed by applications of the high-temperature superconductors.

What accounts for this remarkable range of properties? The flaws in the superconductors suggest the answer: slight modifications of the ideal perovskite architecture often result in new features. There is no one-to-one correlation; a given modification does not automatically produce a particular degree of electrical conductivity. Yet any time the ideal structure is altered, the possibility of new electrical—or other—properties arises.

The Perovskite Structure

In their ideal form perovskites, which are described by the generalized formula ABX_3, consist of cubes made up of three distinct chemical elements (*A*, *B* and *X*) that are present in a ratio of 1 : 1 : 3. The *A* and *B* atoms are metallic cations (ions with a positive charge) and the *X* atoms are nonmetallic anions (ions with a negative charge). An *A* cation—the larger of the two kinds of metals—lies at the center of each cube, the *B* cations occupy all eight corners and the *X* anions lie at the midpoints of the cube's 12 edges [*see illustration on page 76*]. As one might expect, the mineral perovskite (the variety of calcium titanate, $CaTiO_3$, that forms at high temperatures) embodies the ideal structure. Its unit cell, or basic building block, consists of a single cube. The calcium in the *A* position is larger than the titanium in the *B* positions, and oxygen fills all 12 of the *X* sites. The crystal often grows into the shape of a cube or an octahedron, reflecting the symmetry of the atomic structure.

Someone unfamiliar with crystals might well wonder why a crystal composed of units that have one *A* ion, eight *B* ions and 12 *X* ions is said to have the formula ABX_3 rather than AB_8X_{12}. The reason is that each cube is surrounded on all its sides by other cubes. These share corners and edges, and so any given cube is in full possession of only a fraction of the outlying atoms. When such sharing is taken into account, each cube is considered to have one *A* atom, one complete *B* atom and three complete *X* atoms.

A great many elements can combine to form the hundreds of ideal or modified perovskites now known. Barium, potassium and the rare-earth elements (those from cerium through lutetium, numbers 58 through 71 in the periodic table of the elements) are typical of the two dozen elements that can fill the *A* position. An astonishing total of almost 50 different elements—more than half of the stable entries in the periodic table—are known to adopt *B* sites. The *X* positions can be taken up not only by oxygen but also by members of the halogen family of nonmetals: fluorine, chlorine or bromine. Among the dozens of compounds that are known to conform to the ideal perovskite structure are such diverse chemicals as silver zinc fluoride ($AgZnF_3$), cesium cadmium bromide ($CsCdBr_3$), lithium barium fluoride ($LiBaF_3$), potassium iodate (KIO_3) and europium aluminate ($EuAlO_3$).

Like other ceramics, these ideal perovskites are electrical insulators: all atomic sites are filled, and strong ionic bonds—the attractive forces between the cations and the anions—hold the atoms and their electrons tightly in place. As a consequence electrons move through the crystal only with difficulty. The strong bonds in most ideal perovskites also make them rocklike, scratch-resistant, difficult to

ROBERT M. HAZEN is an experimental mineralogist at the Geophysical Laboratory of the Carnegie Institution of Washington, where he has been since 1976. His Ph.D., awarded in 1975, is from Harvard University. Hazen led the team that first deduced the structure and composition of the "1-2-3" high-temperature superconductor, and he recently published *The Breakthrough: The Race for the Superconductor,* a book about the material's discovery and analysis.

deform and hard to melt. Moreover, the fact that the bonds along each of the three axes of the cube are alike leads to a three-dimensional uniformity of properties known as isotropy: the properties of the bulk material—such as compressibility or electrical conductivity—are the same along each of the axes.

Not all perovskites fit this model of the isotropic insulator, however. A number of them deviate slightly from the ideal in shape or composition and hence are less predictable.

Simple Alterations Abound

Many perovskites are somewhat distorted because the central *A* cation is too small in relation to the *B* cations at the corners of the cube. The disparity causes the *X* atoms, and sometimes the *B*'s, to move out of position.

Crystallographers often visualize such movements by thinking of the ideal unit cell not as a simple cube but as a cluster of polyhedrons [*see illustration on next page*]. Each *B* cation, which defines the corner of adjoining cubes, is surrounded by and closely bonded to six anions, one from each of the six cube edges that converge at the corner. The anions define the points of an octahedron. The *A* cation, formerly viewed as being at the center of a cube, is now considered to be surrounded by eight corner-sharing octahedrons, each of which contains a *B* cation in the center. When *A* cations are too small in relation to the *B* cations, the octahedrons, whose axes are aligned in an ideal perovskite, tilt and twist; the framework collapses around the *A* cations, lowering the symmetry and potentially altering the perovskite's optical, elastic, electrical and other physical properties. Dozens of different perovskite tilt patterns have been identified.

In tilted perovskites, as in the ideal forms, the *B* cations may remain at the center of their octahedrons. In some perovskites, however, the *B* cations are slightly shifted. Such "off-centering" of positively charged cations can give perovskite crystals electrical polarity: one end is positively charged and the

ARCHETYPAL PEROVSKITE (*angular chunk*) is a rare mineral formed from calcium titanate ($CaTiO_3$) at high temperatures. Other compounds in the perovskite family generally also conform to the formula ABX_3, in which *A* and *B* are metals and *X* is a nonmetal. The majority of natural perovskites are electrical insulators, but deviations from the standard formula or atomic arrangement can result in other electrical properties. Perovskite itself is an insulator. It was first described in the 1830's by the geologist Gustav Rose, who named it after the Russian mineralogist Count Lev Aleksevich von Perovski.

other end is negatively charged. Moreover, the direction of the off-centering can often be changed simply by subjecting the sample to an electric field. Materials that are both polarized and able to reverse polarity under the influence of an electric field are known as ferroelectrics and have many applications in electronic devices.

One ferroelectric perovskite—the synthetic compound barium titanate ($BaTiO_3$)—is perhaps the best-known electroceramic in commercial use. Its off-centered cations store and release electrical energy quite efficiently: the stronger the applied field is, the more the cations are energized and displaced and the more strongly polarized the crystal becomes. When the electric field is removed, the cations return slowly to their normal positions and release the stored energy.

Barium titanate is often incorporated in capacitors, the elements in electronic circuits that can smooth out an uneven flow of current. These capacitors can store charge from pulses of current and then release the stored charge between pulses, thereby producing a steady direct current. Barium titanate has also found wide application in voltage-surge protectors for computers: when a surge of electricity (from a lightning strike, for example) reaches the crystal, it absorbs the pulse and dissipates it slowly.

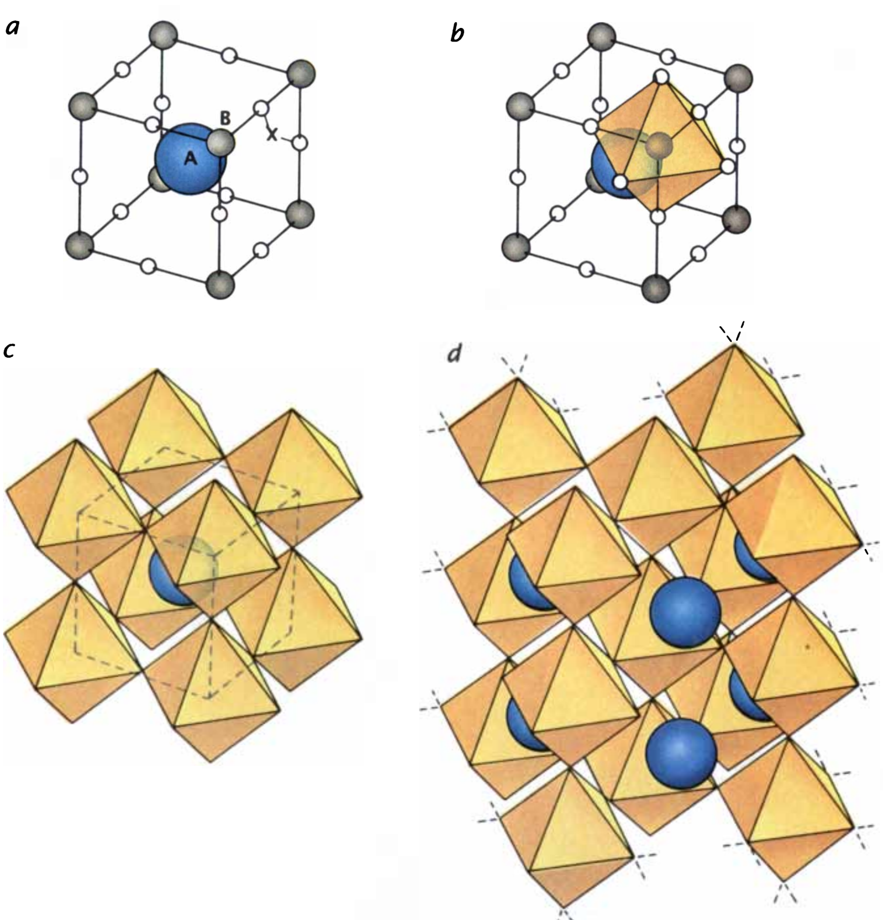
Barium titanate's off-centered cations also contribute to a commercially exploited property known as piezoelectricity. When an external electric field shifts the titanium atoms, the crystal as a whole changes shape: it elongates somewhat. Conversely, mechanical deformation of a barium titanate crystal generates an electric field. Piezoelectricity makes barium titanate valuable in the construction of transducers that convert mechanical energy into electrical energy or vice versa. In loudspeakers, for instance, barium titanate transducers vibrate against a membrane in response to electronic signals, generating sound waves. In microphones such transducers store and then release electrical energy in response to the pressure exerted by sound waves.

Mixed Cations

Perovskites that have off-centered cations or tilted octahedrons often conform to the simple ABX_3 formula. Yet these perovskites and others may also deviate from the ideal composition: their *A* or *B* sites can be filled with two or more types of cations. In calcium uranium oxide (Ca_2CaUO_6), for example, calcium atoms take up all the available *A* sites, but the octahedral *B* sites are equally divided between calcium and uranium ions. The *B* cations "order," or strictly alternate positions, throughout the structure, which is tilted because the calcium ions in the *B* positions are significantly larger than the uranium ions.

The octahedral centers can be filled not only by two different elements but also by differently charged ions of the same element. Barium bismuth oxide, a lustrous bronze semiconductor that is sometimes described by the formula $BaBiO_3$, is such a "mixed valence" perovskite: it has two distinct varieties of bismuth cations, one that has surrendered three valence electrons (electrons in its outermost orbital shell) and one that has surrendered five. The usual designation is therefore $Ba_2Bi^{3+}Bi^{5+}O_6$. As is true for calcium uranium oxide, the octahedrons containing the two types of cations alternate throughout the crystal.

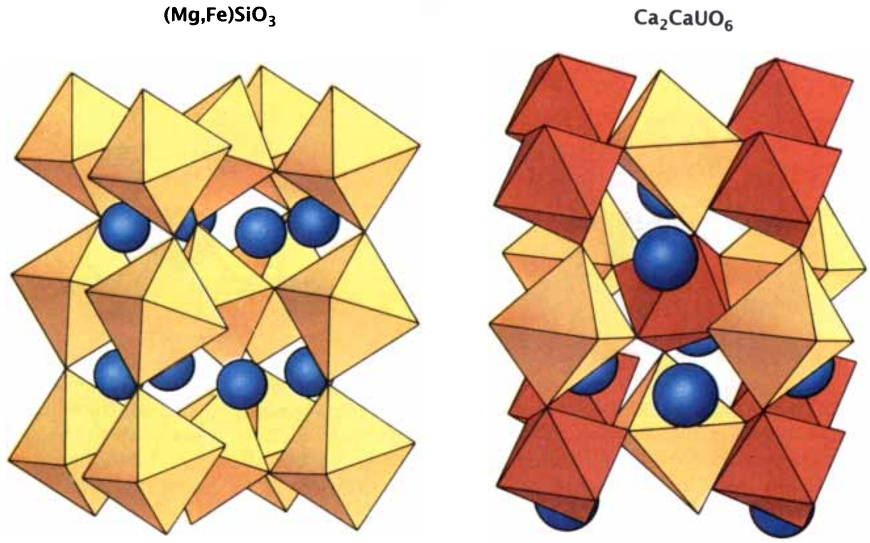
Barium lead oxide ($BaPbO_3$), a black, metal-like conductor, provides a dramatic demonstration of the changes in properties that can result from slight variations in the elements that fill the *B* positions. By replacing the lead with increasing amounts of bismuth, one can create a "continuous series" of compositional variants from barium lead oxide to barium bismuth oxide. As the composition approaches $BaPb_{.8}Bi_{.2}O_3$ the compound becomes a semiconductor. In addition, unlike either barium lead oxide or barium bis-



BASIC STRUCTURAL UNIT of perovskites is a cube (*a*). One metallic atom (*A*) lies at the center, eight smaller metallic atoms (*B*) occupy the corners and 12 nonmetallic atoms (*X*) are the midpoints of the edges. The *A* and *B* atoms are cations, or positively charged ions, and the *X* atoms are negatively charged anions. (Perovskites have the formula ABX_3 instead of AB_8X_{12} because each *B* cation is shared by eight neighboring cubes, and each *X* anion by four cubes.) Crystallographers often replace the cubic model with a polyhedral one (*b*). The six *X* anions that surround each *B* cation and are closely bound to it form the points of an octahedron (*yellow*). In such a model the basic structural unit becomes a group of eight corner-linked octahedrons around an *A* cation (*c*). The bulk crystal is a continuous network of such groupings (*d*). These polyhedral drawings and the ones that follow are based on computer-generated images made by Ross J. Angel of the Geophysical Laboratory of the Carnegie Institution of Washington, who used a modified version of the program STRUPLO.

muth oxide, some of the intermediate compounds become superconducting when they are cooled almost to absolute zero.

B-cation substitutions are put to work in the commercially important group of perovskites known collectively as *PZT*. The designation encompasses lead titanate ($PbTiO_3$) and the continuous series created when zirconium is substituted for titanium until lead zirconate is formed ($PbZrO_3$). All the *PZT* crystals exhibit a remarkably strong piezoelectric effect (a slight compression can result in the storage and release of 100 volts of electric potential); they are found in a wide assortment of devices, including loudspeaker buzzers, electrical relays, pressure gauges, and spark igniters for lawn mowers. *PZT* crystals of differing compositions vibrate at distinct frequencies in response to an electric field. Television designers have capitalized on this compositional effect and routinely employ several *PZT* perovskites as filters to reduce unwanted noise: the crystals intercept incoming electrical impulses and mask undesirable frequencies.

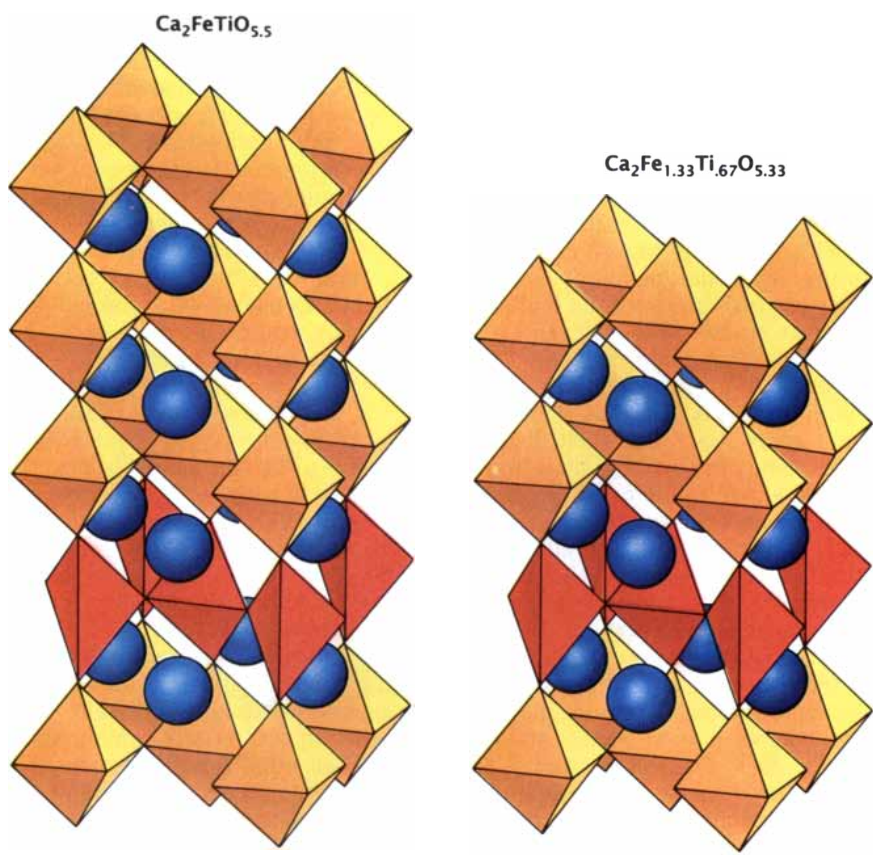


DEVIATIONS from the standard perovskite formula and structure are evident in magnesium-iron silicate and calcium uranium oxide. In the first compound (*left*) two distinct elements (magnesium and iron) can fill the *A* position (*blue*), resulting in the formula $(Mg,Fe)SiO_3$. In calcium uranium oxide (*right*) calcium alone fills the *A* position, but it also alternates with uranium in the *B* sites; the compound's formula is Ca_2CaUO_6. In magnesium-iron silicate the octahedral structural units tilt because both kinds of *A* cations are too small to support the silicon-centered octahedrons in a more upright position. In calcium uranium oxide it is the *B* cations that are mismatched: the octahedrons twist out of position because the calcium-centered octahedrons (*yellow*) are significantly larger than the uranium-centered ones (*red*).

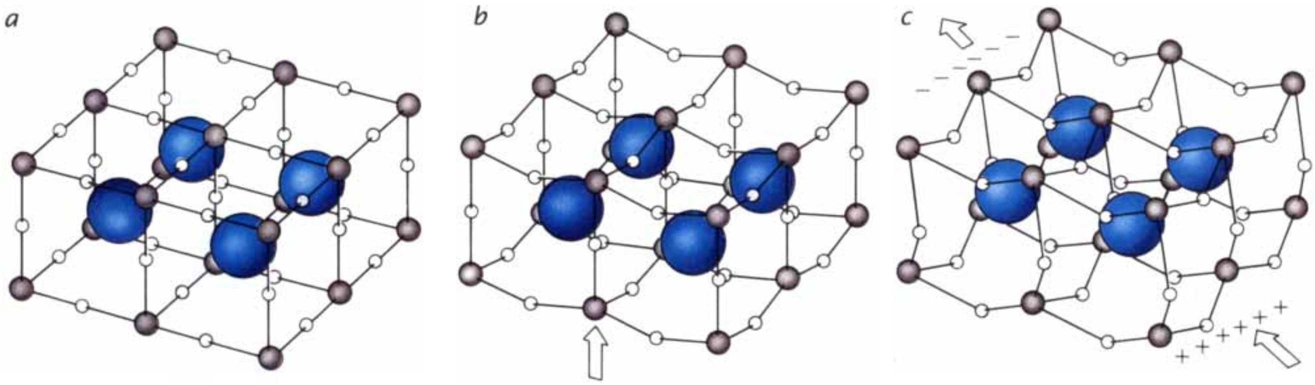
Multiple *A*'s

Another kind of compositional variation—multiple cations in the *A* positions—may control the properties of the earth's interior. At the Geophysical Laboratory my colleagues and I have intensively studied magnesium-iron silicate $(Mg,Fe)SiO_3$, which adopts the perovskite structure at pressures of several hundred thousand atmospheres. The parentheses indicate that the proportions of the magnesium and iron filling the *A* sites can vary. Because both magnesium and iron are rather small, the octahedrons that enclose the silicon atoms at the corners cannot remain upright; instead they tilt.

High-pressure studies suggest that magnesium-iron silicate and calcium silicate ($CaSiO_3$), which also adopts the perovskite structure when it is under great pressure, are probably the predominant minerals in the earth's lower mantle, a region extending from a depth of about 670 kilometers to about 2,900 kilometers and constituting more than half of the earth's volume. Studies done on upper-mantle silicate minerals such as garnets, olivines, spinels and pyroxenes show that at the pressures and temperatures of the earth's deep interior these silicates become transformed, combining with other minerals to form rocks that are dominated by pe-



LAYERED STRUCTURE sometimes arises in perovskites that lack a full complement of oxygen atoms. Entire sheets of octahedrons are replaced by smaller polyhedrons. In two of the oxygen-poor compounds that result when iron replaces some of the titanium in calcium titanate—$Ca_2FeTiO_{5.5}$ (*left*) and $Ca_2Fe_{1.33}Ti_{.67}O_{5.33}$ (*right*)—every fourth and third layer respectively consists of iron-centered tetrahedrons (*red*).



BARIUM TITANATE ($BaTiO_3$) conforms strictly to the standard ABX_3 formula of perovskites, but it deviates from a true cubic structure (*a*) in that its *B* (titanium) cations are slightly off-center (*b*). As a result the crystal is polarized. When it is subjected to an electric field (*c*), the orientation of the titanium ions shifts toward the negatively charged electrode. As the orientation changes, so does the shape of the barium titanate crystal—a phenomenon that is known as piezoelectricity and that is the basis for many electronic devices. Here the shifts of the titanium atoms are greatly exaggerated for clarity.

rovskite-type magnesium-iron silicate and calcium silicate. The studies also suggest that when the iron atoms in magnesium-iron silicate have a charge of +2, the resulting perovskite is an insulator; when the iron atoms have a charge of +3, however, the perovskite may become a more efficient conductor of heat and electricity.

If silicate perovskites rich in +3 iron are indeed abundant in the lower mantle, the earth might be a better conductor than has been thought. The finding of additional evidence supporting this possibility could make it necessary to reexamine prevailing ideas about the earth's initial temperature, the rate at which it has cooled over time and how quickly it is likely to lose heat in the future. My colleagues and I plan to carry out high-pressure and high-temperature studies in the near future to test such speculations.

Iron and magnesium appear randomly in the *A* positions of magnesium-iron silicate, but such disorder pales compared with what is found in several natural perovskites crystallized from volcanic magma. These rocks, which include latrappite, loparite, leushite and several other related perovskites, incorporate a plethora of exotic elements. As a group they are hosts to calcium, sodium, potassium, yttrium, thorium and all 14 rare-earth elements at *A* sites; titanium, niobium, iron, magnesium, manganese and zinc occupy their *B*-cation positions. The perovskite structure is so adaptable that these minerals act like sponges, soaking up various elements that do not fit readily into other mineral structures. The resulting perovskites are exceedingly disordered: a given sample may contain some 20 different *A* cations and a dozen *B* cations, all apparently distributed at random in their respective sites.

Added Complexity

No matter how distorted or compositionally varied they are, all the perovskites described above are stoichiometric: they have a total of two cations for every three anions. A variety of other perovskites or perovskite-related structures are nonstoichiometric: they deviate from the ideal ABX_3 formula by having vacant sites where atoms would normally be.

The compounds in the continuous series from perovskite ($CaTiO_3$) to the oxygen-deficient compound calcium ferrite ($CaFeO_{2.5}$) illustrate well the structural consequences of missing atoms. In the oxygen-poor compounds beyond calcium titanate—such as $Ca_2FeTiO_{5.5}$ and $Ca_2Fe_{1.33}Ti_{.67}O_{5.33}$—the oxygen deficit results in layers consisting of iron *B* cations that are surrounded by four oxygen atoms rather than the usual six. These layers are interspersed with ones featuring normal octahedrons [*see bottom illustration on preceding page*]. The bulk crystal is far from uniform: it grows in flat, platelike crystals, reflecting its layered atomic arrangement.

Such complexities as nonstoichiometry, octahedral tilting, cation off-centering and multiple cations can occur together in almost any combination. Indeed, an infinite number of perovskite variants seems possible, and yet even this picture is incomplete. In a uniform perovskite one can in theory predict the atomic arrangement of one section of a crystal by analyzing another section, even in a crystal that has tilted octahedrons or alternating cations. In reality most perovskites, like other types of crystals, are nonuniform: they are loaded with defects, random faults that cannot be anticipated. Among the many small defects that can occur are a misplaced cube, an anomalously tilted polyhedron, a layer of one type where another type of layer would normally form, a break in an alternating sequence of cations and point defects, which arise when a single atom is missing or is replaced by an errant element.

Perovskite defects can also occur on a larger scale. For example, in a phenomenon known as twinning, large sections of a crystal may be identical with other parts but oriented in a different direction. (The boundaries of such twins often appear as striations on bulk crystals.) An almost ubiquitous feature of perovskites, twinning is likely to occur in any specimen that deviates from the ideal cubic form, and it can have dramatic effects—favorable and unfavorable—on electrical properties. For example, electric fields can cause twin boundaries to shift, a process that consumes energy and potentially reduces the polarization of ferroelectric perovskites. In order to ensure reliable performance for commercial applications, manufacturers sometimes introduce impurities that have the effect of "pinning" twin boundaries and preventing them from moving. For instance, manganese is routinely added to pin the boundaries in barium titanate, which invariably displays twinning.

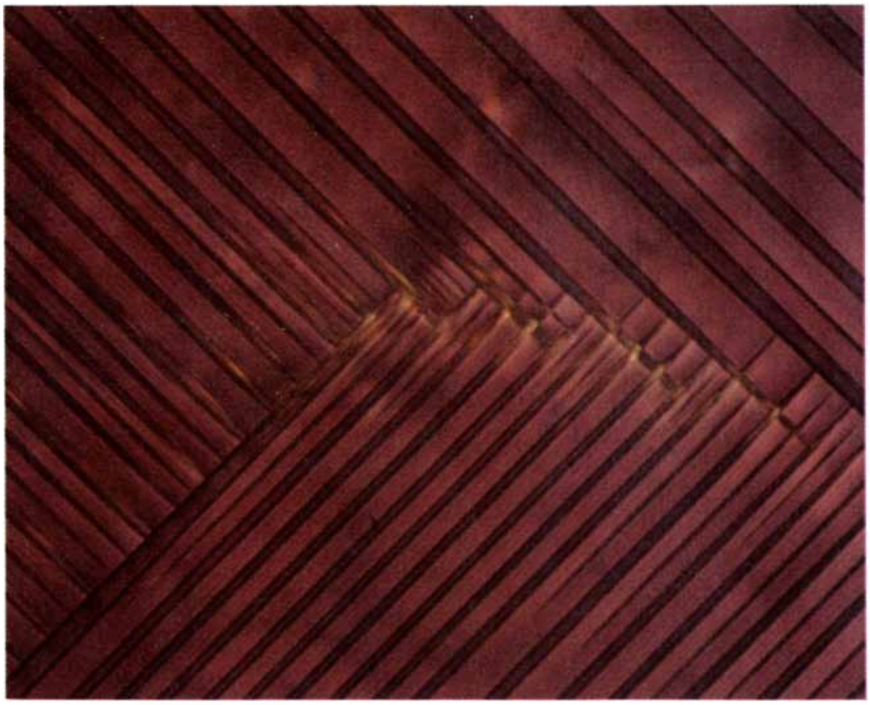
Superconductors: Rich in Defects

In no perovskite are the electrical effects of structural variations or crystal defects more striking than they are in the newest perovskites: the high-temperature superconductors. My intimate acquaintance with these materials began as a result of a request made in February, 1987, by Ching-Wu (Paul) Chu of the University of Houston. Chu and his co-workers at Houston and the University of Alabama at Huntsville had shown beyond a doubt that a substance they had developed was superconducting at a record-breaking temperature, 93 degrees K. They had formed the material by baking a finely ground mixture of yttrium and copper oxides along with barium carbonate, and they knew the ratios of elements in the starting mixture.

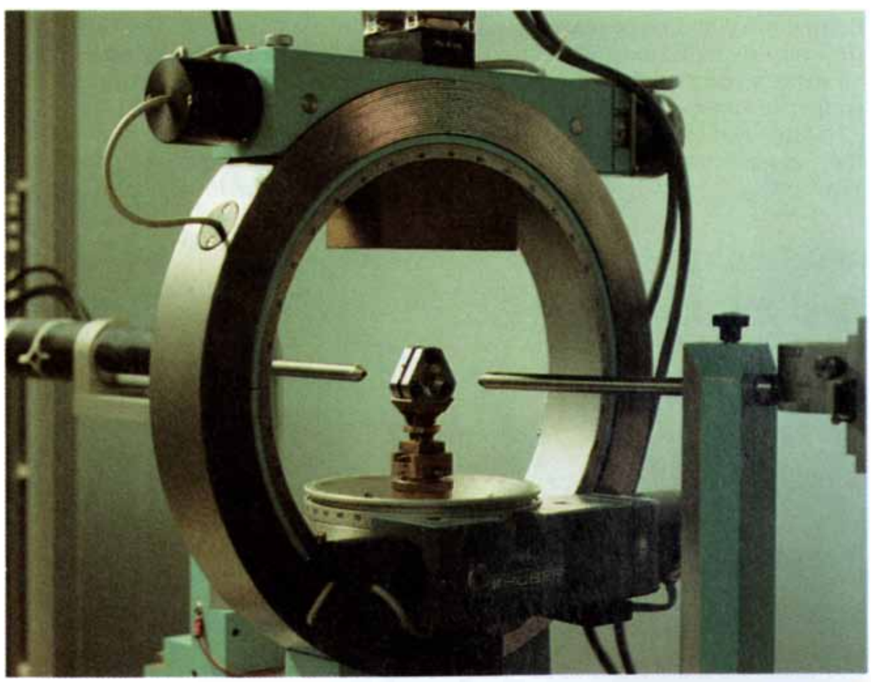
Yet they had a problem. The superconducting sample that emerged from the oven was a mixture of two phases, or distinct chemical compounds. A high-powered optical microscope revealed that the dominant phase, making up perhaps two-thirds of the material, was a deep emerald green, was transparent and appeared to crystallize with rounded edges. The other phase was black, opaque, more regular in shape and square-edged; it was suspected of being the superconductor, because all known conductors are opaque. The problem was that the two phases were so thoroughly intergrown that it was impossible to separate them and analyze their composition by routine techniques. Chu needed to know the composition of the superconducting phase before he could attempt to engineer a pure form or to produce chemical variants that might be superconducting at even higher temperatures.

At the Geophysical Laboratory we often analyze "dirty," or mixed-phase, natural and synthetic rocks. Furthermore, we were not engaged in our own hunt for a superconductor and hence posed no threat to Chu's proprietary interests. Our laboratory was therefore a logical place for Chu to turn for assistance. In addition to me, our team included crystallographers Ross J. Angel, Larry W. Finger, Charles T. Prewitt and Nancy L. Ross, geophysicist Ho-Kwang (David) Mao (Chu's initial contact) and chemical analyst Chris G. Hadidiacos.

We proceeded as we would for any fine-grained rock. We attempted to determine the nature of the elements in each phase by means of an electron microprobe, a device that focuses an intense beam of electrons onto a tiny sample. The energetic electrons excite the atoms in the sample and cause each element to emit X rays of a characteristic frequency. We determined the ratios of the separate elements by measuring the proportion of X rays



CROSS SECTION of a barium titanate crystal reveals a common defect in perovskites: twinning, or the formation of essentially identical domains that are oriented differently. Here the differing stripe patterns in the two large domains reflect differences in the titanium off-centering that is characteristic of barium titanate. The titanium atoms in adjacent light and dark bands are shifted in opposite directions. The sample, which was provided by Robert D. Shannon of E. I. du Pont de Nemours & Company, Inc., appears purple because it was photographed under polarized light.



FOUR-CIRCLE DIFFRACTOMETER can provide information about the structure of a microscopically small crystal. The specimen is maneuvered into the desired orientation by two motorized "arcs" that move in a complete circle (*inner part of large silver "doughnut" and the central pivot*) and by a third arc that rotates the entire doughnut assembly. A fourth arc (*arm extending to the left*) cradles an X-ray detector to measure the positions and intensities of diffracted X rays. This device helped the author and his colleagues to demonstrate that one of the new high-temperature superconductors—yttrium barium copper oxide, or "1-2-3"—is a modified perovskite.

that were emitted at each frequency.

Although the grain size of Chu's superconductor was almost too small for our instrument, we were eventually able to deduce the elemental ratios of the two phases. Most of the finest grains—presumably the green stuff—included yttrium, barium and copper ions in proportions of 2 : 1 : 1 respectively. The important black phase was richer in copper: the metals were present in a ratio of 1 : 2 : 3, which accounts for the superconductor's nickname "1-2-3." (It is also called YBCO, pronounced "yibco," for its constituent elements.)

Our principal uncertainty regarding the composition of the black phase was the exact number of oxygen atoms, because the electron microprobe cannot tell us anything about that element. Yet we could make an educated guess on the basis of a well-accepted natural law holding that positive and negative charges in a crystal have to balance. We knew that the charges of yttrium and barium ions are +3 and +2 respectively, that copper ions can have a charge of +1, +2 or +3 and that all three types of copper ion can be present in a compound. Simple arithmetic then indicated that the black material had a total positive charge of from 10 to 16. The charge of an oxygen ion is −2, and so we knew there had to be between five and eight oxygen atoms for every six metal ions. We settled on an oxygen content of 6.5 because the copper in most compounds is in the +2 state. The black phase, then, had the approximate chemical formula $YBa_2Cu_3O_{6.5}$.

X-ray-diffraction analysis soon enabled us to learn something about the three-dimensional arrangement of the atoms. When X rays strike a crystal, they scatter in characteristic directions at distinctive intensities, providing information about the arrangement of the crystal's constituent atoms. Finding a single crystal large enough for study was difficult because the sample was so fine-grained, but we did in the end obtain a few black specks that were between 30 and 40 micrometers (millionths of a meter) in diameter. Counting ourselves fortunate, we glued a single black bit, too small to see with the unaided eye, to the end of a slender glass fiber and irradiated it with an X-ray beam.

We were thwarted at first because every "single" black speck we tried was actually a combination of two or more minute intergrown crystals, which made diffraction analysis difficult. After several false leads we did nonetheless manage to determine that the material in question had a cube-shaped arrangement of atoms, which were spaced at intervals of about 3.9 angstrom units (3.9×10^{-8} centimeter) along each edge. The arrangement was that of a perovskite.

We were delighted and puzzled: delighted because perovskites are one of the structures we most enjoy studying and puzzled because we had never encountered a perovskite with such a low ratio of oxygen atoms to cations. A normal perovskite would have had nine oxygens for every six cations ($A_3B_3O_9$, or three times ABO_3) and not 6.5 oxygens. The lowest ratio known, in fact, was 7.5 oxygens for every six cations. We realized that this superconducting perovskite was going to have an unusual structure with several missing oxygens.

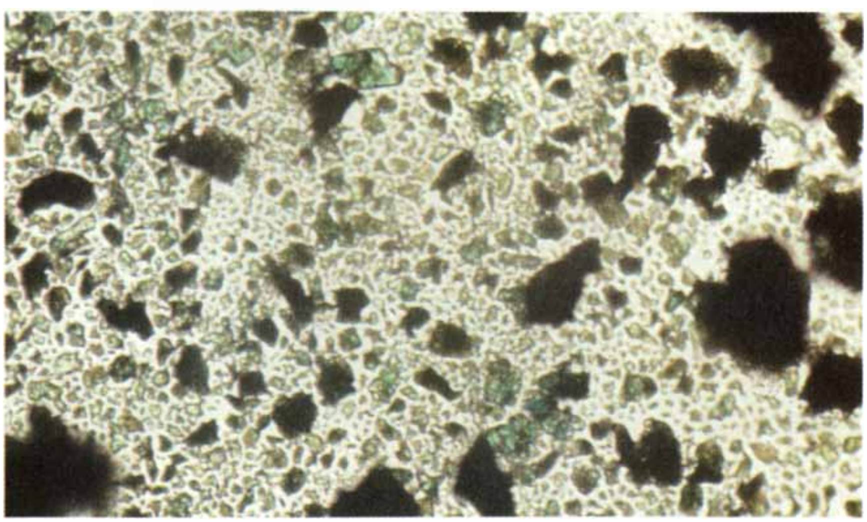
Our next clue to the arrangement of the atoms came from a search for "superstructure" diffraction effects: diffraction patterns that can indicate whether the unit cell is composed of one cube or more. (Although all perovskites are built from individual cube-shaped units, many have unit cells that consist of two or more cubes, each of which has a different composition or arrangement of atoms.) The pattern we discerned indicated that the basic repeating unit of Chu's black superconductor consisted of three cubes. For convenience I shall speak of the cubes in the unit cell as if they are stacked one above another; they could, however, be arranged in a horizontal manner.

Not as Simple as 1-2-3

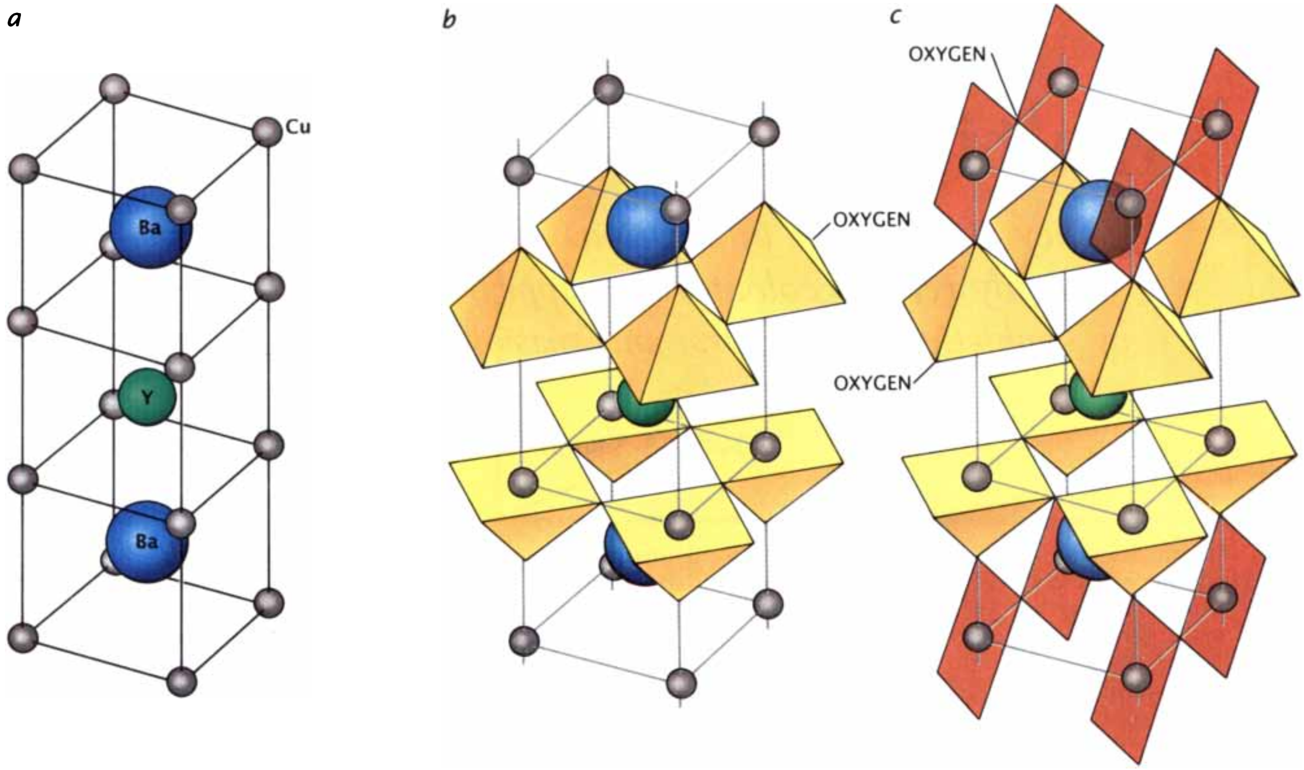
The details of the structure began to take shape in our minds [*see illustration on opposite page*]. We assumed that the relatively large barium and yttrium atoms would fill the *A* sites, whereas the small copper atoms would occupy the *B* sites. The superstructure diffraction effects suggested that barium, which is larger than yttrium, was at the center of the upper and lower cubes and that the yttrium was in the middle cube; such Ba-Y-Ba three-cube units would then be stacked one on top of another throughout the crystal. But where did the 6.5 oxygens belong, and which normally filled oxygen sites were empty?

Additional X-ray-diffraction studies revealed part of the answer, albeit grudgingly. For one thing, the scattering of X rays by oxygen, which is a light atom, is difficult to discern in the context of the much stronger diffraction patterns produced by the heavier barium and yttrium atoms. Our task was also complicated by the intergrown nature of our crystals and by the presence of twinning even in the smallest specimens. Nevertheless, after several days of painstaking measurements a pattern began to emerge.

We found evidence for the presence of oxygen atoms in all available sites on the horizontal planes immediately above and below the yttrium atom (at the boundaries shared with the barium-centered cubes). In contrast, oxygen was clearly absent from the vertical edges of the yttrium cube. That settled, we assigned the remaining oxygens to anion positions that were



HIGH-TEMPERATURE SUPERCONDUCTOR known as 1-2-3 (*black crystals*) was originally produced together with an extraneous substance (*green crystals*). In 1987 at the request of the material's developer, Ching-Wu (Paul) Chu of the University of Houston, the author and his colleagues began the laborious process of separating the two substances and deciphering their chemical compositions and structures.



STRUCTURE OF 1-2-3 was determined in stages. Early studies showed that the unit cell, or smallest repeating unit, of the superconductor consists of three cubes (*a*). Copper (Cu) fills the *B* sites; barium (Ba) occupies the *A* sites in the outer cubes and yttrium (Y) occupies the *A* site of the central cube. But where were the oxygens? Later work showed that 1-2-3 has two main variants. Both lack oxygen on the vertical edges of the yttrium cube, but the variety that is poorer in oxygen, $YBa_2Cu_3O_6$ (*b*), also lacks oxygen in the horizontal planes at the top and bottom of the unit cell. The copper atoms in those planes are in linear coordination: each is tightly bound to one oxygen above it and one below. The other type, $YBa_2Cu_3O_7$ (*c*), is a better superconductor. Two oxygen atoms lie in both the top and bottom plane, and so the copper atoms in those planes fall at the center of squares formed by four oxygen atoms. Another model of $YBa_2Cu_3O_7$ appears on the cover of this issue.

still "open" in the two barium cubes, leaving about half of those positions unfilled so that on the average each unit cell had a total of 6.5 complete oxygen ions.

Several weeks later workers at the Argonne National Laboratory, employing a technique known as neutron-powder diffraction, determined the specific positions of the remaining oxygens. They also showed that 1-2-3 could have as many as seven oxygen atoms. Indeed, it is now known that yttrium barium copper oxide sometimes includes six oxygens, sometimes seven and sometimes a fraction in between. Apparently it is the more oxygen-rich varieties that are superconductors.

In addition to being highly deficient in oxygen, 1-2-3 is also an unusual perovskite in that the oxygen atoms near the barium cations can, with time, hop back and forth between anion sites in the same plane, altering the crystal structure. Moreover, the presence of many empty anion sites allows the material to take in extra oxygen atoms when the oxygen content of the environment increases.

Of particular interest is the finding that the copper atoms are not surrounded by the usual octahedral cages. Instead they are mostly in fourfold square-planar or fivefold square-pyramid coordination: they are closely bonded to the four or five surrounding oxygen atoms, which form a square and a square-based pyramid respectively. In both arrangements the copper atoms lie in the same plane as the square. Since 1-2-3 and other high-temperature superconductors are superconducting in directions parallel to the copper planes, many theorists now believe this unusual planar arrangement contributes to the remarkable electronic properties of the materials. Exactly how it does so remains to be clarified, however.

In spite of its unique features, 1-2-3 has many of the characteristics of other nonideal perovskites, such as ordered cations, missing oxygen atoms, layering, twinning, numerous point defects and errors in the order of stacked layers. Such deviations underscore the formidable challenges perovskites present to anyone who attempts to analyze them or engineer custom-made varieties. Yet, as the superconductor story strikingly demonstrates, the deviations that challenge also provide tremendous opportunity for the development of revolutionary electronic materials.

FURTHER READING

PIEZOELECTRIC CERAMICS. Bernard Jaffe, William R. Cook, Jr., and Hans Jaffe. Academic Press, Inc., 1971.

THE CLASSIFICATION OF TILTED OCTAHEDRA IN PEROVSKITES. A. M. Glazer in *Acta Crystallographica,* Section B, Part 11, Vol. B28, pages 3384–3392; November 15, 1972.

CRYSTAL STRUCTURES: A WORKING APPROACH. Helen D. Megaw. W. B. Saunders Company, 1973.

STRUCTURE AND CRYSTAL CHEMISTRY OF PEROVSKITE-TYPE $MgSiO_3$. Takehiko Yagi, Ho-Kwang Mao and Peter M. Bell in *Physics and Chemistry of Minerals,* Vol. 3, No. 2, pages 97–110; August 28, 1978.

THE BREAKTHROUGH: THE RACE FOR THE SUPERCONDUCTOR. Robert M. Hazen. Summit Books, 1988.

Bacteria as Multicellular Organisms

They differentiate into various cell types and form highly regular colonies that appear to be guided by sophisticated temporal and spatial control systems

by James A. Shapiro

Without bacteria, life on earth could not exist in its present form. Bacteria are key players in many geochemical processes, including the fundamental nitrogen, carbon and sulfur cycles, which are critical to the circulation of life's basic elements. If these processes were to grind to a halt, the planet's soils, waters and atmosphere would become inhospitable for life. Yet in spite of such global importance, the notion has persisted that bacteria are simple unicellular microbes.

That view is now being challenged. Investigators are finding that in many ways an individual bacterium is more analogous to a component cell of a multicellular organism than it is to a free-living, autonomous organism. Bacteria form complex communities, hunt prey in groups and secrete chemical trails for the directed movement of thousands of individuals.

Already at the beginning of this century investigators had evidence that bacteria live communally in the soil. Martinus Beijerinck of the Netherlands discovered that *Rhizobium* bacteria infect the roots of leguminous plants, where they form organized multicellular structures that function as factories for nitrogen production. At about the same time Sergei Winogradsky, working in Paris, elucidated the role bacteria play as decomposers of cellulose in the global carbon cycle. Winogradsky was also one of the first microbiologists to observe bacteria directly in the soil, where he found that few exist as isolated cells; most live in groups adhered to soil particles. Similar group behavior was well known in the laboratory, where bacteria formed distinctive colonies on petri dishes or adhered as organized populations to the walls of culture flasks.

In spite of these early observations, the image of bacteria as unicellular organisms has persisted over the years. In large part this can be attributed to medical bacteriology. Disease-causing organisms are commonly identified by isolating a single cell of the suspected pathogen, growing a culture from that cell and showing that the resulting pure culture gives rise to the disease in question. The possibility that infections of the human body involve multicellular aggregations of bacteria is normally not even considered.

Indeed, many existing theories of bacterial growth, physiology and genetics are formulated exclusively in terms of the isolated bacterium. From an epistemological standpoint this emphasis on the single cell is curious. In practice most research is carried out on cell populations. An enzyme measurement, for example, may be based on an extract from 100 million cells, but conclusions based on the results are often made under the assumption that every bacterium in a population is more or less the same. Such a premise may simplify the interpretation of experimental results, but it is a simplification that is likely to prove invalid in many cases. How exceptional—or how common—are multicellular features in bacteria? In investigating this question I have concluded that most—perhaps virtually all—bacteria lead multicellular lives.

Examples of multicellularity among the bacteria abound; indeed, some of the complex biochemical processes performed by bacteria could not be carried out as effectively without organized groups. Photosynthesis is a process that illustrates this point in several ways. Photosynthetic bacteria, like green plants, rely on solar energy to convert carbon dioxide into organic chemicals. One group of photosynthetic bacteria, known as the cyanobacteria, often grow as connected chains of cells or as intertwined mats; they contain a form of chlorophyll and in many ways resemble multicellular algae. For many years, in fact, the cyanobacteria were thought to be members of the plant kingdom. The multicellular organization aids in light harvesting but yields other benefits too.

Anabaena, an inhabitant of freshwater ponds, is one of the best-known of the photosynthetic bacteria. *Anabaena* is capable of both photosynthesis and nitrogen fixation. These two biochemical processes are incompatible within a single cell because oxygen, produced during photosynthesis, inactivates the nitrogenase required for nitrogen fixation. When nitrogen compounds are abundant, *Anabaena* is strictly photosynthetic and its cells are all alike. When nitrogen levels are low, however, specialized cells called heterocysts are produced. The heterocysts lack chlorophyll but synthesize nitrogenase, an enzyme with which they are able to convert nitrogen gas into a usable form.

James W. Golden and Robert Hasel-

JAMES A. SHAPIRO is professor of microbiology at the University of Chicago. He earned his bachelor's degree in English at Harvard College in 1964 and was awarded a Marshall Scholarship for postgraduate study at the University of Cambridge, where he received his Ph.D. in genetics. He spent a year as a postdoctoral fellow at the Pasteur Institute in Paris, almost two years as visiting professor at the University of Havana and several months at Tel-Aviv University; he was involved in the U.S.-U.S.S.R. microbiology exchange program from 1975 to 1978. In 1973 he joined the faculty at Chicago. This is Shapiro's second article for SCIENTIFIC AMERICAN.

korn of the University of Chicago have shown that differentiation in *Anabaena* involves a form of controlled genetic engineering. In the course of heterocyst differentiation a specific DNA rearrangement occurs that results in the creation of a complete coding sequence for one of the subunits of nitrogenase. Such a rearrangement occurs only in cells that are differentiating to form heterocysts. Comparable DNA rearrangements are involved in the formation of specialized immune-system cells in vertebrates.

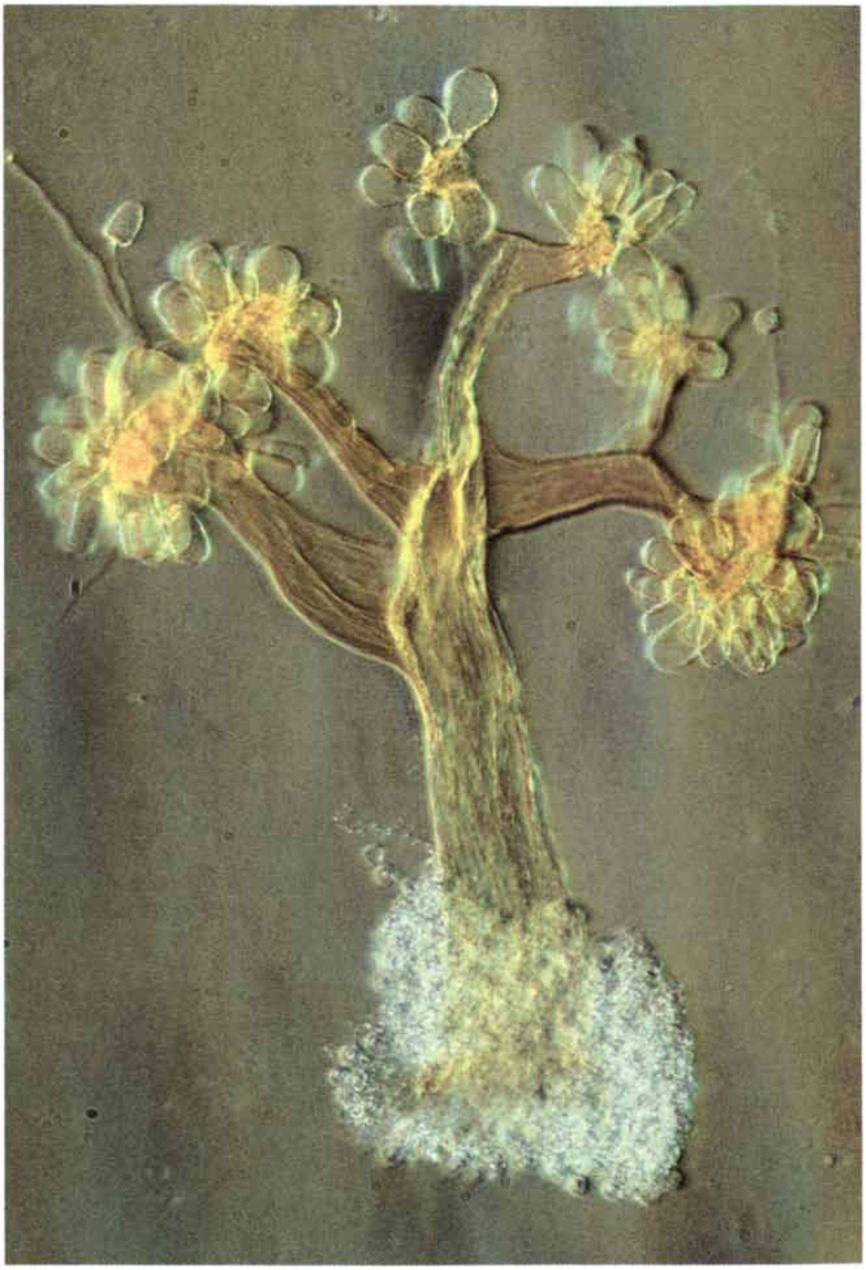
In addition there are submicroscopic channels within each *Anabaena* filament that connect the two kinds of cells. The transport of cellular products (fixed nitrogen to the photosynthetic cells and photosynthetic products to the heterocysts) takes place by way of these channels. In overall character, then, it can be said that *Anabaena* functions more like a multicellular organism than a unicellular one: it relies on division of labor among its cells to carry out specialized and incompatible chemical processes.

More spectacular examples of multicellular behavior can be found among the Myxobacteria, the most morphologically complex of all bacteria. Their elaborate fruiting bodies rival those of fungi and slime molds and have long been an object of scientific curiosity. Myxobacteria are social creatures par excellence and their intriguing, almost psychedelic patterns of aggregation and movement have been recorded in a fascinating series of time-lapse motion pictures produced by Hans Reichenbach of the Society for Biotechnological Research in Braunschweig and his collaborators at the Institute for Scientific Film (I.W.F.) in Göttingen.

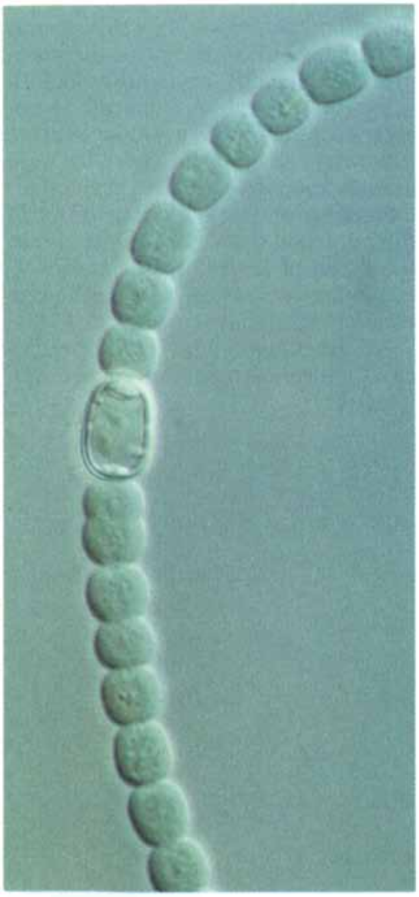
Unlike many bacteria, which periodically enter a dormant stage as individual spores, many Myxobacteria never exist as single cells. Instead they enter dormancy in the form of a multicellular cyst that eventually germinates and spawns a ready-made population of thousands of individuals. Each cyst founds a new population; as the bacteria become more numerous and dense, a number of sophisticated events specific to multicellularity take place. Trails of extracellular slime are secreted and serve as highways for the directed movement of thousands of cells, rhythmic waves pulse through the entire population, streams of bacteria move to and from the center and edges of a spreading colony, and bacteria aggregate at specific places within the colony to construct cysts or, in some species, to form elaborate fruiting bodies. Movement is highly coordinated: when the population migrates over agar, it displaces itself as an intact unit. When an individual cell moves a few microns beyond the edge, it quickly pops back into place as though drawn by an elastic thread.

Even those species of Myxobacteria that do enter dormancy as single-cell spores are social throughout much of their life cycle. When two cells of *Myxococcus virescens* meet, for example, they go through a characteristic ritual: they align themselves side by side and either move off in the same direction or rub alongside each other before separating. Daughter cells can be observed taking part in a similar routine following cell division. These bacteria literally keep in touch with each other!

Predator-prey relationships exist in the microbial world just as they do in the world of larger organisms. Several predatory species of Myxobacteria feed by secreting enzymes that dissolve the outer cell layer of other microorganisms; when the cells burst,



MULTICELLULAR FRUITING BODY of *Chondromyces crocatus*, a species in the group Myxobacteria, is magnified 268 diameters. The structure consists of a central stalk that branches to form specialized clusters of single-cell spores. When the clusters burst, the spores disperse to form new colonies. The micrograph was made by Hans Reichenbach of the Society for Biotechnological Research in Braunschweig.



ANABAENA, a photosynthetic cyanobacterium, forms filaments of cells in freshwater ponds. Most of the cells are photosynthetic, but when nitrogen levels are low, heterocysts develop. The heterocysts, which are capable of nitrogen fixation but not photosynthesis, are larger than the photosynthetic cells and have special storage granules for nitrogen-rich compounds. The Nomarski micrograph was made by the author; the cells are enlarged 1,625 diameters.

the myxobacteria absorb their contents. One species, *Myxococcus xanthus,* has evolved a specialized method of prey capture in response to its aquatic environment. In water it cannot release its digestive enzymes because they would immediately be diluted, as would the prey's nutrients. Jeffrey C. Burnham and his colleagues at the Medical College of Ohio have shown that instead *M. xanthus* constructs spherical colonies containing millions of bacteria. The colonies surround suitable prey organisms, trapping them in pockets on the surface of the sphere, where both the digestive enzymes and the prey contents can be effectively corralled.

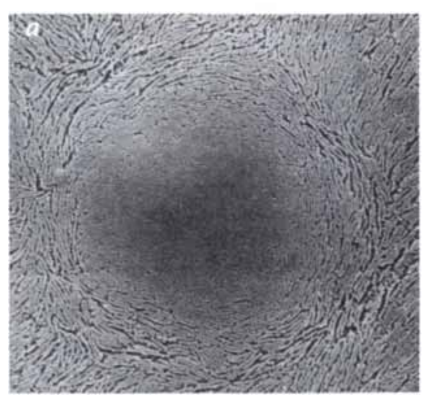
Over the past decade A. Dale Kaiser of the Stanford University Medical School and his colleagues have studied the genetic basis of communication and movement in *M. xanthus.* By searching for mutants that have lost the ability to spread or to form fruiting bodies, they have identified specific regions of the organism's DNA that control for aggregation, motility and differentiation. They found that motility in *M. xanthus* is under the control of two different systems. The *A* (for adventurous) system enables individual cells to move across the substrate; the *S* (for social) system controls the movement of groups of cells. If either system is defective, the cells are still able to spread, but they do so abnormally. If both systems are defective, the cells cannot spread at all. The two systems are surprisingly complex and a great deal of specific genetic information is required for each to be operative. A mutation affecting any one of 23 different genetic loci will eliminate the *A* motility system; at least 10 different loci control the *S* system.

Kaiser and his colleagues have also begun to unravel some of the ways *M. xanthus* cells communicate with one another. They accomplished this by combining different mutant cells that were defective for the same trait but in which the defect resulted from mutations at different loci. They found that when two motility-defective mutants are combined in the same petri dish, they regain motility as long as both mutants stay together. Similarly, if two sporulation-defective mutants are mixed in culture, they will form normal fruiting bodies and sporulate. In some instances such mixed-cell complementation is now known to be mediated by the production of extracellular substances; in other cases it may result either from direct cell-to-cell contact or from the physical presence of two complementary cell types.

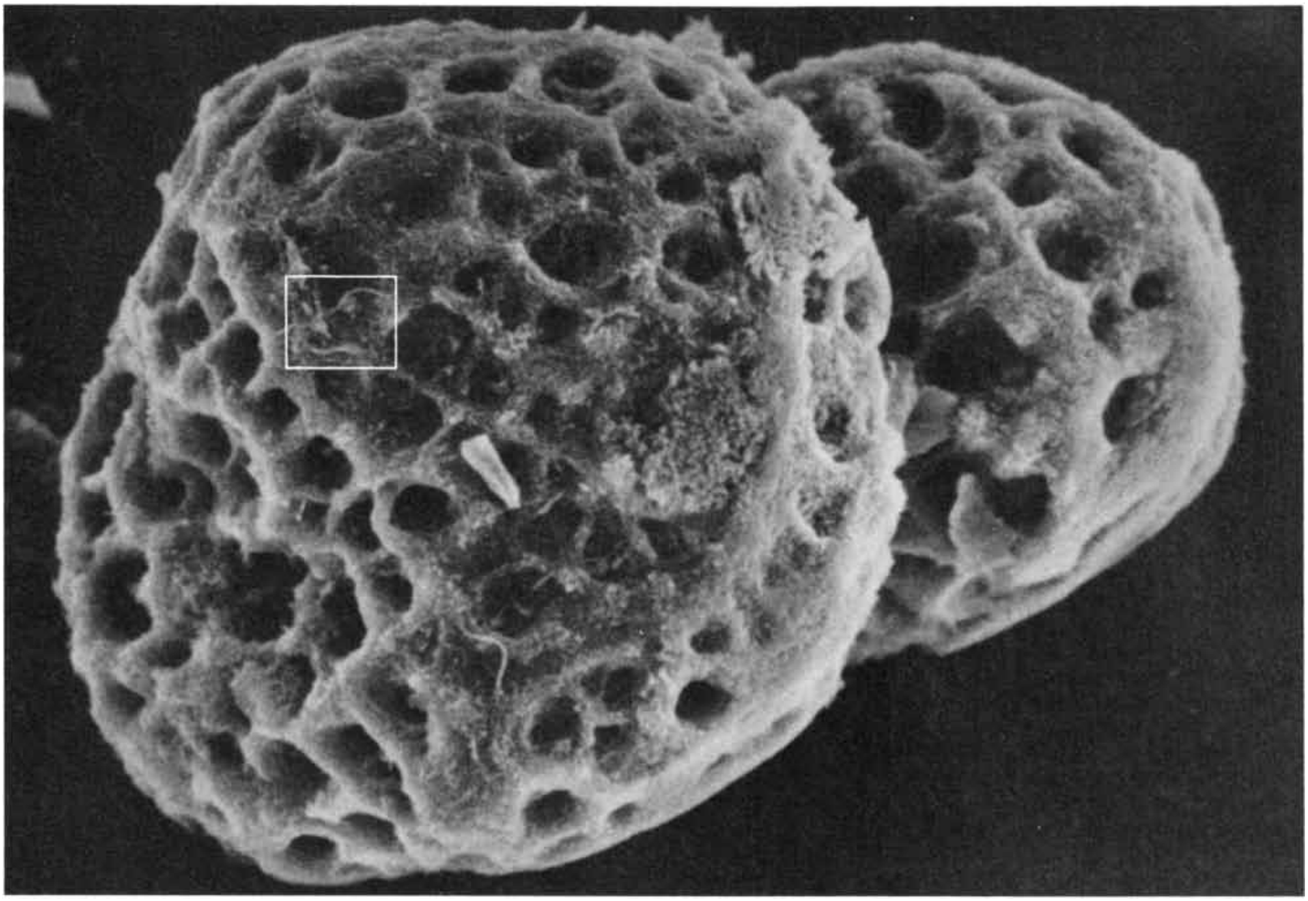
My own studies of multicellular behavior in bacteria began, as is often the case, as the result of a chance observation. A little over five years ago I was experimenting with a genetic-engineering tool, designed by Malcolm J. Casdaban of the University of Chicago and his students, in order to study enzyme expression in *Pseudomonas.* The technique enabled me to join the genes for certain *Pseudomonas putida* enzymes with the DNA sequence from *Escherichia coli* that encodes the enzyme beta-galactosidase. The advantage of beta-galactosidase as a genetic-engineering tool is that it causes certain chemicals to turn color when they are exposed to it. If the beta-galactosidase sequence is linked to the gene for the enzyme under study and the recombinant DNA is inserted into *Pseudomonas,* then when the bacteria are grown on agar containing those chemicals, the amount and distribution of color is a direct measure of the gene's expression.

When I plated my recombinant strains of *P. putida* on such indicator agar, I was astonished to find that every colony exhibited a characteristic flowerlike pattern of staining. I repeated the experiment with engineered *E. coli* and with strains of naturally pigmented bacteria, such as *Pseudomonas cepacia, Serratia marcescens* and *Chromobacterium violaceum.* Each produced its own unique flowerlike pattern. The fact that different strains and species produced distinctive colonies (including species that were naturally colored and not subjected to genetic manipulation) gave me reason to believe colony growth in bacteria is a highly regulated process and is under some form of temporal control. Subsequent studies confirmed that hypothesis. It is now clear to me that colony organization follows certain general rules, which help to explain the existence of general patterns.

The colonies tend to assume a circular configuration, growing outward by adding cells to the perimeter. As a colony spreads across the agar, it is apparent that the pattern of growth consists of both concentric and radial elements. The concentric elements are rings that encircle the colony; the radial elements, or sectors, look like slices of pie. Each sector consists of outwardly growing progeny descended from a common ancestor. Some grow better than others and expand, whereas other sectors merely hold their own or even disappear as the colony gets bigger.



FORMATION OF FRUITING BODY in the Myxobacteria follows a characteristic sequence of steps, as is shown from the far left in these scanning electron micro-

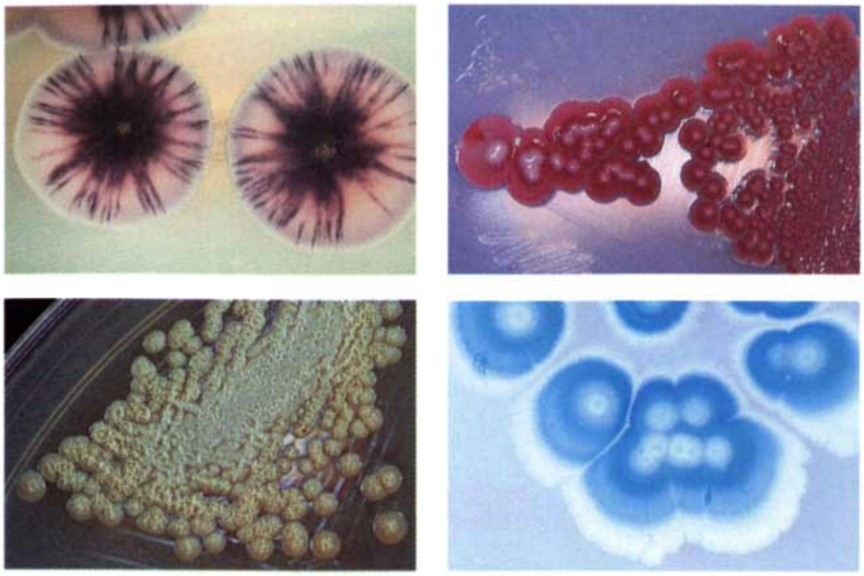


PREDATORY SPHERES are formed by millions of individual cells of *Myxococcus xanthus* as a means of capturing prey in an aquatic environment. Microscopic prey, such as the cyanobacterium *Phormidium luridum* (*inset*), stick to the colony and are eventually digested within pockets on the sphere's surface. The spheres are enlarged 440 diameters in this scanning electron micrograph made by Jeffrey C. Burnham, Susan A. Collart and Barbara W. Highison of the Medical College of Ohio.

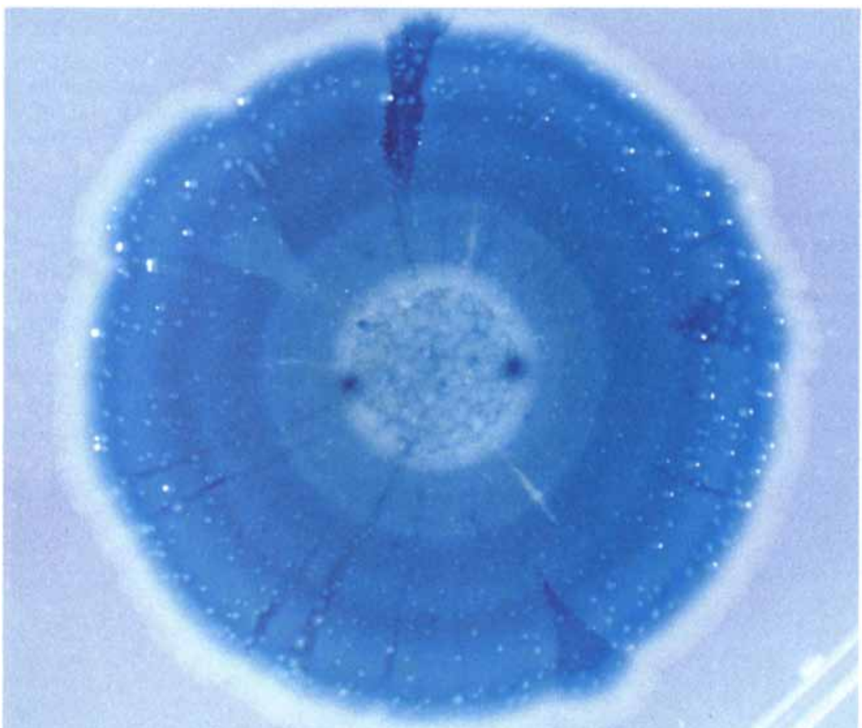
In many cases it is possible to select individual cells from different sectors, grow them in culture and show that the cells in one sectorial culture have heritable properties differing from those of the cells in another sectorial culture; often cells from different sectors can be distinguished on the basis of differences in their DNA. Recently N. Patrick Higgins of the University of Alabama Medical School and I found that differences in DNA between distinct sectors and concentric zones can sometimes be visualized directly by picking colonies up on filter paper, extracting their DNA in situ and then applying radioactive probes to detect specific sequences.

The concentric elements in a colony are less familiar than the sectors and hence more puzzling. The cells in a concentric zone or ring share a common property (such as the level of expression of beta-galactosidase activity) but are not related by common ancestry; they are directly related to bacteria in the preceding and succeed-

graphs tracing the morphogenesis of *Stigmatella aurantiaca.* At first cells form aggregation centers. As the centers accumulate bacteria they bubble upward until the vertical stalk is complete. In some species, such as this unidentified one from the Indiana University campus, the stalk is elaborately branched (*far right*). The *S. aurantiaca* fruiting body is enlarged 450 diameters. Micrographs were made by Gabriela M. Vasquez, Frank Qualls and David White of Indiana University.



FLOWERLIKE COLONIES can be produced by streaking bacterial cultures over an agar plate. Each colony (when it is not crowded) assumes a pattern characteristic of its strain. *Chromobacterium violaceum* (*top left*) naturally produces the pigment violacein and forms purple colonies. *Serratia marcescens* (*top right*) synthesizes prodigiosin; it forms bright red colonies once thought to be drops of blood. *Pseudomonas cepacia* (*bottom left*) is yellow and has a unique surface texture, the result of cell aggregation at the surface. An *Escherichia coli* colony (*bottom right*) carries genetically engineered DNA sequences encoding the enzyme beta-galactosidase; where the enzyme is expressed the colony turns blue. The colonies are at various magnifications.



E. COLI **colony was grown from a single drop of culture that contained thousands of cells. The highly regular, intricate pattern of pigmented rings is characteristic of certain genetically engineered *E. coli* colonies stained for beta-galactosidase activity. Pie-shaped sectors, within which the control of enzyme synthesis has changed, are also apparent. The sectors at five o'clock and ten o'clock have curved edges, indicating that the bacteria within these regions spread faster than the rest of the colony. The concentric rings that run through the ten o'clock sector are displaced outward, suggesting that changes in enzyme activity occurred at similar times both inside and outside the sector. This colony was approximately a centimeter in diameter.**

ing zones, not to one another. If the organization of the colony into distinct concentric zones cannot be explained by heredity, how might it be explained? Some system must exist that bestows common properties on bacteria within a ring and distinguishes them from bacteria in other rings.

One set of clues to the origins of concentric patterns lies in the different ways the sectorial and concentric elements interact with one another. Photographs of colonies often show that concentric rings persist through sectors that grow faster than the rest of the colony. The resulting pattern contains rings that are stretched outward. The stretching shows that the rings are formed not at specific positions on the agar (that is, at particular distances from the center) but at a specific time in the course of colony development. This suggests bacteria have biological clocks that enable them in some way to program cellular differentiation at specific times during development. The rhythmic pulsations of spreading myxobacterial colonies also reflect the operation of a clocklike mechanism. Both biological clocks and the temporal control of development, previously unknown in bacteria, are important features of higher organisms.

Examination of the surface textures of a colony indicates that cellular differentiation also takes place at the level of cellular aggregation. When light is reflected from a colony, various surface textures that are as organized as the pigmentation patterns and also show radial and concentric elements become visible. In many cases the two patterns coincide: a sector defined by color may also display a novel surface structure, and a ring may stand out on the basis of both its distinctive color and its topography.

Clearly colony organization involves more than the distribution of cells with simple biochemical differences. In order to study structural patterns more precisely, I turned to the scanning electron microscope, which visualizes surfaces at high magnification with great depth of field. The scanning electron micrographs revealed that colonies of *P. putida* and *E. coli* are made up of highly differentiated cells that often form distinctive multicellular arrays coincident with the macroscopic organization of the colony. They also revealed that each colony secretes extracellular materials, some of which form a skin or framework over its surface.

It was clear from these studies that

biochemical activity within a bacterial colony is highly organized and spatially restricted: cells in different regions of the colony have different shapes and biochemical properties. In order to identify the unknown factors that control multicellular growth in bacteria, I began to study swarm colonies. As the name implies, these are colonies that grow quickly and cover a large surface area, two characteristics that make them ideal for laboratory experimentation.

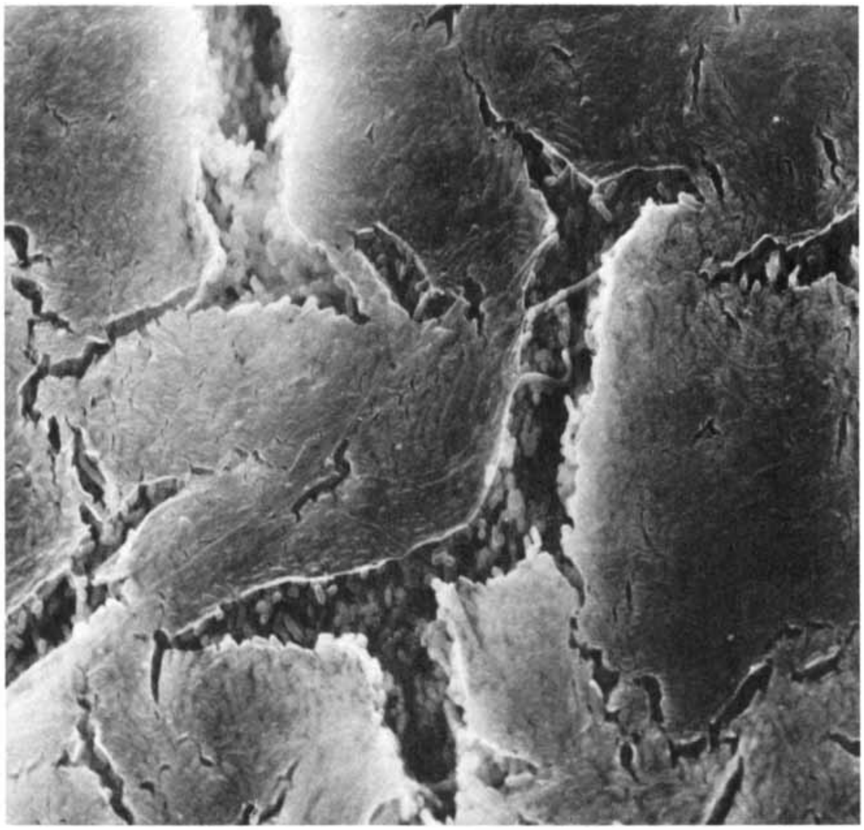
Swarm behavior can be observed in many taxonomically distant species of bacteria. I have focused on one species: *Proteus mirabilis.* Like the Greek god Proteus, this bacterium assumes different forms, producing striking configurations in a petri dish. Over the years two key features of *Proteus* colonies have been noted. One is that there are at least two very different types of cells in a colony: long swarmer cells covered with hundreds of flagellae and short nonswarmer cells with few flagellae. The second feature is that colony development occurs as a tightly programmed rhythmic process.

Swarm colonies develop from an initial population of short nonswarmer cells. As the short cells divide, long swarmer cells begin to appear. They migrate to the periphery of the colony, where they assemble in groups and then pioneer the expansion of the colony by moving out in a series of swirls. The flagellae that cover the surface of swarmer cells rotate and in so doing somehow propel the cells. (It is easy to understand how the spinning flagellae propel a cell through liquid; how such delicate structures are able to propel bacteria over the highly viscous agar surface, however, remains a total mystery.) Observing swarm colonies under a microscope, one sees thousands of flagellae on dozens of swarmer cells moving in synchrony, creating oscillating waves as the cells spread outward from the periphery. Such exquisite coordination prompted Alexander Fleming, the discoverer of penicillin, to wonder in print if *Proteus* could possibly have a nervous system!

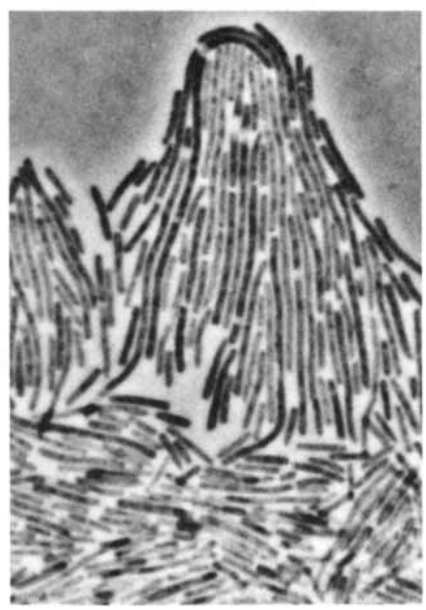
By recording *Proteus* morphogenesis with a time-lapse video camera I have been able to identify distinct periodicities in colony growth, confirming the findings of earlier workers. I discovered that intense activity within a swarm colony can be seen both at the advancing edge, where swarmer cells are rapidly moving outward, and inside the edge, where cell division and streaming activities continue—even when the swarmer cells have stopped advancing. Swarmer cells do not spread indefinitely; they stop after moving a set distance and start spreading again only after a delay, which may be as long as a few hours, depending on conditions such as temperature and the composition of the agar. Swarming is strictly a multicellular activity; an individual cell that gets separated from the rest is unable to advance over the agar unless it is engulfed by another swarmer group, at which time it starts to move again.

Activity inside the edge has its own periodicities and rhythms but is connected to the expansion of the colony as a whole. When the swarmer cells finish one phase of spreading, for example, a series of visible waves composed of more densely aggregated cells moves from inside the swarm zone toward the perimeter. Then there is a thickening of the cell mass in the recently colonized zone and a bubbling of the surface inside the perimeter. These and other exquisitely choreographed postmigration processes produce elaborate textures in the form of terraces on the surface of a fully developed swarm colony.

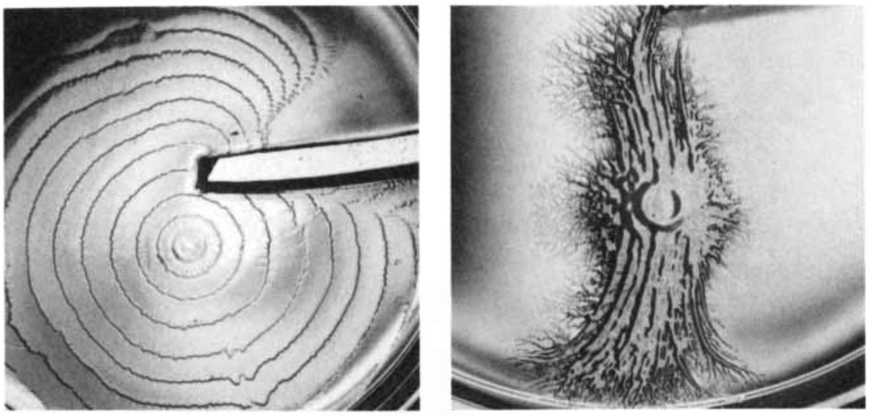
Watching the swarming process, I



SCANNING ELECTRON MICROGRAPH of a *Pseudomonas putida* colony reveals that extracellular material covers its surface like a skin. This superstructure may in some way facilitate communication among cells of the colony. One long, curved cell can be seen bridging a crack in the skin; such cells may also be involved in intercellular communication within the colony. The colony is enlarged 2,300 diameters.



LONG SWARMER CELLS can be seen at the edge of a *Proteus mirabilis* colony. The cells are preparing to move across the agar as a group; they do so by rotating their flagellae in synchrony. The cells are enlarged 600 diameters in this micrograph made by S. A. Sturdza of the Cantacuzino Institute in Bucharest.

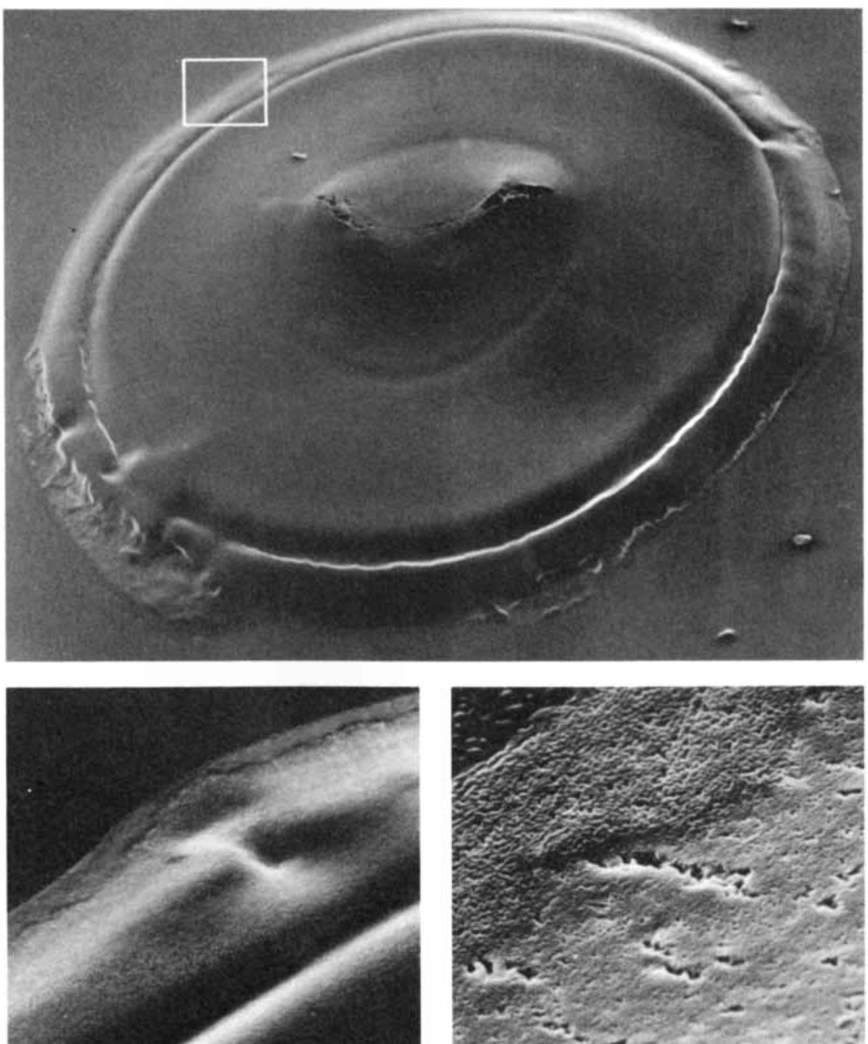


PROTEUS MIRABILIS **mutants show geometries that provide clues to the way colony spreading is controlled. The mutant (*left*) forms regular repeating terraces in a nine-centimeter petri dish. If a trench is cut in the agar, however, spreading stops after a few cycles. The fact that swarming is blocked in the shadowed zone indicates that a chemical signal must emanate from the center. Morphogenesis appears to be under whole-colony control; it is not regulated solely by the migrating edge. The mutant (*right*) has lost its circular symmetry and has a markedly different pattern. The bacteria began by forming thick columns along stress lines in the agar; after these columns grew to a certain size they ramified into smaller perpendicular processes that in turn formed smaller branches. This pattern suggests that although the mutant lacked the ability to produce circular colonies, it retained some kind of directional control. The length of this colony was about four centimeters.**

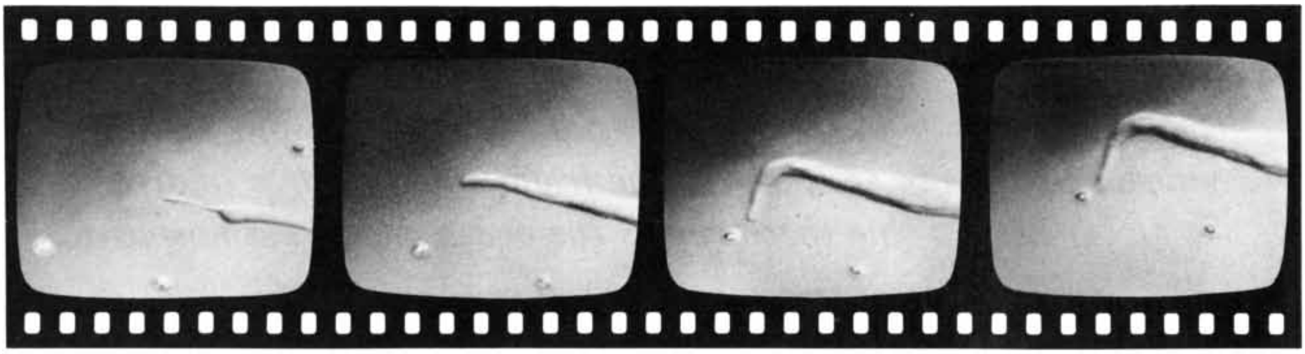
wondered whether it would be possible to learn about the systems that control this intricate and regular behavior. In biological research important clues can come from looking at the response of organisms to unusual circumstances. If conditions in the petri dish changed, would a swarm colony alter its behavior?

Two kinds of evidence suggest that swarming behavior is indeed regulated. One is that when swarming is interrupted by chemical or physical obstacles, or by interference with the growing cell mass, specific morphological responses take place. Observation of the geometries of the swarm colonies after they have been artificially manipulated (or following spontaneous accidents that deform the regular outlines of the colonies) makes it clear that both temporal and directional controls influence colony growth. In particular, chemicals diffusing through the agar appear to play important roles in guiding colony spread.

The second line of evidence is that morphogenesis is under hereditary control. For one thing, each naturally isolated strain of *P. mirabilis* has its own characteristic mode of swarm-colony development. For another, one can obtain from these natural isolates various mutants that can still spread but that have geometries markedly different from those of their progenitors. Some mutants form periodic swarm terraces that are more closely spaced than those of their parents, whereas others have no terraces at all. One particularly striking mutant has lost its circular symmetry altogether and grows in a branching pattern influenced by stress lines in the agar. Clearly the genes that regulate morphogenesis have been altered in these mutants, whereas the ability to swarm has not been eliminated. A detailed biochemical explanation for how morphogenetic control systems might operate in *Proteus* (or in *Myxococcus* or *E. coli,* for that matter) has not yet



DISTINCTIVE MORPHOLOGY of cells in different zones is demonstrated in a series of scanning electron micrographs of an *E. coli* colony. The 68-hour colony (*top*), formed from an initial population of 100,000 cells, was five millimeters in diameter. When the leading edge is enlarged 100 diameters (*bottom left*), a distinct boundary is visible between two groups of differentiated cells. At greater magnification (750 diameters) the outermost zone (*bottom right*) is seen to be made up of large cells arranged irregularly, whereas the inner zone has smaller cells grouped in roughly parallel arrays.



PURPOSEFUL MOVEMENT of *M. xanthus* cells is shown in a series of frames on a video monitor. At the bottom of the screen is a latex bead, five micrometers in diameter, toward which the cells, collectively called a flare, will orient. The tip of the flare turns toward the bead at 18 minutes (*second frame*); at 33 minutes (*fourth frame*) the flare reaches it. After sensing that the bead is inedible the flare will move on. These images were made by Martin Dworkin of the University of Minnesota.

been found. Nevertheless, the hereditary specificity of the developmental phenomena suggest that studying these systems will yield important lessons about coordinating the behaviors of large numbers of bacteria in a spatially and temporally defined way.

The view that bacteria are sentient creatures, able to receive, process and respond meaningfully to external signals, has been gaining ground over the past two decades as investigators spend more time exploring the mysteries of bacterial behavior.

Martin Dworkin of the University of Minnesota has recently provided a graphic demonstration of multicellular responsiveness in the microbial predator *Myxococcus xanthus.* He found that roaming flares, or groups, of *M. xanthus* cells perceive clumps of prey bacteria (or even glass or plastic beads) on an agar surface, make sharp turns toward the objects and then move directly to them. Once there, the *Myxococcus* flares are able to tell whether the objects are edible or not. If the objects are edible, the flares stay to feed; if they are not, the flares turn away and continue their searching behavior. Such purposeful behavior has traditionally been thought to operate only in larger organisms.

What practical value, if any, do these findings have? The biotechnology industry, eager to turn to genetically engineered bacteria as factories for the production of complex biochemicals, will undoubtedly benefit from the knowledge that bacterial cells specialize and control protein synthesis with the aid of intercellular communication signals. In the field of biodegradation the application of bacteria to remove toxic chemicals from polluted soils and water sources may be enhanced by greater understanding of multicellular processes. It may even be possible to introduce multicellularity characteristics that optimize productivity or improve the capacity of specific strains to degrade synthetic compounds. Understanding the behavior of bacteria may also make it easier to monitor the release of genetically engineered organisms.

In medicine greater understanding of bacterial behavior may lead to increased efficacy of drug treatments. J. William F. Costerton and his colleagues at the University of Calgary in Alberta recently described a patient who suffered from recurrent bloodstream infections because a colony of *Staphylococcus aureus* had formed on the lead of his cardiac pacemaker. Individual cells would break away from the parent colony periodically and infect his bloodstream. Although the individual cells were sensitive to penicillin, the colony itself was drug-resistant: like many bacterial colonies, it was protected by a coating of extracellular slime. In order to end the chronic infection the only solution was to remove the pacemaker (and its colony).

In other cases of bacterial pathogenesis there is a clear correlation between the tendency of cells to aggregate and their ability to establish an infection. It has been known for 25 years that the organism that causes gonorrhea, *Neisseria gonorrhoeae,* forms several types of colonies on laboratory media. Cells from one type are virulent, whereas those from another are not. Since a successful infection requires that the disease organism colonize its host, I suspect pathogenesis is directly related to this organism's ability to form multicellular organizations (as reflected in the kinds of colonies it builds).

Bacteriology's early pioneers, such as Louis Pasteur, recognized that there are many lessons to be learned from these smallest of living cells. Today much more is known about the intricacy and complexity of this group of organisms. If, as I have proposed here, bacteria possess elaborate developmental and behavioral capabilities typical of higher organisms, then it is likely that detailed explanations of how these small cells communicate will influence views of information processing in all organisms.

Although bacteria are tiny, they display biochemical, structural and behavioral complexities that outstrip scientific description. In keeping with the current microelectronics revolution, it may make more sense to equate their small size with sophistication rather than with simplicity. There is little reason to doubt that insights gained from studying the interactions of billions of bacteria, living together in a volume of less than a few cubic millimeters, will enhance understanding of all forms of life.

FURTHER READING

NATURE OF THE SWARMING PHENOMENON IN *PROTEUS.* Fred D. Williams and Robert H. Schwartzhoff in *Annual Review of Microbiology,* Vol. 32, pages 101-122; 1978.

MYXOBACTERIA: DEVELOPMENT AND CELL INTERACTIONS. Edited by Eugene Rosenberg. Springer-Verlag, New York, 1984.

ORGANIZATION OF DEVELOPING *ESCHERICHIA COLI* COLONIES VIEWED BY SCANNING ELECTRON MICROSCOPY. James A. Shapiro in *Journal of Bacteriology,* Vol. 169, No. 1, pages 142-156; January, 1987.

The following films are available in the U.S. from Audio-Visual Services, Pennsylvania State University:

BACILLUS CIRCULANS—AUFBAU UND VERHALTEN (*B. CIRCULANS*—GROWTH AND BEHAVIOR). Silent film, E183. Institut für den Wissenschaftlichen Film.

PROTEUS—BEWEGUNGSVERHALTEN (*PROTEUS* SWARMING BEHAVIOR). Silent film, E271. Institut für den Wissenschaftlichen Film.

Polynyas in the Southern Ocean

They are vast gaps in the sea ice around Antarctica. By exposing enormous areas of seawater to the frigid air, they help to drive the global heat engine that couples the ocean and the atmosphere

by Arnold L. Gordon and Josefino C. Comiso

During the austral winter (the months between June and September) as much as 20 million square kilometers of ocean surrounding Antarctica—an area about twice the size of the continental U.S.—is covered by ice. For more than two centuries, beginning with the voyages of Captain James Cook in the late 18th century, explorers, whalers and scientists charted the outer fringes of the ice pack from on board ship. Nevertheless, except for reports from the few ships that survived after being trapped in the pack ice, not much was known about the ice cover itself.

Since 1973, however, satellites bearing passive microwave sensors have made it possible to survey the ice cover routinely from space, and these observations have brought about a marked change in investigators' view of the Antarctic ice. One of the most surprising findings of the satellite era is that the ice cover is not at all a continuous blanket. Within its borders there are many small breaks, on the scale of from one to 10 kilometers, called leads. More surprisingly, there are sometimes vast regions—up to 350,000 square kilometers in area—that are completely free of ice.

ARNOLD L. GORDON and JOSEFINO C. COMISO combine sea- and satellite-based research in their study of polynyas. Gordon is professor of physical oceanography at Columbia University and is on the senior staff of Columbia's Lamont-Doherty Geological Observatory. His research is focused on large-scale circulation and mixing in the ocean and their relation to the global climate. Comiso is a physical scientist at the National Aeronautics and Space Administration's Goddard Space Flight Center. He went to Goddard in 1979 after doing research in experimental particle physics at various institutions. At Goddard he has helped to demonstrate the utility of passive-microwave and infrared data from satellites in the study of oceanic and atmospheric processes.

In these regions, called polynyas, the surface waters of the Southern Ocean (the ocean surrounding Antarctica) are bared to the frigid polar atmosphere. Polynyas and their effects are only incompletely understood, but it now appears they are both a result of the dramatic interaction of ocean and atmosphere that takes place in the Antarctic and a major participant in it. The exchanges of energy, water and gases between the ocean and the atmosphere around Antarctica have a major role in determining the large-scale motion, temperature and chemical composition of the ocean and atmosphere throughout the globe. Polynyas accelerate these processes by exposing the surface of the Southern Ocean directly to the atmosphere.

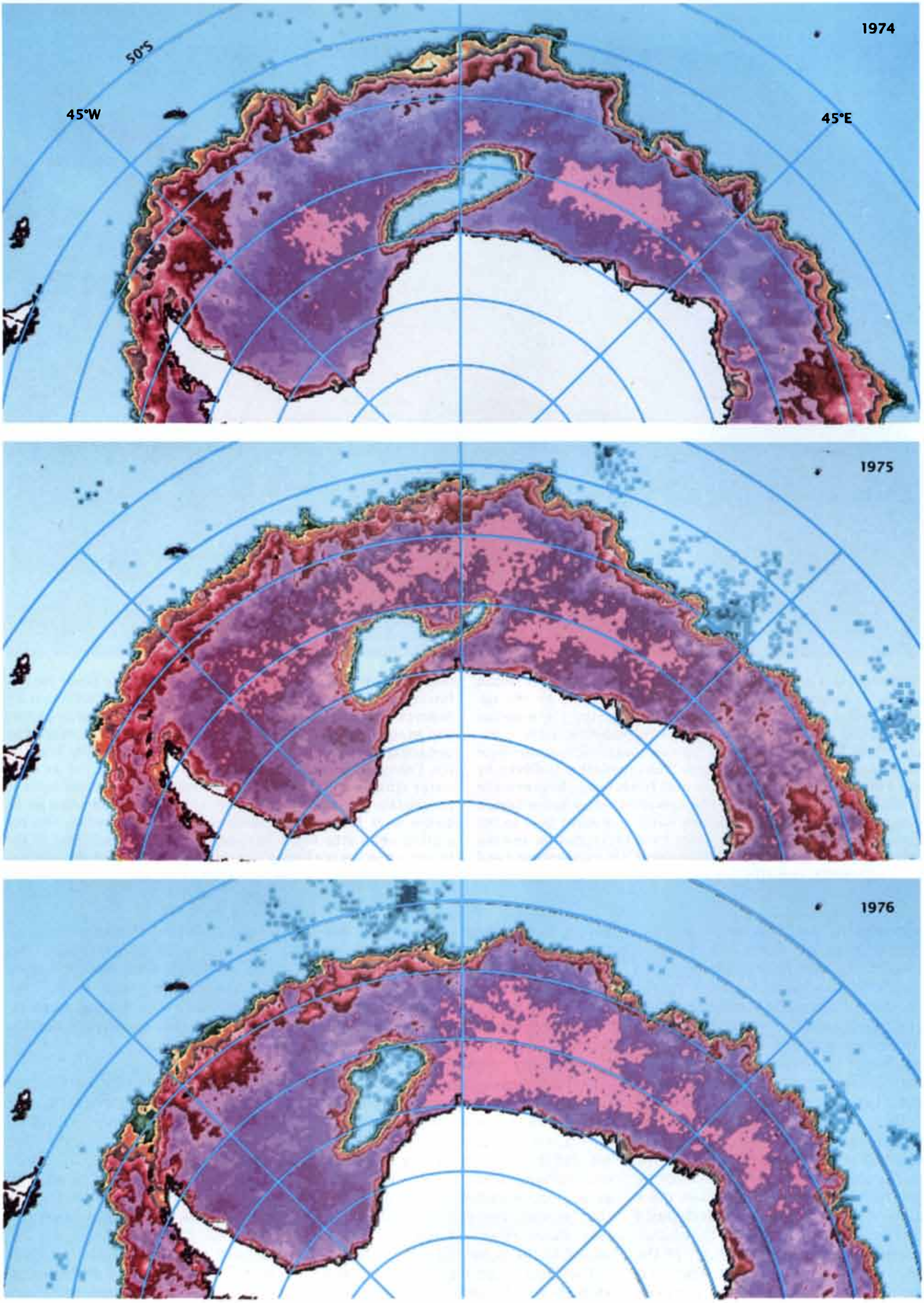
There are two general kinds of polynyas: coastal polynyas and open-ocean polynyas. Coastal polynyas develop when strong local winds originating on the Antarctic continent blow ice away from the shoreline as it freezes, leaving an ice-free area between the coast and the ice pack. Even before satellite data were available, investigators had known about the existence of coastal polynyas. They were very surprised, however, to learn from satellite observations how widespread coastal polynyas are and how large they can become. (In some coastal polynyas the distance from the shoreline to the border of the ice pack is as much as 50 to 100 kilometers.)

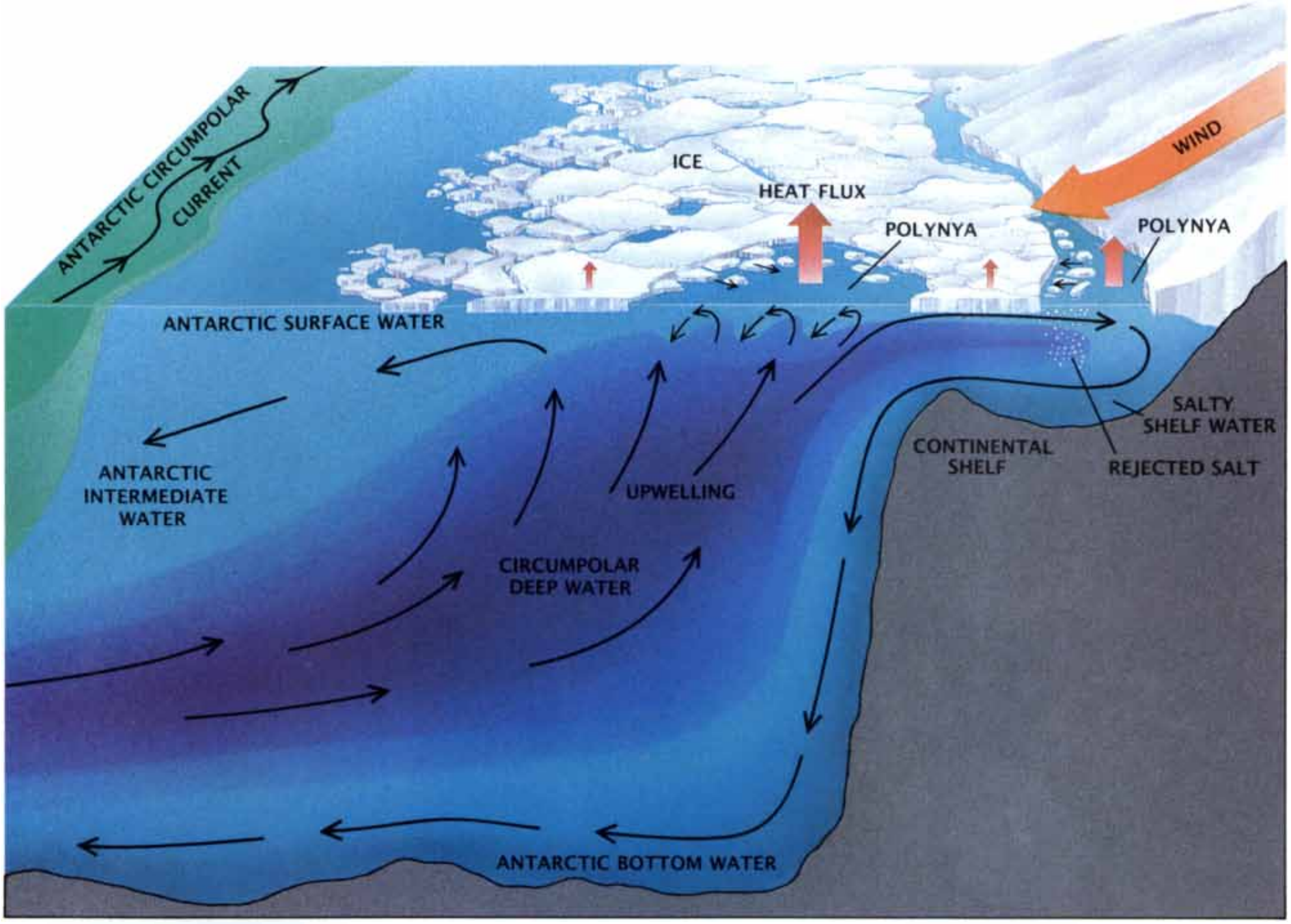
Open-ocean polynyas form within the body of the ice cover, far from the coast. Some of these are by far the largest and longest-lived polynyas. It is not yet fully understood what forces create and sustain open-ocean polynyas, but ship and satellite data gathered by us and by other investigators are enabling oceanographers to devise reasonable hypotheses.

In order to understand the specific forces that create and sustain polynyas and the effects polynyas can have, one must first understand the role the Southern Ocean plays in the general circulation of the world ocean and in the global climate as a whole. Because the Southern Ocean has extensive open connections to the rest of the world ocean, and because it is much larger than the polar ocean in the Arctic region, the Antarctic plays a larger role in the global ocean circulation than the Arctic does.

When solar radiation warms the tropical and subtropical oceans, the resulting heat diffuses downward into deeper water. Heat from the sun also causes relatively high rates of evaporation, which raises the salinity of the surface waters; the excess salt, like the heat, diffuses into the depths. The warmth and salinity of the deep water are also enhanced by deep-reaching mixing events in the North Atlantic, in which relatively warm, high-salinity surface water is mixed downward. The resulting mass of warm, salty deep water gradually spreads toward the Southern Ocean. In the Antarctic this water, now called the Circumpolar Deep Water, rises toward the surface. There it gives up some heat to the

WEDDELL POLYNYA, an enormous ice-free region amid the ice cover near the Weddell Sea, formed during three consecutive Southern Hemisphere winters. In these satellite images, which were made in September of 1974 (*top*), 1975 (*middle*) and 1976 (*bottom*), the white region represents the Antarctic landmass and colored regions represent ocean areas covered by various concentrations of sea ice. Pink and purple regions are almost completely covered by ice and light blue regions are ice-free. In summer the polynya disappeared with the melting of the ice cover. At its largest the polynya measured about 350 by 1,000 kilometers. It had measurable effects on the temperature of the underlying ocean at depths as great as 2,500 meters.


1974
50°S
45°W
45°E
1975
1976



MERIDIONAL CIRCULATION PATTERN of the Southern Ocean (the ocean surrounding Antarctica) is dominated by the upwelling of a warm, salty water mass called the Circumpolar Deep Water and its transformation into Antarctic Surface Water, which ultimately sinks to become Antarctic Intermediate Water and Antarctic Bottom Water. The circulation is driven by wind and the exchange of heat and fresh water between the ocean and the atmosphere. The upwelled water is converted into cold, relatively fresh surface water as it loses heat to the atmosphere and gains fresh water from precipitation and the melting of ice. Some of the surface water moves northward and gradually sinks, contributing to the formation of Antarctic Intermediate Water. Another fraction of the surface water moves toward the pole; it eventually sinks and flows northward as Antarctic Bottom Water. The circulation is influenced by polynyas. In so-called open-ocean polynyas (*center*) heat is vented to the atmosphere rapidly because there is no insulating layer of ice. Convection currents, in which warm water rises as cold water sinks, accelerate the movement of warm water toward the surface. So-called coastal polynyas (*right*) form when ice is blown away from the continent as fast as it freezes. The resulting open area vents heat rapidly, and the salt that is rejected when sea ice forms increases the density of the surface water and accelerates the formation of Antarctic Bottom Water.

atmosphere and becomes a cold, dense mass of water, which sinks toward the sea floor. This so-called Antarctic Bottom Water then moves northward along the ocean floor, traveling well beyond the Equator; it is the major mass of bottom water in the world. As it moves north it gradually mixes upward, and the cycle of warming and cooling begins again.

The Southern Ocean is thus one part of an enormous heat engine that drives the motions of much of the world ocean. In addition, the massive overturning of water that takes place in the Antarctic provides a mechanism by which the chemistry of the atmosphere can affect the chemistry of the deep ocean and vice versa. The overturning in a sense ventilates the deep ocean, by drawing water to the surface, allowing it to approach equilibrium with the temperature and composition of the atmosphere and forcing it downward again. The process cools the deep ocean, lowers its salinity and restores oxygen levels that have been depleted by organisms. The deep-ocean overturning may also be significant in establishing a rough balance between the oceanic concentrations of such dissolved gases as carbon dioxide and the levels of those gases in the atmosphere. Hence the overturning is an important factor in such events as the "greenhouse warming" that may take place as carbon dioxide gas is added to the atmosphere.

The heat engine's effects do not stop there. The overall rate at which heat is transferred from the Equator to the pole helps to determine the degree to which temperature varies with latitude on a global scale. This in turn has a part in driving many of the processes that govern the earth's weather and climate.

Perhaps the most important forces driving overturning in the Southern Ocean are buoyancy and the wind. Buoyancy gives the ocean a layered structure, in which masses of water are stratified according to their density (which in turn is determined by the water's temperature and salinity). The wind, on the other hand, tends to disturb this stratification. In addition, the exchange of heat and fresh water between the surface water and the atmosphere alters the density of the surface water, and so it can stimulate convection currents (in which buoyant water moves upward and denser water moves downward) that lead to a rearrangement of the stratification.

The top layer of water in the Southern Ocean, extending to a depth of about 100 meters, is called the Antarctic Surface Water. The Antarctic Surface Water is cold because of its proximity to the extremely cold Antarctic atmosphere, and it is relatively fresh because in the Southern Ocean precipitation (and the melting of Antarctic glacial ice) exceeds evaporation.

North of about 65 degrees south latitude, prevailing winds push much of the surface water farther to the north—away from the pole—while south of that latitude other winds push surface water to the south, toward the pole. The resulting divergence at 65 degrees south causes an upwelling: deeper water is lifted to replace the water that has been pushed north or south. The effect is to bring the warmer, saltier layer below the Antarctic Surface Water up toward the surface, where it is converted into Antarctic Surface Water.

The source of the warmer water is the Circumpolar Deep Water, which is drawn to the Southern Ocean in part by the wind-driven upwelling. Although the Circumpolar Deep Water, which extends to a depth of roughly 2,000 meters, is warmer than the surface water, it is still slightly denser because it has relatively high salinity. For that reason it tends to remain below the surface water.

Nevertheless, the surface water and the deep water south of about 60 degrees south latitude are so similar in density that their stratification pattern is only marginally stable. Very small forces can disturb it. Those forces are provided by the wind and by changes in the freshwater balance of the surface layer. Strong winds can greatly enhance the mixing of the two layers, particularly if ice floes are present to help create turbulence. The formation of sea ice in the winter also has a role in this mixing: when seawater freezes, the ice crystals reject much of the seawater's salt, expelling it in a concentrated brine that increases the salinity of the surface layer and thereby makes it more susceptible to mixing with the underlying deep water.

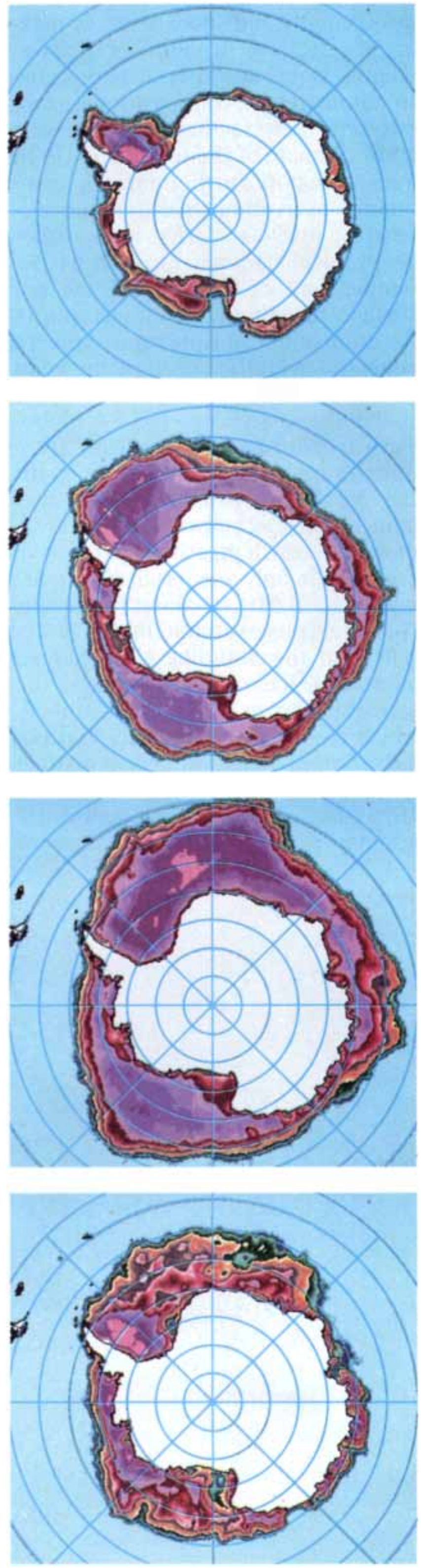
When the Circumpolar Deep Water mixes upward as a result of these processes, it vents heat to the atmosphere and cools, becoming denser. At the same time much of the surface water continues to be pushed either toward or away from the pole by the prevailing winds. The poleward-flowing portion continues to increase in density as it mixes with the brine rejected by newly forming sea ice. When this surface water reaches the Antarctic continental shelf, it sinks, flowing off the continental shelf to the ocean bottom, where it forms the Antarctic Bottom Water. In this way the effects of buoyancy and wind have combined to draw warm deep water from northern latitudes, transfer its heat to the Antarctic atmosphere and send it sinking toward the ocean floor.

Although the formation of the ice cover acts to accelerate this heat engine, the presence of the ice itself paradoxically tends to impede the deep-ocean overturning. The ice cover insulates the ocean surface, inhibiting the venting of oceanic heat. Some heat is still conducted to the atmosphere through the ice, but conduction is not nearly as efficient as the direct exposure of open water to the atmosphere would be; a layer of ice one meter thick (somewhat thicker than most of the Antarctic ice cover) can reduce the heat loss of the winter ocean by a factor of as much as 10 to 100.

How do polynyas, in which the ice cover is breached, affect these processes? This question would be impossible to address without data from satellites. Satellite observations provide consistent, repetitive and long-term coverage, and they also make it possible for investigators to follow changes that take place over time frames as short as a week or even a single day.

Satellite-borne sensors can operate in a variety of wavelengths and observing modes. The sensor that has shown itself to be the most versatile and useful for observations in polar regions is the passive microwave radiometer, which detects the radiation having wavelengths between one millimeter and one meter that is emitted naturally by various materials. This radiation is emitted during dark periods as well as during daylight hours and passes through most cloud cover, making it possible to observe a given region at any time of day and through almost any kind of weather. The microwave radiometer is ideal for monitoring sea-ice cover, because at some microwave wavelengths the emissivity of ice and water is very different.

The amount of radiation emitted in a given wavelength by an area of sea ice depends on a number of variables, including the ice's temperature, thick-



EXTENT of the Antarctic ice cover varies from season to season. These satellite images were made in (*from top to bottom*) March, June, September and December of 1984—respectively the southern summer, fall, winter and spring. In the southern summer ice covers about four million square kilometers of ocean, and in the winter it covers about 20 million.

ness, salinity and snow cover. By making observations in a number of wavelengths simultaneously, it is possible to eliminate the complications introduced by these factors and to derive a good estimate of the total ice cover in a given region. Furthermore, by combining observations in different wavelengths, one can derive many other geophysical parameters, such as the temperature of the ice and the sea surface, the amount of water vapor in the atmosphere, the position of the ice edge and even the speed of the wind over the ocean.

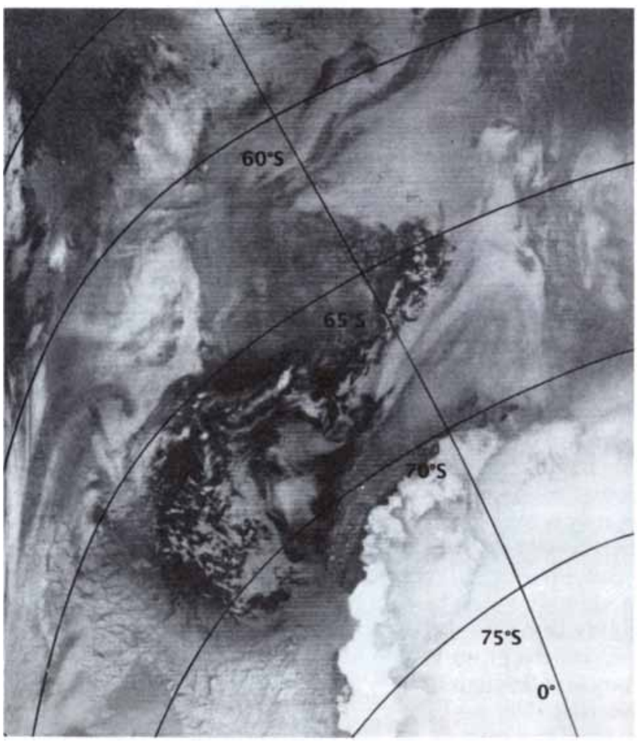
The first spaceborne passive microwave sensor was the Electrically Scanning Microwave Radiometer on board the *Nimbus 5* satellite, which was launched in December, 1972. This radiometer, which measured microwave radiation in only one frequency, transmitted good data for about four years. For later observations investigators relied on the Scanning Multichannel Microwave Radiometer on board the satellite *Nimbus 7*. This instrument, which recorded the intensity and polarization of radiation in five distinct frequencies, was in operation from 1978 until 1987, when the microwave scanner was turned off because of irregularities in the satellite's orientation. In 1987 a new sensor, called the Special Sensor Microwave Imager, was launched as part of the Defense Meteorological Satellite Program, ensuring continuous satellite coverage.

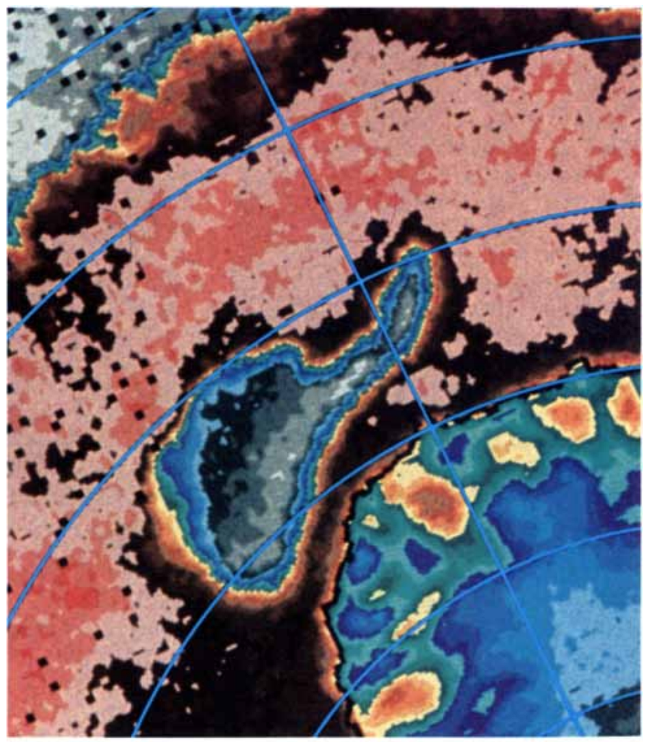
The existence and extent of polynyas are among the most interesting findings to come from satellite observation of the Antarctic. As we have mentioned, there are two kinds of polynyas: coastal polynyas and open-ocean polynyas. Satellite data, shipboard observations and theoretical analysis have now shown that the two kinds of polynyas are the result of very different causes and have quite different effects on the deep-ocean overturning.

The coastal polynyas are essentially sea-ice factories. They appear when local winds push newly forming sea ice away from the continent as quickly as it freezes. This exposes an area of open ocean on which more ice can form, continuing the process. Coastal polynyas are also called latent-heat polynyas, because the heat they release to the atmosphere is primarily in the form of the "latent heat" that liquid water gives off while it hovers at a steady temperature in the process of freezing. The heat flux to the atmosphere due to latent-heat polynyas is thought to be in excess of 300 watts per square meter, enough to add a 10-centimeter-thick layer of new ice (which is subsequently removed by the wind) to the coastal region every day. The ice generated in latent-heat polynyas may supply much of the sea ice in the adjacent ocean.

Meanwhile, as the ice forms within the polynya, the brine it rejects raises the salinity of the water overlying the continental shelf and helps to drive the formation of the Antarctic Bottom Water. According to rough calculations based on satellite data, the formation of ice in coastal polynyas leaves behind enough salt to form bottom water from Antarctic Surface Water at the rate of about 10 million cubic meters per second.

The enormous scale at which coastal polynyas create ice, cool the ocean and add salt to the Antarctic Bottom Water took oceanographers completely by surprise when it was first deduced from satellite observations. The knowledge we have gained about coastal polynyas has helped to explain several previously puzzling features





SATELLITE IMAGES of the Weddell Polynya made in August, 1975, illustrate the advantages of passive-microwave imagery (*right*) over conventional infrared imagery (*left*). In the infrared image the polynya is visible as the dark area at the center, but cloud cover makes it impossible to determine the polynya's extent. In the passive-microwave image it is possible not only to determine the precise shape of the polynya but also to identify such important features as the border between the Antarctic ice cover and the open ocean (*boundary between blue and gray at top left*) and the edge of the continental landmass (*orange border at bottom right*). In this image the various colors represent the "brightness" of the microwave radiation emitted by materials on the earth's surface. White and gray indicate the faintest emission, which signifies areas of open water.

of the Antarctic ocean, such as the high salinity of the waters overlying the continental shelf (which had been measured during the austral summers) and the large quantities of Antarctic Bottom Water that flow from the Antarctic into the world ocean.

The forces that create and sustain open-ocean polynyas are somewhat more complex than those underlying coastal polynyas. Open-ocean polynyas probably form because of convection cells: vertical circulation patterns in which, over an area a few kilometers across, warm water (in this case water from the Circumpolar Deep Water) rises to the surface, cools and then sinks, to be replaced by more rising warm water. In open-ocean polynyas the continuously rising warm water would melt any ice on the surface and prevent further ice from forming. Open-ocean polynyas are also called sensible-heat polynyas, because the process in which they give up heat is directly associated with a change in temperature—it can be sensed.

Because the layered structure of the Southern Ocean is only marginally stable, isolated convection cells can form under the sea ice in many places. Why do only some of them lead to polynyas? The answer may have to do with spatial scale. When convection is triggered under a cover of sea ice, the initial burst of heat rising to the surface melts most if not all of the ice immediately above the cell. This creates a film of fresh water at the surface. The fresh water is less dense than the warm salty water, and so it is stable: it damps out the convection, because the warm water cannot rise past it to the surface. Even if that initial sheet of fresh water were not sufficient to damp out the convection, other sections of ice would flow into the polynya and melt, perhaps eventually providing enough fresh water to halt the convection.

If the polynya initially had a large area, however, it could be self-sustaining. The surface area of a polynya, within which convection occurs, is proportional to the square of its radius. The amount of ice that can float into a polynya is proportional to its perimeter, which is directly proportional to the radius. Thus if a polynya had a radius above some minimum threshold, ice could not flow in quickly enough to stop the convection and the polynya would survive.

For reasons that have to do with the fluid dynamics of the rotating earth, the convection cells in the Southern Ocean cannot have a diameter greater than about 10 to 30 kilometers. A large, self-sustaining polynya would therefore probably have to consist of a number of convection cells pressed together "shoulder to shoulder."

Sensible-heat polynyas have a number of important effects. For one, their convection cells draw up warm water from the Circumpolar Deep Water and accelerate the process in which heat is transferred to the atmosphere. It is not yet known how important this effect is in the overall action of the global heat engine, because it is not known whether large polynyas form frequently enough to be climatically significant. The effect could be substantial, however, because the ocean sheds heat from open water so much more efficiently than through ice.

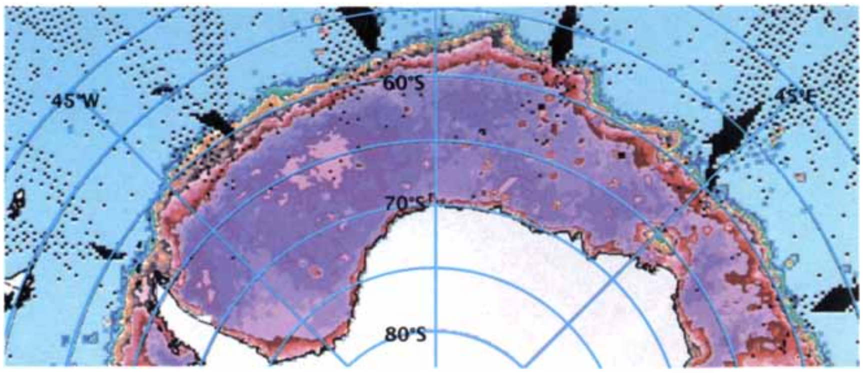
Sensible-heat polynyas also promote chemical interaction between the ocean and the atmosphere. Unlike water under the ice pack, water exposed in a polynya can exchange gases relatively freely with the atmosphere, absorbing some and releasing others, before it sinks again. Although the rapid rate of overturning within a polynya would prevent any specific water parcel from attaining full equilibrium with the atmosphere, it would still allow significant exchange of gases. The existence of large sensible-heat polynyas could thus alter the nature of the ocean-atmosphere interaction and thereby have a great effect on the chemistry and climate of the atmosphere and the deep ocean worldwide.

Much of our knowledge concerning open-ocean polynyas has come from studying one spectacular polynya that formed in the mid-1970's near the Weddell Sea. The Weddell Polynya was remarkable for several reasons. One reason was its sheer size: at its largest it measured some 350 by 1,000 kilometers.

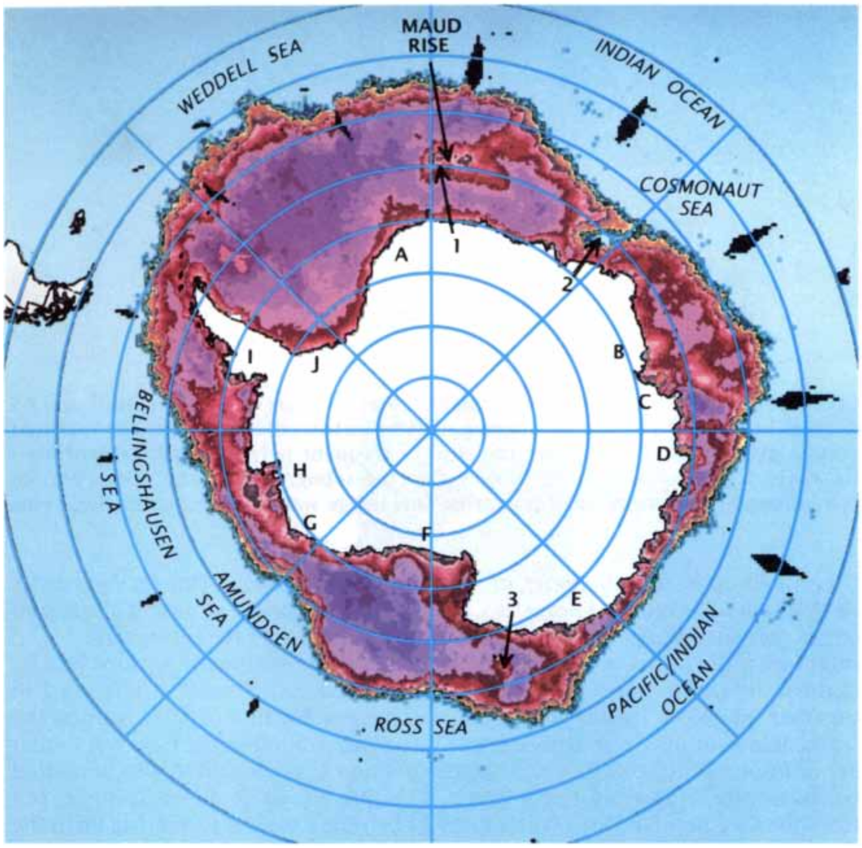
Another remarkable feature was its persistence. It was first observed in the austral winter of 1974. During the following summer the ice cover in that region of the Southern Ocean melted completely, as it normally does, but when the ice cover formed again in the winter of 1975, the Weddell Polynya re-formed as well. It had moved slightly to the west (in keeping with the prevailing surface current in that region), but its shape and size were much the same as they were in 1974. The following year saw the same sequence of events: when the ice cover re-formed in the winter of 1976, the polynya reappeared, slightly to the west of its 1975 position and somewhat smaller.

In 1977 the Weddell Polynya did not appear, but its persistence over the three previous winters indicates the existence during the summer of some "memory" mechanism, some large-scale residual effect that could recreate the polynya after the summer melt. It is likely that the key to this memory mechanism is found in salinity. When the ice cover melted in the austral spring and summer, areas outside the polynya would have been covered by a thin layer of relatively fresh water. Because of its buoyancy, the fresh water would have stayed on the surface and not mixed with the water below. In the area where the polynya had been, however, the surface would have been salty, because there would have been no sea ice to melt there.

The following winter, when the cold atmosphere started to remove heat from the surface water, the layered structure of the waters in the region of the polynya would have become



COSMONAUT POLYNYA, in the Cosmonaut Sea at 45 degrees east longitude and 65 degrees south latitude, marks a region where polynyas often appear. The fact that certain geographic locations are the sites of frequent polynyas and areas of poor ice cover lends support to the theory that the topography of the ocean bottom can influence the formation of polynyas. This image was made in September, 1986.



POLAR VIEW reveals several open-ocean polynyas (*1–3*) and coastal polynyas (*A–J*). The image shows the average ice cover during a three-day span in the winter of 1980.

unstable: the salty surface water, on becoming colder and therefore denser, would have sunk and warmer water from lower layers would have risen, establishing new convection cells and re-creating the polynya. In other areas the relatively fresh surface water would simply have frozen, reestablishing the Antarctic ice cover.

The effects of the Weddell Polynya were in keeping with its size. The polynya left a clear imprint on waters far below the surface. For example, measurements of temperature at various depths made in the area of the polynya during the summers of 1973 and 1977 (that is, before and after the occurrence of the polynya) reveal dramatic changes in the temperature of the deep water. The changes seem particularly dramatic when one bears in mind how very stable the deep waters of the ocean are over short time scales. The temperatures measured in the region in 1977 were colder than those measured in 1973 by as much as .8 degree, in depths all the way down to 2,500 meters.

The missing heat had probably been carried to the surface by convection in the polynya. Reasonable estimates suggest that the rate of convective overturning necessary to transfer that much heat could have been as high as six million cubic meters per second during the winter, when the polynya was active. In the process the polynya could have produced as much as half of the Antarctic Bottom Water that flowed out of the Weddell Sea.

What caused the Weddell Polynya to form exactly where it did? The most plausible answer holds that the topography of the ocean bottom played a role. The polynya's first appearance was directly above a feature called the Maud Rise, which is an underwater seamount. Here the bottom is some 3,500 meters closer to the surface than it is in other regions. Perhaps the warm Circumpolar Deep Water is deflected toward the surface as it passes over the Maud Rise. This would "precondition" the region so that other events, such as greater than average mixing of the water by windstorms, would be more likely to bring warm, salty water into the surface layer, triggering the formation of convection cells.

Satellite data from the years since 1976 yield some evidence for the hypothesis. In many of those years the microwave measurements reveal relatively small polynyas or areas of very thin and patchy ice cover in the vicinity of the Maud Rise. Further evidence is found in an area called the Cosmonaut Sea near 65 degrees south latitude and 45 degrees east longitude, where the ocean circulation, perhaps in response to the topography of the bottom, also seems to be preconditioned for the formation of convection cells. In the Cosmonaut Sea, as over the Maud Rise, small polynyas and areas of diminished ice cover often occur. These polynyas in general do not survive an entire winter.

To test this hypothesis, one of us (Gordon) helped organize efforts to push through the ice cover and reach the Maud Rise region in winter. Our first attempt was in 1981 on the Soviet ship *Mikhail Somov.* On this trip we were not able to penetrate all the way to the Maud Rise, although we got close enough to be able to gather extremely useful data on the interactions between the deep water and the surface layer below the ice cover. We found an unexpectedly large entrainment of deep water into the winter surface layer, suggesting an active vertical flux of deep water.

We succeeded in reaching the Maud Rise on our next attempt, in the austral winter of 1986, as part of a multinational expedition on the West German vessel *Polarstern* of the Alfred Wegener Institute for Polar and Marine Research in Bremerhaven. Although the analysis of our data from that expedition is incomplete, we found there was indeed an upwelling of deep water over the Maud Rise. The fact that there was an upwelling in 1986, even in the absence of a polynya, indicated that the topography of the ocean bottom was probably deflecting the flow of Circumpolar Deep Water upward. In addition to elucidating mechanisms associated with the formation of the Weddell Polynya, the *Polarstern* voyage represents a relatively new and exciting way to study the ocean: combining large-scale satellite observations of the surface with local, ship-based research.

The study of polynyas is leading oceanographers to a new understanding of the mechanisms at work in the Southern Ocean. As a result it is bringing about new ways of considering the processes that govern much of the world's climate, and it raises important new questions. For example, does the overturning in large polynyas, such as the Weddell Polynya, alter the rate at which carbon dioxide is exchanged between the ocean and the atmosphere? The deep water that wells up south of 60 degrees south latitude is high in carbon dioxide and releases significant quantities of it to the atmosphere. Because polynyas accelerate the upward flux of Circumpolar Deep Water, one might expect they would enhance the rate of carbon dioxide release, perhaps adding to the projected global greenhouse warming. On the other hand, as we mentioned above, the rate of overturning in polynyas is so great that no single water parcel spends enough time near the surface to exchange much carbon dioxide with the atmosphere. The net effect of polynyas on atmospheric carbon dioxide levels is therefore uncertain. Conversely, it is also not known what effect the projected warming might have on the frequency or extent of future polynyas.

How important are polynyas generally in the deep-ocean overturning that takes place in the Antarctic? We know the Weddell Polynya had quite a pronounced effect, cooling a great volume of ocean and giving surface water access to the atmosphere over a large area. Are such large polynyas rare events, or do they occur relatively often? And how much of the overturning that takes place near the continental shelf is mediated directly by the coastal polynyas? Oceanographers now have the tools to answer these important questions and many others. The combination of comprehensive, long-term observations by satellite and detailed observations from the field in winter is making it possible to construct ever more precise models of these vital and far-reaching processes in the Southern Ocean.

FURTHER READING

WEDDELL DEEP WATER VARIABILITY. Arnold L. Gordon in *Journal of Marine Research,* Vol. 40, Supplement, pages 119-217; 1982.

ANTARCTIC SEA ICE, 1973-1976: SATELLITE PASSIVE-MICROWAVE OBSERVATIONS. H. Jay Zwally, Josefino C. Comiso, Claire L. Parkinson, William J. Campbell, Frank D. Carsey and Per Gloersen. NASA Special Publication 459; 1983.

ANTARCTIC OFFSHORE LEADS AND POLYNYAS AND OCEANOGRAPHIC EFFECTS. H. Jay Zwally, Josefino C. Comiso and Arnold L. Gordon in *Oceanology of the Antarctic Continental Shelf,* edited by Stanley S. Jacobs. American Geophysical Union, 1985.

RECURRING POLYNYAS OVER THE COSMONAUT SEA AND THE MAUD RISE. J. C. Comiso and A. L. Gordon in *Journal of Geophysical Research,* Vol. 92, No. C3, pages 2819-2833; March 15, 1987.

The Neurobiology of Feeding in Leeches

A single neurotransmitter, serotonin, orchestrates feeding behavior in the medicinal leech. The discovery may illuminate how neurochemicals control behavior in other animals

by Charles M. Lent and Michael H. Dickinson

In *The African Queen* Humphrey Bogart, finding leeches clinging to his body, expressed a popular sentiment when he exclaimed, "If there's anything in the world I hate, it's leeches—the filthy little devils!" Yet to a neurobiologist the bloodsucking worm is a thing of beauty. Its nervous system is simple and highly organized, and its neurons are large, readily identifiable and accessible to microelectrodes. These features make the leech a particularly useful animal in which to study the activity of specific neurons. Moreover—with a certain poetic justice—the animal's repugnant feeding habits have turned out to provide the vital clues enabling our laboratory to discover the function of an important group of neurons.

The leech's scientific name, *Hirudo medicinalis,* reflects the animal's medical utility in the once widespread practice of bloodletting, or phlebotomy. Leeching is mentioned in the works of the second-century Greek physician Galen and its methods were detailed by the medieval Islamic philosopher Avicenna. By the Middle Ages it had developed into such a common treatment for most maladies that the word "leech" assumed multiple meanings: any medical procedure, poultice, drug—even the physician himself.

The popularity of leeching peaked in the late 18th and early 19th centuries. Some 30 to 40 million leeches were imported each year into France alone. The demand for *Hirudo* was so great that collectors fished them to near-extinction throughout western Europe, and the leech population in that region remains depleted to this day. In the late 19th and early 20th centuries leeching declined markedly and became confined primarily to removing blood from black eyes.

Many folk remedies are now being rediscovered and, given the protracted history of leeching, it is not entirely surprising to learn of its renaissance in modern medicine. In Europe and America today leeches are postoperative adjuncts to plastic surgery and the surgical reattachment of amputated extremities. Leeches increase the success of these operations by removing excess blood and thereby preventing the tissues from dying while the capillaries regrow. In addition leeches are the source of a powerful anticoagulant, hirudin, the gene for which has now been cloned, and their saliva may suppress the spread of tumors.

In spite of the long association between medical science and *Hirudo,* little was known until recently of the natural feeding behavior of leeches. We therefore undertook to examine this behavior in detail, and in the process we uncovered crucial hints to the functional role of a specific group of neurons in the nervous system of the leech.

Leeches are segmented worms (annelids) that evolved from earthworms. They have a sucker at each end: the one at the head houses the mouth, and the larger one at the tail is employed in crawling. In our studies we found that hungry leeches rest at the edge of a pond, at the water surface. They are alerted by wave ripples such as those made by mammals—the primary source of blood meals. An alerted leech orients its head toward the wave source and initiates undulatory swimming. Leeches do not always hit their target, but their aim is alarmingly accurate: most swim within 25 degrees of a wave source.

The leech stops swimming when it makes contact with an object, which it then explores, crawling with its suckers much like an inchworm. On finding a warm region, the leech bites. (We confirmed the leech's attraction to warmth by noticing that the animals clustered under a beaker of warm water placed on an aquarium lid.) If a bite fails to draw blood, the leech explores further and bites another warm region. To study biting behavior we put leeches on a sheet of paraffin lying on a warm surface and then count the number of bites in the wax. The animal's three serrated jaws create distinctive marks reminiscent of the Mercedes-Benz emblem. We found biting occurs most frequently when the paraffin is at between 35 and 40 degrees Celsius—the range of mammalian body temperatures.

The feeding apparatus of *Hirudo* is composed of various muscles and glands. Rhythmic contractions of the jaw muscles move the three semicircular, toothed jaws back and forth, cutting the skin of the host. Saliva is secreted from ducts that alternate with the teeth. This arrangement injects saliva, with its lubricating mucus and clot-inhibiting hirudin, deep within incisions. As blood flows into the leech's buccal cavity, it is pumped into the crop by rhythmic contractions of

CHARLES M. LENT is professor of biology at Utah State University. He got a doctorate in marine biology from the University of Delaware in 1967. His interest in the neuronal basis of behavior led to postdoctoral work on leeches with Gunther S. Stent at the University of California, Berkeley. He has held faculty positions at Beloit College, the State University of New York at Stony Brook and Brown University. MICHAEL H. DICKINSON was an undergraduate in Lent's laboratory at Brown, where he made major contributions to the research reported in this article. He is now in a doctoral program in neurophysiology at the University of Washington.

the pharynx, in synchrony with the biting movements.

In the laboratory we feed leeches on a mixture of red blood cells and tissue-culture fluid. The warm mixture is placed in a test tube, which is sealed with a paraffin sheet through which the leeches bite. They feed readily and ingest blood for about 30 minutes, filling the crop and its diverticula. Ingestion is stereotyped and compulsive: strong stimuli, such as forceful pulling or even cutting into the body, do not deter this tenacious worm once it begins to feed. Massive amounts of blood are ingested—from seven to nine times the weight of the leech itself, among the largest meals of any animal. Not surprisingly, the leech feeds rarely. In the laboratory satiation often lasts for a full year.

After it has fed, the leech is distended and barely able to move. Satiated leeches behave quite differently from hungry ones. They seek deeper water, tend to crawl into crevices of rocks and logs and do not exhibit the spontaneous swimming of hungry leeches. Rather than biting a warm surface, satiated leeches avoid it, lifting their heads and crawling away.

What could account for this dramatic change in behavior? We wondered whether distension of the body wall might be the stimulus to terminate ingestion. When we distended leeches as they fed by injecting saline solution into their crops, they always stopped feeding immediately. We found next that if we cut through the body wall of satiated leeches and emptied their crops, the animals would resume biting if they were placed on a warm surface. Lastly we found that if we cut into feeding leeches, allowing the blood to escape, their ingestive behavior was prolonged for hours. The leech's behavioral state, then, alternates between hunger and satiation, and the switch between the two activities is distension.

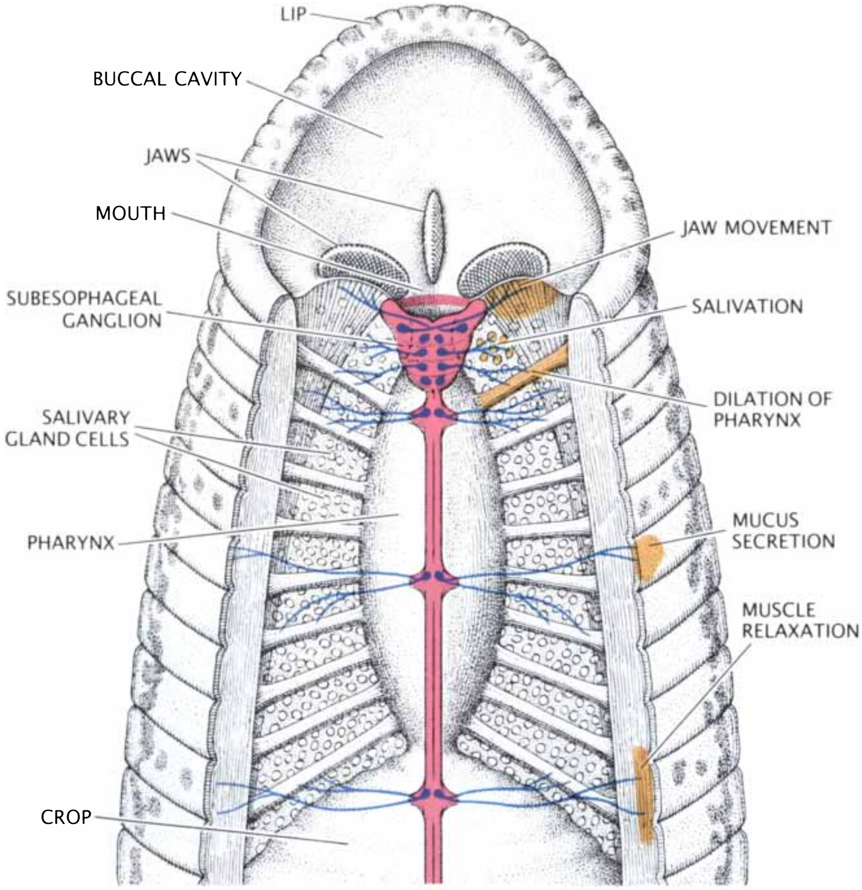
Having uncovered the principal components of the leech's feeding habits, we were ready to delve into the neuronal basis of the behavior. In leeches and other segmented worms each segment contains one ganglion, or cluster of nerve cells. The leech's central nervous system consists of 32 ganglia, each communicating with the rest of the body by way of two pairs of lateral roots [*see bottom illustration on page 101*]. The 21 ganglia of the midbody are linked along the longitudinal connectives like beads on a string. Of the remaining ganglia, four are fused in the head to form the subesophageal ganglion and seven are fused in the tail to form the caudal ganglion [see "The Nervous System of the Leech," by John G. Nicholls and David Van Essen; SCIENTIFIC AMERICAN, January, 1974].

Each ganglion includes some 400 neurons. In leeches the sensory and motor neurons send their axons—long cell extensions that carry nerve impulses—into the lateral roots, where the signals are relayed to sensory detectors and to effector cells: muscles and glands. The leech's interneurons (neurons that communicate exclusively with other neurons) have axons extending only to the connectives.

The nerve cells of all animals, including leeches, share fundamental physiological properties. Neuronal membranes are electrically polarized: negatively charged ions within the cells cause the electric potential of the cell interior to be between −50 and −80 millivolts with respect to the outside. If positive ions flow into the cell, its membrane is depolarized; when the inside-outside voltage difference decreases to a critical value, called the threshold, the cell fires an electric impulse, or action potential. Action potentials are rapidly conducted along axons. An impulse of a motor neuron, for example, travels from the ganglion to a muscle and makes it contract, whereas impulses of sensory neurons travel inward from the periphery.

Neurons communicate with one an-

LEECHES cling to Humphrey Bogart's torso in a memorable scene from *The African Queen.* Katharine Hepburn applies salt to them, a procedure that causes them to regurgitate and release their grip. Bogie mentions that this will prevent them from leaving their heads embedded in his skin—something that leeches in fact do not do.



VENTRAL VIEW of a dissected leech head shows the animal's nervous system and feeding organs. Of the nerve cells only the large serotonin neurons are depicted (*blue*). The subesophageal ganglion contains two large lateral cells near the edges and four pairs of Retzius cells near the center. The other ganglia contain two Retzius cells each and are linked by connectives along the length of the body. Neurons in each ganglion extend axons to a pair of lateral roots leading to peripheral glands and muscles. Impulses from the Retzius cells and large lateral cells release serotonin at peripheral organs (*yellow*), evoking jaw movement, salivation, contractions of pharyngeal muscles, mucus secretions and softening of the muscles in the body wall.

other at synapses, buttonlike structures at the terminals of axons. There are two types of synapse, electrical and chemical. Electrical synapses connect the cytoplasm of neurons, allowing ions to flow directly between them. This coupling brings the cells toward the same electric potential and hence to similar levels of excitability. Such synapses conduct in either direction between cells, and in the leech they play a key role in synchronizing the activity of a specific group of neurons.

Chemical synapses, on the other hand, conduct information by an extracellular route and only in one direction. An impulse triggers the release of a chemical from the terminals of the presynaptic cell (the one in which the impulse originated). This "neurotransmitter" diffuses to the postsynaptic cell, binds to its receptors and alters its membrane potential.

Chemical synapses can be excitatory, depolarizing the postsynaptic cell so that it is more likely to fire an impulse, or inhibitory, decreasing the likelihood. Individual neurons can excite some cells and inhibit others, depending on the properties of the postsynaptic receptors. The various excitatory and inhibitory synaptic inputs to a neuron are integrated to establish its moment-to-moment membrane potential; that is, the inputs together set the level of the cell's excitability. Elegant studies have shown that in the leech, synaptic activity between individual neurons produces some simple, stereotyped behaviors: reflexive muscle contraction, oscillatory swimming and rhythmic beating of the heart.

The array of neurotransmitters in leech neurons is similar to that found in mammalian brains. It includes acetylcholine, gamma-aminobutyric acid (GABA), certain peptides and three monoamines: octopamine, dopamine and serotonin. We became intrigued by the possible functions of serotonin when it was discovered some years ago that the largest leech neurons, the Retzius cells, contain an abundance of the chemical.

These cells, first described in 1891 by the Swedish anatomist Gustaf Retzius, were the first to be identified from animal to animal. Each ganglion houses a pair of Retzius cells whose axons enter the lateral roots, branch repeatedly and project widely among peripheral organs of the body. The cells have the enzymes to synthesize serotonin, and their membranes actively accumulate it. Of the remaining neurons that project axons to the periphery, only one other pair contain serotonin: the large lateral cells in the first segment of the nervous system.

In addition, leech ganglia incorporate four classes of interneurons that have high concentrations of serotonin. Two classes are paired and positioned on the lateral edges of most ganglia. The other two are represented unequally in ganglia along the length of the leech. Posteromedial interneurons usually occur in pairs in frontal ganglia, singly in the midbody and are often absent altogether from posterior ganglia. In addition the first seven ganglia have extra pairs of anteromedial interneurons. As a result there are about 10 serotonin neurons per ganglion toward the front of the animal, and the number decreases to about five cells toward the rear.

We recently employed a sensitive technique, high-pressure liquid chromatography, to learn whether these differences in cell number correspond to serotonin levels within ganglia. It turns out that they do. Indeed, the anterior ganglia contain about five times as much serotonin as the posterior ganglia. Hindsight allows us to suggest these differences point to there being some function for serotonin and serotonin neurons that is expressed by the head—such as feeding.

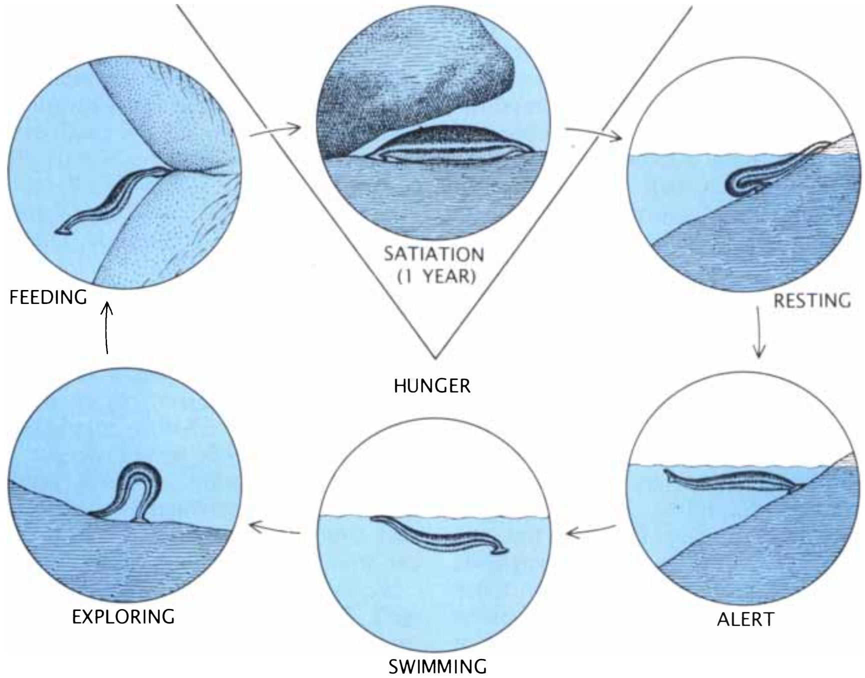
For years, however, the function of Retzius cells remained mysterious in spite of intensive study. Yet because the size of biological structures often reflects their functional importance, it was difficult to believe cells so large would not have a major role. Then about 15 years ago one of us (Lent) uncovered a seemingly trivial role for the cells in the secretion of mucus. Working with dissected preparations consisting of a segmental ganglion connected by its lateral roots to a section of midbody wall, we applied depolarizing currents to Retzius-cell membranes, increasing the frequency of the cells' impulses. The impulses were conducted to the body, where they evoked mucus secretion from skin glands. We then removed the ganglion and found we could still induce secretions by bathing the body wall in serotonin solutions.

In 1978 Adrian J. R. Mason and Lucy D. Leake of Portsmouth Polytechnic in England showed that serotonin from Retzius cells also softens body-wall muscles and thereby increases the ability of the body wall to distend. Two years later William B. Kristan, Jr., and Michael P. Nusbaum of the University of California at San Diego found that serotonin from the lateral interneurons is involved in generating the undulating muscular contractions that characterize swimming in leeches. Then about five years ago Cameron G. Marshall in our laboratory established that serotonin is the only neurotransmitter to reliably stimulate leech salivary gland cells to secrete saliva.

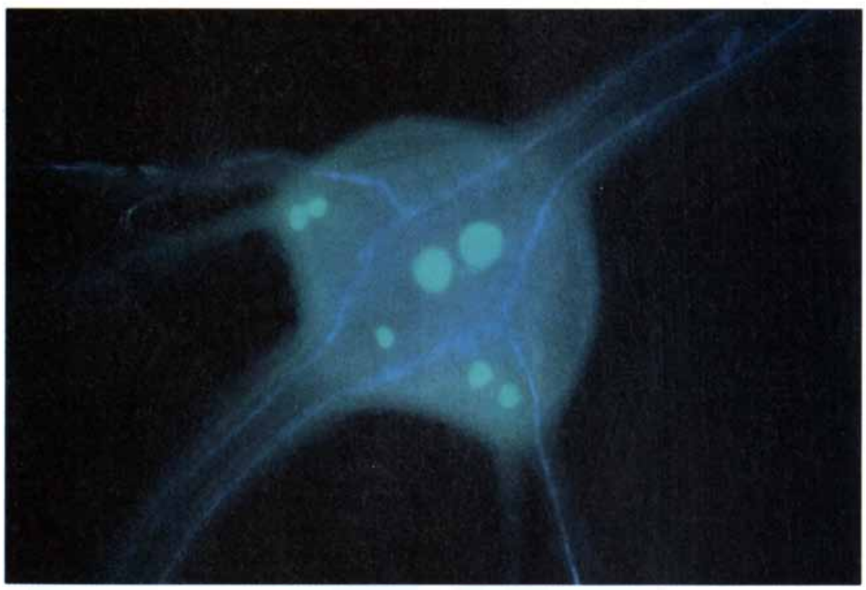
Leech serotonin neurons have still another salient characteristic. When Bryan M. Frazer and one of us (Lent) were at the State University of New York at Stony Brook, we discovered all serotonin neurons are intercoupled by electrical synapses and many of them share synaptic inputs from a common source. The electrical intercoupling maintains the cells at similar levels of excitability, and the shared synapses excite or inhibit them simultaneously. Both kinds of connections tend to synchronize the firing of impulses by these chemically related neurons.

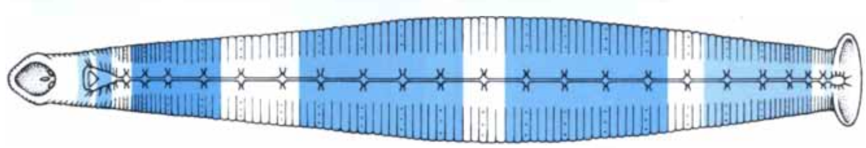
One would expect such synchronized neurons to carry out a coordinated function, and yet these serotonin neurons exhibit a curious diversity of roles. In particular, the links between serotonin, salivation and swimming led us to ask the simple question: Why would a leech swim and salivate? Both activities are associated with feeding (leeches swim to their prey and salivate while ingesting), and so we hypothesized that serotonin is implicated in feeding. This led us to devise a series of experiments to test the role of serotonin in leech feeding behavior and physiology. It was the turning point in our work.

In a first set of experiments we looked for effects of serotonin on the behavior of whole animals. If leeches are bathed in serotonin, the behaviors associated with hunger are greatly enhanced: hungry leeches begin swimming into wave ripples in half the time it normally takes them, and their biting frequency increases by



FEEDING CYCLE of the medicinal leech alternates between hunger and satiation. The hunger phase consists of two subphases, appetite and ingestion. During the appetitive phase the leech swims toward wave sources. Once it finds a warm-blooded host, it will feed for half an hour, consuming up to nine times its weight in blood. Satiated leeches do not bite; they avoid warm surfaces and hide under rocks or in crevices.





SEROTONIN causes neurons in a midbody ganglion of the leech to glow brightly in a fluorescence photomicrograph (*top*). Two large Retzius cells are seen at the center. Two pairs of lateral interneurons appear near the edges. The small cell below the Retzius cells is a posteromedial interneuron. Frontal ganglia contain two such interneurons and an extra pair of anteromedial interneurons, but caudal ganglia may have only the Retzius cells and the two lateral interneurons. Such differences in the number of serotonin neurons account for the higher serotonin levels in anterior and midbody segments, as is indicated by varying shades of color (*bottom*). This is consistent with serotonin's role in feeding, a behavior dominated by the head.

two-thirds. The serotonin bath also augments the volume of ingested blood by one-third, raising the size of the meal to more than 10 times the animal's body weight! Most important, the serotonin evokes biting by satiated leeches, which normally do not bite. In other words, serotonin can change the leech's feeding behavior qualitatively as well as quantitatively.

We next investigated the effects of serotonin on the organs directly involved in feeding behavior. When we flooded partially dissected leech heads with serotonin, we observed biting movements of the jaws, a secretion of saliva and rhythmic contractions of the pharynx. These three responses continued to be evoked by serotonin after we removed the ganglia and connectives, which suggests the transmitter acts directly on these organs. In addition we found warming the lips of the leech heads by from three to five degrees C. evoked the same responses, but only if the nervous system remained intact.

Before we could ascribe these functions to serotonin neurons, we had to demonstrate that impulses of such neurons are both sufficient to evoke the physiological responses of feeding and, indeed, necessary for these responses. To test the first requirement we prepared leech heads so that the anterior Retzius cells and large lateral cells could be impaled with microelectrodes. We observed that low-frequency stimulation of the large lateral cells produces pharyngeal contractions and that higher frequencies evoke rhythmic pumping. Stimulation of Retzius cells increases salivation and at higher frequencies produces twitches in the jaw muscles. Electrically stimulating the serotonin effector neurons to fire, then, evokes the same feeding-related responses as exposure to serotonin itself.

To demonstrate that the cells are in fact necessary for feeding responses, we needed a way to remove the effector neurons from the behavioral circuit. We chose to inject leeches with the neurotoxin 5,7-dihydroxytryptamine (5,7-DHT). The toxin is accumulated actively inside serotonin neurons, where it is oxidized in a reaction that produces destructive free radicals and brown phenolic compounds.

We injected hungry, biting leeches with 5,7-DHT. When we examined them a few days later, we found they could swim and crawl normally but would not bite warm surfaces. In fact, they acted as if they were satiated and drew their heads rapidly away. We found their Retzius cells and large lateral cells were brown and irregularly shaped. Moreover, the serotonin levels in the brown Retzius cells were depleted by more than 90 percent.

To examine the precise effects of the toxin on individual cells, we inserted microelectrodes into specific sensory and motor neurons. These cells displayed normal impulses and synaptic potentials. We thought the toxin had destroyed the serotonin neurons, and so when we turned to the Retzius cells and large lateral cells, we were astonished to find them firing action potentials. On closer examination we found their membrane receptors, synaptic inputs and peripheral axons also appeared normal.

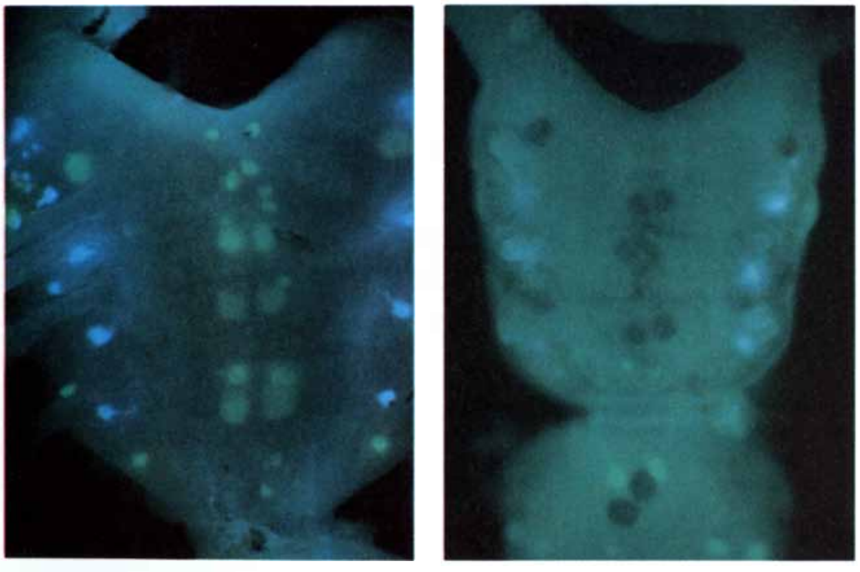
Apparently 5,7-DHT does not kill these neurons but merely depletes their serotonin: the toxin induces a precise neurochemical lesion. We reasoned that if the toxin halts feeding behavior by depleting serotonin, replacement of the neurotransmitter ought to reverse the toxin's behavioral effect. And indeed, when we bathed the toxin-treated leeches in serotonin, they resumed biting when presented with a warm surface.

We next examined preparations dissected from toxin-treated animals. We found that neither warming the lips nor electrically stimulating the Retzius cells and large lateral cells at high frequency would evoke jaw movement, salivation or pumping. When the preparations were flooded with serotonin, all three physiological responses resumed. We can conclude that these large serotonin effector neurons are necessary for the expression of leech feeding behavior.

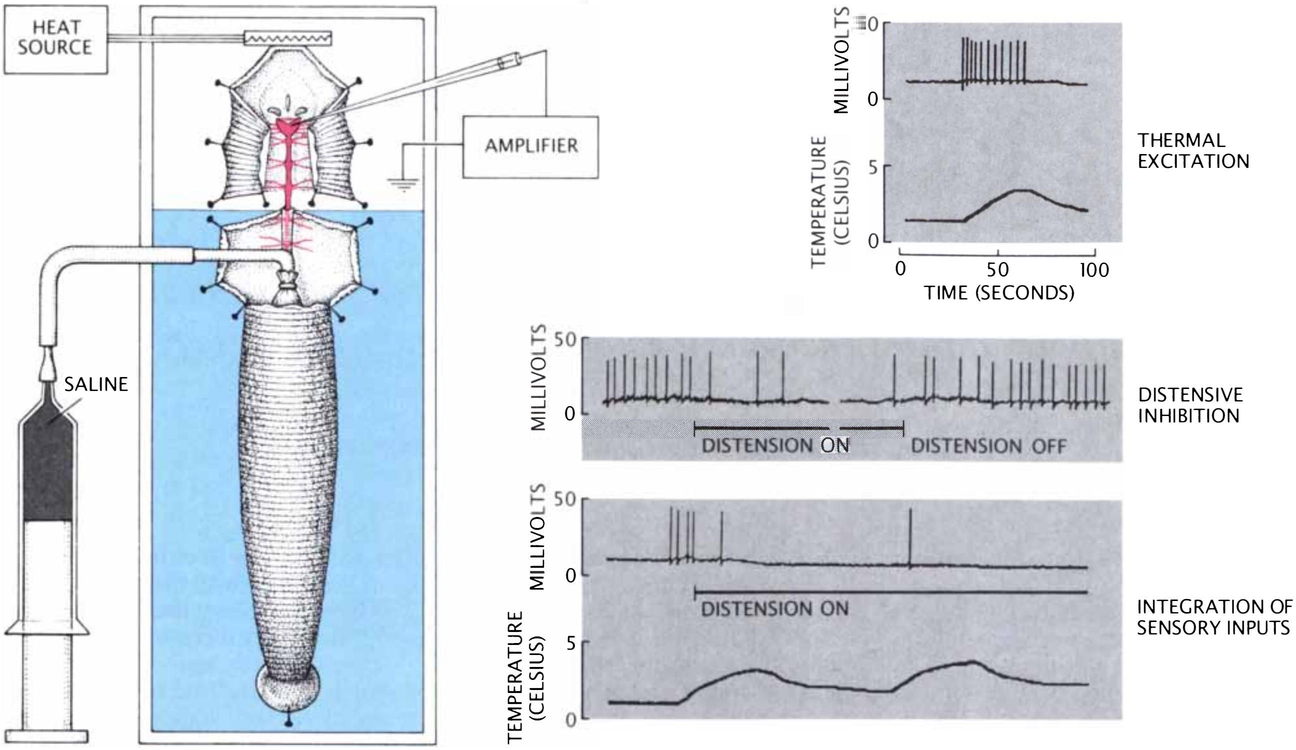
These experiments establish the pivotal role of neuronal serotonin in generating feeding behavior. It follows logically that stimuli, such as warmth, that initiate feeding ought to excite serotonin neurons to fire impulses. Conversely, stimuli that terminate feeding, such as distension of the body wall, ought to inhibit the cells' activity or interfere with their release of serotonin.

We studied the effects of warmth and body-wall distension on serotonin neurons in partially intact preparations. We found that when the lip is warmed, the Retzius cells and large lateral cells are synaptically excited and respond with high-frequency impulses. Moreover, the response to warmth is specific to the lip, not to other parts of the skin, and generally the stimulus affects only the serotonin neurons. The firing rate of the cells usually increases as the temperature at the lip rises. A stimulus that evokes feeding in leeches, then, also excites their serotonin neurons to fire impulses at high frequency.

To test the effects of distension we filled the crops of these preparations with saline solution. The Retzius cells and large lateral cells were inhibited for as long as the crops remained distended. Impulse activity resumed



SUBESOPHAGEAL GANGLION has four pairs of Retzius cells at the center and a pair of large lateral cells toward the upper sides, which appear green in a fluorescence micrograph (*left*). Neurons containing the neurotransmitter dopamine appear blue. When treated with the toxin 5,7-dihydroxytryptamine, the serotonin neurons turn brown and appear deformed (*right*): the toxin depletes serotonin. It does not kill the cells, however. Toxin-treated leeches crawl and swim normally but will not bite.



WARMTH applied to the lip of a partially dissected leech stimulates the large lateral cells to fire rapidly (*top trace*). The effect of distension was tested by filling the crop with saline and then emptying it. Distension inhibits the large lateral cells. When distension is relieved, the cells resume firing (*middle trace*). When warmth and distension are applied together, the synaptic inputs are integrated: the inhibitory effect of distension dominates the excitatory effect of warmth (*bottom trace*).

when the saline was removed from the crop. We concluded that the stimulus (distension) that terminates leech feeding inhibits the cells from firing impulses and therefore effectively blocks the release of serotonin from the axon terminals. Furthermore, the cells integrate their synaptic inputs so that the inhibitory effect of distension overrides the excitatory effect of warmth. Hence the leech will not feed when its body is distended.

All known serotonin functions in the leech, then, fit into a scheme whereby the chemical orchestrates feeding behavior, from finding prey to ingesting a blood meal. We have also found that behavior affects serotonin levels in the leech nervous system. Hungry leeches have more serotonin in their ganglia than satiated leeches. Ingestion is rapidly followed by a drop of from 25 to 30 percent in the serotonin content of anterior ganglia, and the levels remain low as long as the body wall is distended. When distension is relieved, the levels of serotonin begin to increase and the leech resumes its feeding behavior.

We find it intriguing that appetitive behaviors, such as swimming to the prey, do not deplete serotonin levels in those interneurons that are involved in these behaviors. Kent T. Keyser of Stony Brook and Joyce Ono of the City of Hope, together with one of us (Lent), measured serotonin in the lateral interneurons at concentrations exceeding 100 millimolar, among the highest transmitter levels measured in any neuron. The high levels may ensure that serotonin is not depleted during the hungry leech's repeated attempts to find a meal.

The Retzius cells and large lateral cells, in contrast to interneurons, are connected mainly with ingestion: pharyngeal pumping, biting, salivation, body-wall distensibility and mucus secretion (which turns out to help the animal cling to prey). The act of ingestion depletes ganglionic serotonin as the leech's behavior changes from hunger to satiation. We suspect prolonged excitation of these cells during ingestion activates a transport of serotonin along axons, thereby depleting the serotonin in the cells.

A single neurotransmitter expresses feeding behavior in leeches by exciting the specific nerve cells and organs that together produce a complex behavior at the organismal level. Only a few neurons are involved—a remarkable example of neuronal economy that enables these animals, which have only 400 neurons per ganglion, to hunt and prey on mammals. Remarkably, serotonin effector neurons also control ingestive muscles and salivary glands in distantly related leech species that feed with a proboscis rather than jaws. Indeed, the behavioral function of serotonin may transcend leeches: serotonin evokes salivation, biting and pharyngeal contractions in animals of three other phyla: nematodes, mollusks and insects. Hence the discovery of a clear behavioral role for these chemically related neurons in the leech may help to illuminate patterns in the evolution of behavioral neurochemicals.

FURTHER READING

FEEDING BEHAVIOR OF THE MEDICINAL LEECH, *HIRUDO MEDICINALIS* L. Michael H. Dickinson and Charles M. Lent in *Journal of Comparative Physiology A*, Vol. 154, No. 4, pages 449-455; 1984.

SEROTONIN INTEGRATES THE FEEDING BEHAVIOR OF THE MEDICINAL LEECH. Charles M. Lent and Michael H. Dickinson in *Journal of Comparative Physiology A*, Vol. 154, No. 4, pages 457-471; 1984.

LEECH BIOLOGY AND BEHAVIOR. Roy T. Sawyer. Oxford University Press, 1986.

ON THE TERMINATION OF INGESTIVE BEHAVIOR BY THE MEDICINAL LEECH. C. M. Lent and M. H. Dickinson in *The Journal of Experimental Biology*, Vol. 131, No. 1, pages 1-15; September, 1987.

Early Iron Smelting in Central Africa

More than 2,500 years ago the people near Lake Victoria began smelting iron in tall furnaces that produced a remarkable heat. The authors unravel the workings of this ancient technology

by Francis Van Noten and Jan Raymaekers

In the early 1950's people of the Bahunde tribe in southern Zaire made a curious find while digging for sand not far from their village. What they found was a group of clay objects resembling rough bricks, many of them decorated with circular or linear impressions. Two investigators from the Institute for Scientific Research in Central Africa, Jean Hiernaux and Emma Maquet, were called in to examine the serendipitously uncovered material. Their excavation showed that the bricks had formed part of a tall furnace for the smelting of iron. When the Bahunde people were confronted with that interpretation, they could see no connection between such furnaces and their own culture. They were inclined to believe the furnaces had been built by people with a more sophisticated iron-smelting technique than their own.

The Bahunde were right. Discoveries in the past decade indicate that the tall furnaces represent a remarkable Early Iron Age technology that was widespread in the interlacustrine region of central Africa (which includes eastern Zaire, Rwanda, Burundi, northern Tanzania, southwestern Kenya and Uganda). Employing only simple materials, the Iron Age smelters were able to generate temperatures as high as 1,400 degrees Celsius, providing an efficient means of obtaining iron from its ore. The smelting technology was not only efficient but also long-lived: it appeared in the interlacustrine region sometime during the first millennium B.C. and persisted in many areas until the beginning of the 20th century. Although its origins remain obscure, systematic excavations combined with ethnographic work have begun to yield a detailed picture of how the tall furnaces actually worked.

FRANCIS VAN NOTEN and JAN RAYMAEKERS are keenly interested in subjects that combine archaeology and ethnography. Van Noten is head of the department of general archaeology and anthropology at the Royal Museums for Art and History in Brussels. He has done archaeological fieldwork in both Africa and Belgium since 1965. Raymaekers is a research associate at the Royal Museum for Central Africa at Tervuren in Belgium.

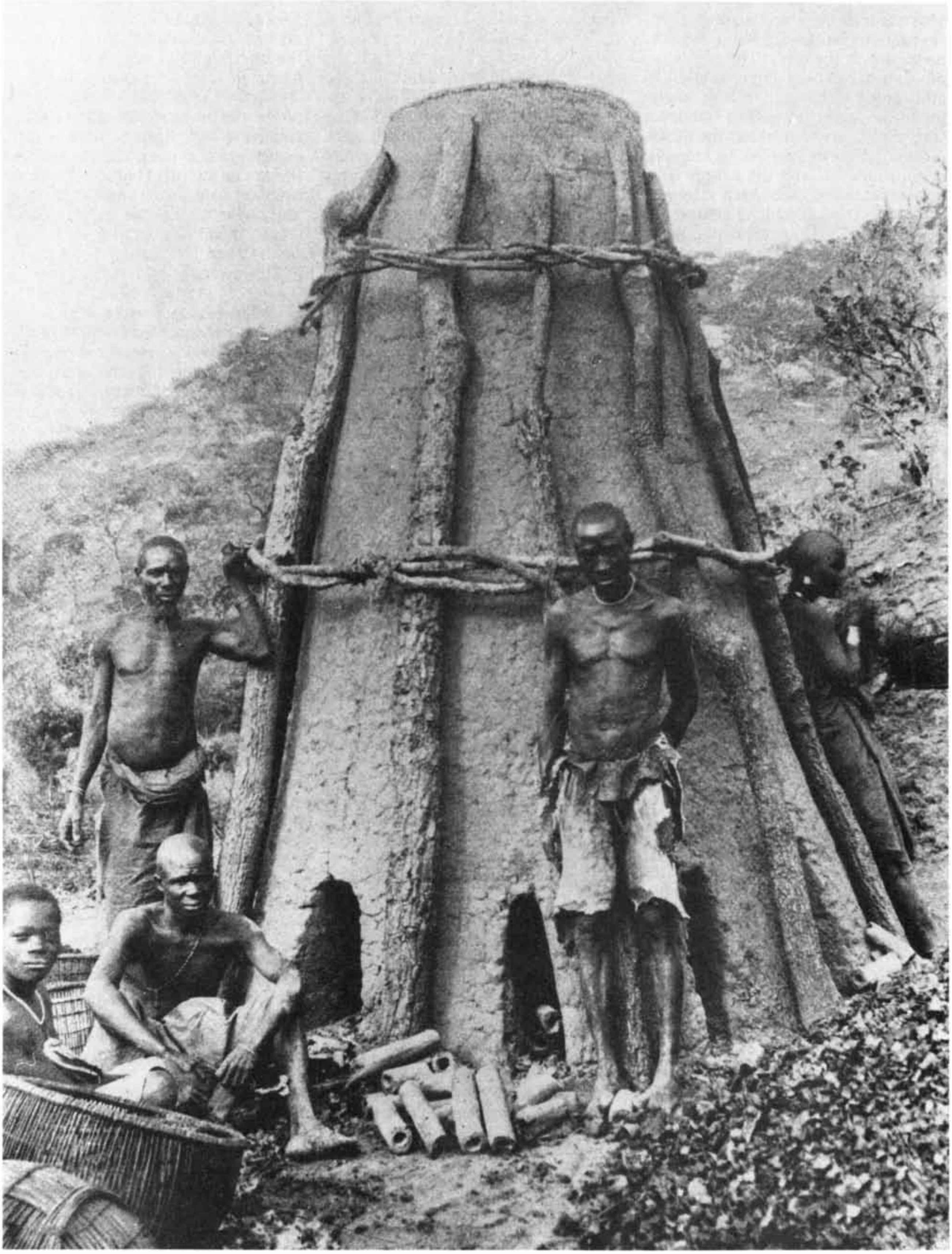
The people of the interlacustrine region (which lies between Lake Victoria and a chain of smaller lakes) are descendants of immigrants who spoke languages of the Bantu family. The Bantu-speaking migrants came originally from an area in what is now Cameroon called the Grassfields. Beginning thousands of years ago they migrated southward, ultimately occupying an area that stretches from southern Africa all the way to Cameroon and Kenya. Within that area some 650 languages of the Bantu family are spoken. Our work has been concentrated on the interlacustrine region, where recent studies indicate a surprising uniformity of Early Iron Age technology.

That uniformity is both spatial and temporal. Spatially, the recent discoveries show that the Early Iron Age was quite homogeneous throughout the interlacustrine region. What is more, the timing of the appearance of Iron Age technology is quite similar in the areas studied so far. At the time of Hiernaux and Maquet's initial excavation, radiocarbon-dating techniques had not been widely applied, and so the first finds could not be reliably placed in time. As radiocarbon methods were more widely employed, however, a range of sites in Rwanda, Burundi, Uganda, the Sudan, Kenya, Zaire and Tanzania were dated, and it became clear that in all these areas Early Iron Age technology had appeared by about the seventh century B.C.

The Early Iron Age culture in the region is characterized not only by the iron-making techology but also by a specific type of pottery that is almost always found in conjunction with it. One of us (Van Noten) has given the term Early Iron Age Industrial Complex to the combination of pottery and iron-smelting methods. After its appearance in the first millennium B.C. the industrial complex persisted for more than 1,000 years. It reached a peak in the first half of the first millennium A.D. before being replaced in the eighth century by the richer, more highly stratified societies of the Late Iron Age.

Now, the industrial complex, which represents the beginning of Iron Age technology in the interlacustrine region, has created problems for prehistorians. In the 1960's it was widely held that the dispersal of ironmaking methods was associated with the migration of Bantu-speaking people from the Grassfields in Cameroon. Yet the dates from the seventh century B.C. heralding the appearance of Iron Age technology in the interlacustrine region are among the earliest in sub-Saharan Africa; they are certainly just as early as the dates from the Grassfields. It is currently thought that the Bantu-speaking people began spreading out of the Grassfields long before Iron Age methods appeared there. Iron Age technology may have been acquired on the way south, or it may have arisen independently in the interlacustrine region (although this seems less likely).

Even if we cannot fully explain its origins, we can provide a sense of the remarkable technology itself. That



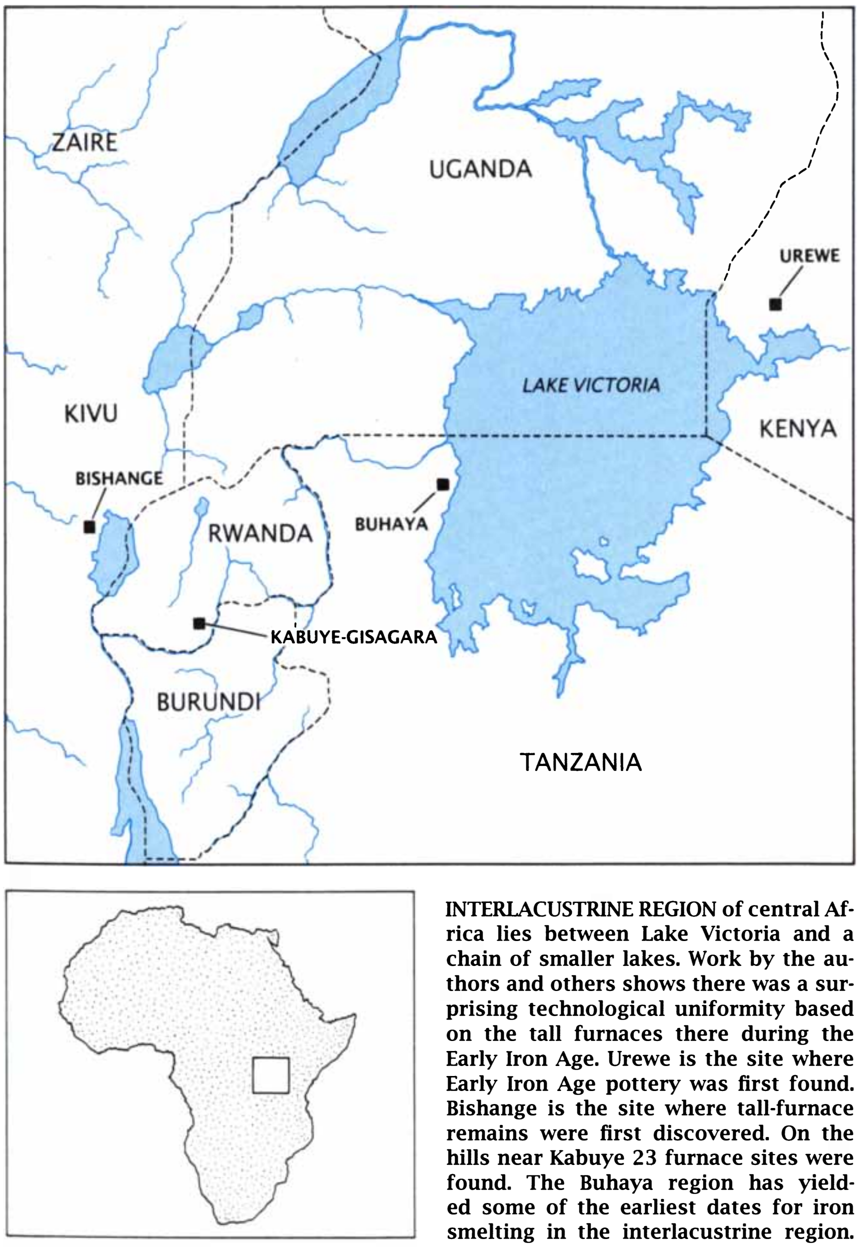
TALL FURNACE for smelting iron was built by the Fipa tribe of Tanzania. The shaft is made of plaster-covered clay; it sits over a large pit. The scaffolding is for loading charcoal and iron ore. The clay pipes around the base are called tuyeres; they convey oxygen to the fire to raise the temperature. Such furnaces came into use in central Africa in the seventh century B.C. and were still in use in some areas in 1914, when this photograph was made by the Belgian ethnographer R. P. Wyckaert.

sense comes from archaeological investigations on the one hand and ethnography on the other. The ethnography in turn has two components: early ethnographic accounts of iron smelting in the region as well as "ethnoreconstruction," which entails the duplication of original Iron Age techniques in collaboration with the people who now inhabit the area. Such ethnoreconstruction is possible in central Africa because Iron Age techniques were practiced within the memory of the oldest living people. In some parts of the region those memories were vivid enough so that during World War II, when imported iron was in short supply, the original Iron Age smelting techniques could be revived.

Much of our work was focused on an area called Kabuye in southern Rwanda, where a remarkably dense concentration of furnaces has been uncovered: 23 sites within a radius of three kilometers. All the furnaces were uncovered accidentally by erosion or construction. When the remains were excavated, the same items generally came to light. Among them were a pit (frequently with a burned outer rim), brick fragments, clay pipes called tuyeres, slag (wastes from smelting), charcoal and in a few cases some iron ore. Occasionally fragments of pottery in a typical Iron Age style were found in the same pit.

The first stage of iron smelting, of course, is to obtain a sufficient quantity of the required raw material: hematite, or iron ore. Hematite contains iron as an oxide (Fe_2O_3) that occurs in crystals or in a reddish earthy form. In the interlacustrine region hematite is found in lumps on certain hillsides, and in the Early Iron Age the gathering of ore was probably limited to surface collecting and digging; there is little evidence for mining. The presence of the ore in Rwanda is attested to even by place-names: the most important town near the Kabuye Hills is called either Butare or Ubutare, meaning "stone" and "ironstone" respectively in Kinyarwanda, the local language.

After the ore was collected it had to be prepared for smelting. One of us (Van Noten) observed the preparation of the ore during an ethnoreconstruction of iron smelting among the Madi people of the northern region of Zaire in the 1970's. There the ore was preheated for 30 minutes in a very hot fire. The Madi gave several reasons for doing so. One was that the firing of the hematite made it easier to separate the iron ore from the other associated materials. A second reason was that heating made the ore brittle and easier to break into small pieces, which can be smelted more effectively. The pieces cannot be too small, however, because they would suffocate the fire in the furnace. Hence the broken ore is sieved to remove the smallest pieces and the dust.



INTERLACUSTRINE REGION of central Africa lies between Lake Victoria and a chain of smaller lakes. Work by the authors and others shows there was a surprising technological uniformity based on the tall furnaces there during the Early Iron Age. Urewe is the site where Early Iron Age pottery was first found. Bishange is the site where tall-furnace remains were first discovered. On the hills near Kabuye 23 furnace sites were found. The Buhaya region has yielded some of the earliest dates for iron smelting in the interlacustrine region.

The pretreatment of the ore was not the only form of preparation required for the smelting. The fuel for the furnace itself was charcoal, which was made from trees selected for the purpose. Microscopic analysis of fragments of charcoal found in the base of Early Iron Age furnaces shows that the smelters chose their trees with care. They preferred trees, such as the thornbush tree (*Zizyphus sp.*), that have solid wood and a high silica content and therefore burn slowly. The charcoal found in the furnaces also shows that the people of the Early Iron Age collected the wood in their immediate surroundings—the grassy savanna—rather than journeying to the thick mountain forests that were some distance away.

The quantity of wood needed in the smelting was considerable. Indeed, the demand for wood may ultimately have been great enough to alter the environment in which the smelters lived. In 1978 an experimental smelting was carried out near the town of Gisagara in the Kabuye region. In a single firing of the small experimental furnace some 95 kilograms of charcoal was consumed. To produce that much charcoal required 15 trees, each with a diameter of from 14 to 16 centimeters.

This example makes it clear that over a period of several centuries the consumption of wood for iron smelting might (along with herding and farming) have led to deforestation. Deforestation would in turn have made smelting impossible, which may explain why no furnaces are found at Kabuye after A.D. 500: perhaps the area was abandoned when the trees ran out.

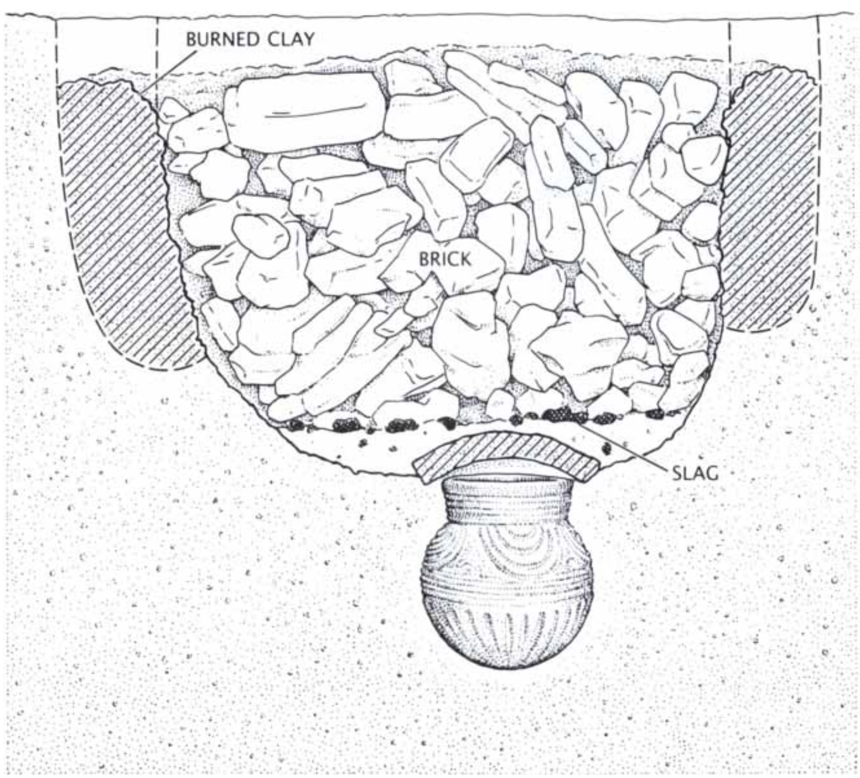
Be that as it may, the means by which the charcoal was actually prepared are well understood, both from historic accounts and from ethnoreconstructions. In Rwanda the traditional method began with the sinking of a pit one to two meters wide and 60 to 80 centimeters deep. Grasses and branches were burned to obtain a layer of warm ash, which was then covered with medium-size sticks. The entire mass was covered with a layer of wet grasses or leaves and allowed to burn slowly. The charcoal, which keeps the original form of the sticks because of the slow and careful burning, is ready in 24 hours.

While the charcoal was being prepared, the furnace was probably being made ready. Two types of furnace are known from the interlacustrine region. The first was simply a bowl in the ground, and it is well known from ethnographic accounts. The second type has a tall shaft erected over the pit to hold the charcoal and ore. Like the low furnaces, these high ones are known from ethnography. Indeed, such accounts indicate that the two types of furnace were employed side by side, with the tall furnace serving for the initial smelting and the low one for refining the product of the initial work. In the archaeological record of Early Iron Age times in the interlacustrine region, however, only the remains of the tall furnaces are found, and it is on those that we shall concentrate here.

In constructing a tall furnace, the initial stage was the digging of the pit. It is not easy to tell precisely how large the pit was, because in many instances soil erosion and human occupation have cut off the top of the bowl. The bowls that have retained their burned rim suggest that the diameter of the pit was in some instances about 1.4 meters and the depth about 75 centimeters. The bottom and sides of the bowl were sometimes treated; in some cases the treatment consisted in smoothing clay over them. At least one of the furnaces in Kabuye (called Kabuye I) was lined with wood poles.

Such accoutrements undoubtedly

REMAINS OF BOWL that lay under the shaft of a tall furnace were discovered at a site called Kabuye IV. The bowl is about 70 centimeters across. (Each division of the arrow is 10 centimeters long.) The fragments within the bowl are the remains of the bricks that made up the shaft and the tuyeres that supplied oxygen to the fire.



CROSS SECTION THROUGH BOWL depicts the site called Kabuye II. The burned clay at the left and right sides of the bowl are the foundations of the walls of the shaft. Slag is the waste from the smelting process. The pot below the bottom of the bowl held material that undoubtedly had magical significance for the smelting process.






had a practical function, but in several areas items have been found at the bottom of the pit that were not likely to have been utilitarian. In Buhaya in northwestern Tanzania the item was simply a piece of smelted iron that had been placed in a small hole at the bottom of the bowl under a sandstone cover. In Bishange in Kivu, in eastern Zaire, Hiernaux and Maquet found potsherds at the center of a bowl. Their hypothesis was that the sherds were the remains of a pot intended to collect molten iron at the bottom of the furnace. As we shall describe, however, the smelting process in the tall furnaces of the Early Iron Age was not compatible with such a function.

The most spectacular examples of these nonutilitarian objects were found in Rwanda. At both the site called Kabuye II and the one called Mutwarubona II a small pot and its lid were buried under the lowest point of the bowl. Both pots were in archaeological terms atypical: they differ from the utilitarian pottery that is commonly found in association with the furnaces. Their anomalous form suggests that they were made specially for the smelting; perhaps they were not made by the potters who made the usual pots. In historic times in the interlacustrine region pottery was made by women, whereas smelting was done by men. It is possible that these special pots were made by the male smelters as part of a ritual preparation for the smelting.

That there were ritual preparations is underscored by something that was observed in the ethnoreconstruction among the Madi of Zaire. Although the Madi had been asked to perform all the smelting operations in view of the observers, in fact the smelters had dug and prepared the pit before the observers arrived. Below the base of the bowl a hole had been dug and filled. When we insisted, the hole was opened, revealing a standing log and neatly cut pieces of various kinds of wood. That assemblage was crowned with a wreath of lianas (a woody vine) and a construction made from three sticks. All of that had been covered with leaves and two more pieces of wood and then buried. Permission to excavate this material was refused after the smelting experiment, strongly suggesting that the buried objects had magical significance.

On a more practical level, after the bowl was complete it may have been filled with stems of papyrus and grasses. The stems probably did not serve as fuel, since slag found in ancient furnaces bears traces of unburned stems and even fresh leaves. The function of the grasses is not well understood; they may have separated iron from slag by supporting the iron as the molten slag dripped through. Such a method of separation applies largely to the tall furnaces; in the low ones built by the Madi as an experiment slag was often allowed to drip into a pit below the bowl and collected there. If the slag was indeed separated from the smelted iron by grasses, then the pottery found under the bowl in the tall furnaces could not have collected molten iron, because the iron remained higher up in the furnace and never ran down.

When the bowl had been dug, it was time to prepare the clay for the construction of the shaft. The clay used in Rwanda during the Early Iron Age contains a considerable quantity of minerals such as quartz, mica and feldspar. Such minerals helped to temper the clay, giving it the toughness needed to withstand the high heat of the furnace. One good source of mineral-rich clay is anthills, and some ethnographers have recorded clay being dug from anthills, kneaded, dried and then pounded flat before being used. After the clay had been prepared in this way, additional materials such as bits and pieces of broken pots were sometimes added for further tempering. Then the building of the shaft could begin.

In order to create the right environment for the smelting process, the walls of the shaft had to be excellent insulators. One reason is that the smelting entailed not only the separation of the slag from the iron but also the chemical reduction of the iron from the oxide form to metallic iron (Fe). Such a reduction requires that the oxygen in the atmosphere be kept to a minimum, and that is one of the functions served by the walls of the tall furnace. It seems clear from excavations in the interlacustrine region that a reducing atmosphere was achieved by a building technique based on a combination of bricks and long, ropelike coils of clay.

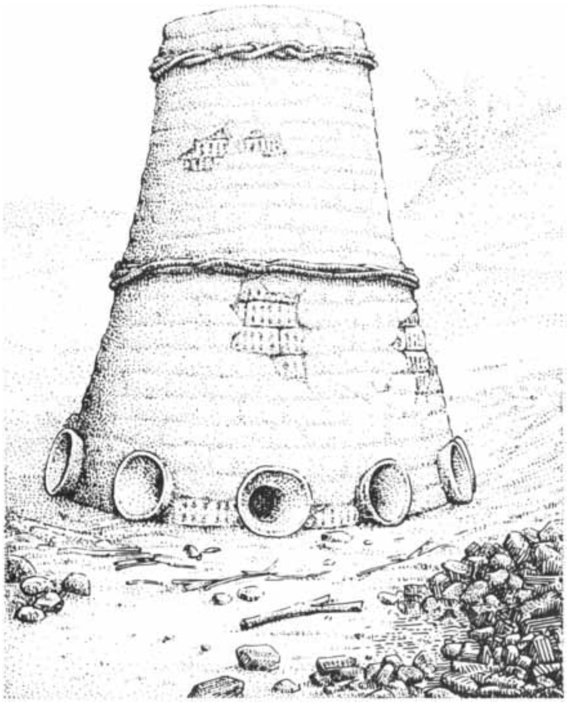
The individual bricks, which were sunbaked, seem to have formed the lower part of the shaft. One reason for such an arrangement was that individual bricks would form a structure that can expand and contract more readily without cracking than one made of coils; this flexibility would be of greatest significance near the base of the shaft, where the temperatures were highest. The bricks, which formed a solid foundation for the upper part of the shaft, do not appear to have been fired beforehand. Instead they seem to have been sun-dried, then put in the wall of the shaft, where they were fired during the smelting process. Paleomagnetic analysis—based on the fact that when clay is subjected to high temperatures, it partially melts and mineral crystals within it align themselves in the prevailing magnetic field—shows that some bricks were fired several times in different positions, suggesting they came out of an old furnace in good shape and were later reused.

The upper part of the wall was erected by means of a technique analogous to "coil-building" methods, which are of ancient origin and are still used for making ceramic vessels. In coil-building of pottery the starting point is a disk of wet clay. Additional clay is then rolled into long, thin snakelike shapes that are coiled around and pressed into place as the walls of the pot rise. When the vessel is complete, both the inside and the outside of the wall are smoothed out and decorated. On a considerably larger scale (the tallest shafts are more than three meters tall), the method of adding coils was employed to construct the tall iron-smelting furnaces.

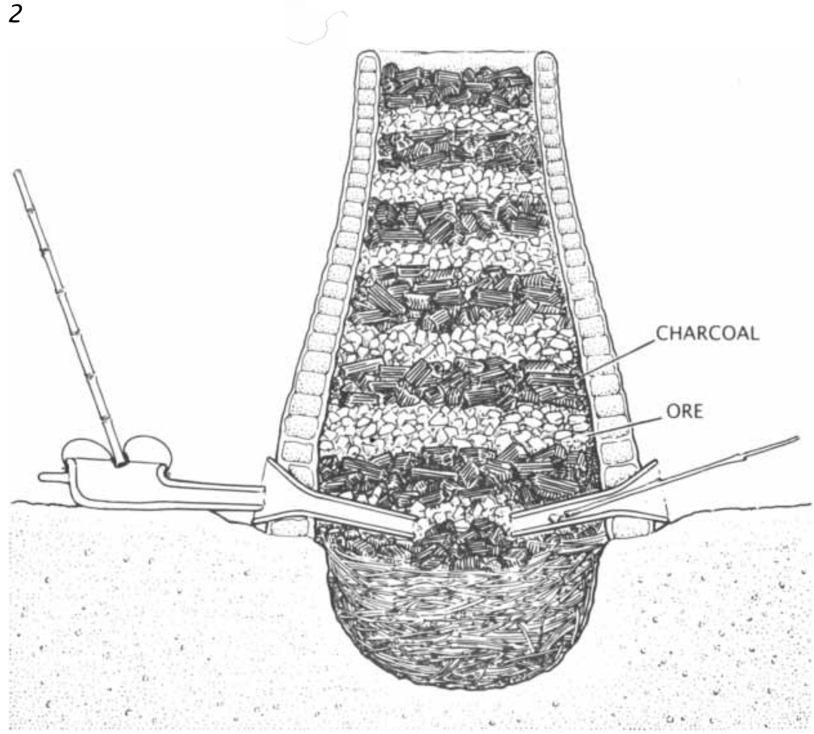
Archaeological analysis has provided much detail about how the coiling was actually done. The superposed layers did not stick together, as they would have if they had been laid down wet. Hence one must conclude that the building up was a gradual process, with one layer being allowed to dry briefly before the next was added. The builders appear to have worked primarily in a counterclockwise direction. Each time a fresh layer was added it was pressed into place with the fingers, leaving a pattern of finger marks in the fresh clay. Some archaeologists have taken that pattern for decoration, but in our view it is simply the result of the technique by which the shaft was constructed. Reinforcing

SIX STAGES OF IRON PROCESSING are shown among the Madi of northern Zaire. The fire is lighted (*top left*). The funnel-shaped pipe is a tuyere. The fire is burning well (*top right*). Slag runs out of the bowl into a pit below (*middle left*). The smelting is finished: the tuyere has been removed and the iron is present as a spongy block in the furnace (*middle right*). Fragments of iron are removed from the furnace (*bottom left*). The iron is reheated and hammered to remove charcoal and dirt (*bottom right*).

1



2



SMELTING IN TALL FURNACES of the Early Iron Age has been reconstructed by the authors. The shaft was plastered and wrapped with vines to hold it together (*1*). Tuyeres were placed radially around the base. The shaft was loaded with layers of

our view is the fact that when the shaft was complete, it was plastered inside and out, which would obscure any "decoration" that had been pressed into the coils.

Plastering strengthened the wall and improved its insulating qualities, thereby contributing to the reducing atmosphere within. What is more, replastering would have made it possible to prepare a furnace for another firing or to renovate one that had fallen into disuse. That is what happened in Zambia during World War II, when imports of iron were curtailed. To compensate for the absence of imports the Lungu people replastered their furnaces and began smelting once again with Iron Age techniques.

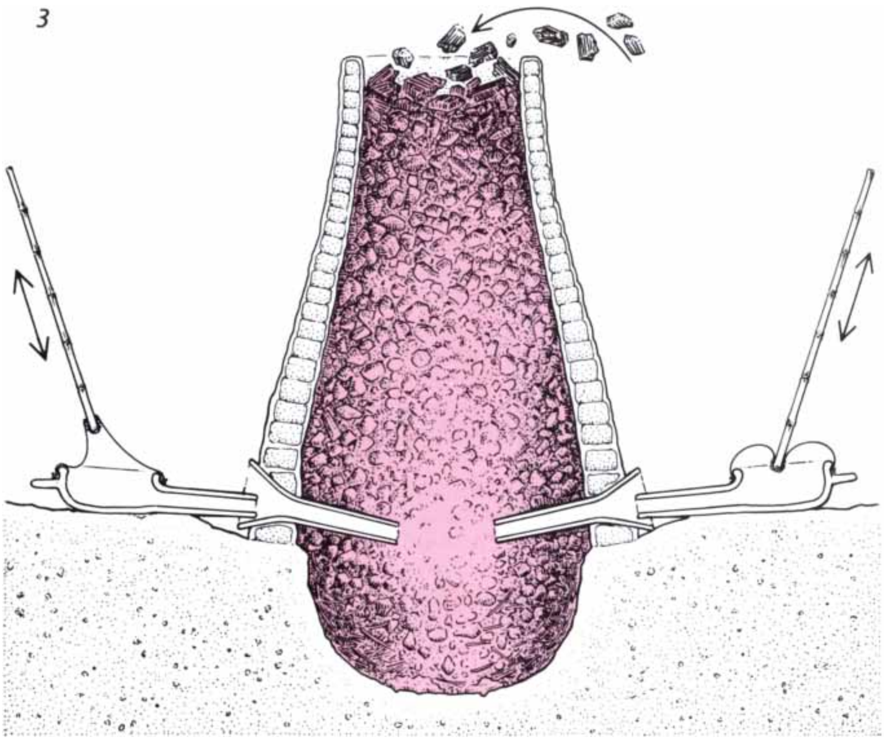
The lifetime of a furnace depended on its type. The low furnaces served only once before being dismantled to make possible collection of the iron. Tall furnaces, however, represented a larger investment of labor power. The tall furnaces used until recently by the Fipa in Tanzania stand as much as two or three meters high. Those built during the Early Iron Age were probably two meters or more in height. It would have been wasteful to rebuild such a structure each time it was fired, and so the furnaces were in all likelihood reused for many years. Vines and larger sticks formed a scaffolding that helped to hold them together and also provided a platform for loading. Large slabs of burned clay found with the furnace remains may have been doors that enabled the smelters to remove the smelted iron without dismantling the furnace.

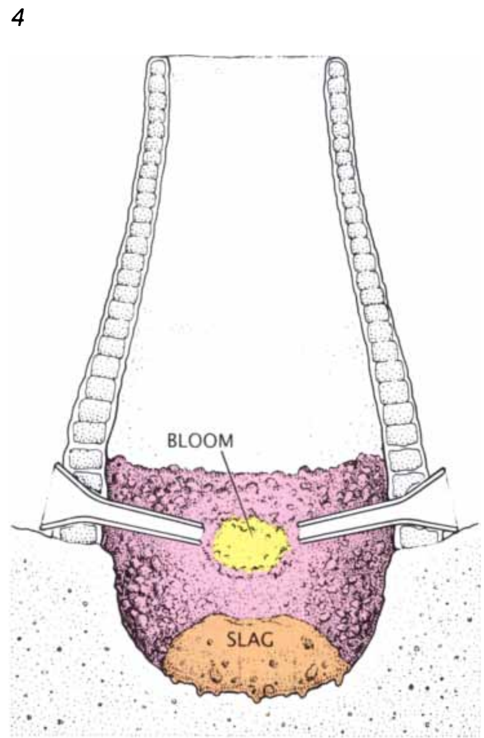
The need to provide an effective reducing atmosphere partly accounts for the shape of the furnaces: the tall, narrow furnace could be loaded with ore and charcoal, filling the space and keeping the amount of oxygen within to a minimum. Yet oxygen was needed for combustion, and so a means had to be found of supplying it where it was needed for the fire without allowing it to diffuse through the entire structure. The simplest method is to provide holes at the bottom through which the air can pass. Alternatively, the draft can be induced by human action. Recent archaeological research suggests that induced draft was employed in the tall furnaces.

Central to the system of induced draft were the tuyeres. These clay pipes are known from ethnographic descriptions and also from archaeological excavations. The tuyeres all seem to have been made in a similar way, according to ethnographic reports. Carefully prepared clay was pressed around wood poles; after the clay had dried slightly the pole was pulled out, leaving a hollow clay pipe. The clay pipes, sometimes linked in a series, were inserted into the base of the furnace. The outer end of the pipe may have been shaped like a funnel to facilitate the entry of air. Such a funnel-shaped endpiece was discovered in the wall of a furnace excavated at Gisagara, reinforcing the notion that induced draft was employed in Iron Age furnaces.

It is quite likely that bellows were inserted into the open end of the tuyeres. No bellows have yet been found in an archaeological context, but they are attested to abundantly in the region's ethnographic record. The bellows in use today consist of a large bowl, formed from clay or carved from wood, with a hollow shaft. The open mouth of the vessel is covered with an animal skin; a stick is sometimes attached to the skin. Moving the stick up and down forces air from the bellows through the tuyeres deep into the center of the furnace.

Once all the preparations were complete, the furnace was ready for loading. By the time the loading began the walls of the shaft must have been fairly hard, because no pieces of char-





charcoal and iron ore (*2*). Bellows were inserted into the openings of the tuyeres. The fire was lighted with a burning coal. More charcoal was added from above (*3*). The result of the smelting was a spongy "bloom" of iron and slag (*4*).

coal or ore are found incorporated in the bricks. The loading itself consisted in putting charcoal and ore into the shaft, probably in alternating layers that filled the tall structure more or less completely. Because there were no large openings at the bottom of the furnace, the loading of charcoal and ore had to be carried out from above by means of scaffolding.

The loaded furnace was lighted with a piece of glowing charcoal inserted through a tuyere into the center of the furnace at the base. As the fire began to heat up, the smelting process got under way. The smelting actually includes two parts: the reduction of iron oxide to metallic iron and the separation of the iron from the associated minerals, which run off as slag. Chemically, the reduction begins first. Oxygen, introduced through the tuyeres, combines with carbon in the charcoal to yield carbon monoxide (CO). It is the carbon monoxide that actually extracts the oxygen from the iron oxide, leaving carbon dioxide (CO_2), which escapes up the shaft, and iron. The reduction begins at a temperature of about 800 degrees C.

The second stage, the separation of the slag, begins at about 1,150 degrees C., the temperature at which the unwanted minerals begin to liquefy. The melting point of iron itself is about 1,540 degrees. Since it was not beneficial to melt the iron, part of the technical expertise of the smelters must have been to keep the working temperature of the furnace between 1,150 and 1,540 degrees—remarkable temperatures given the simple materials that went into the furnace. The outcome of the reduction was a spongy-looking "bloom" of iron. The bloom, containing charcoal and slag among other impurities, was removed and then reheated and hammered to purify it. Thereafter the iron was shaped into weapons, tools and ornaments.

Although the technical aspects of smelting are becoming clearer, its social and cultural context is unfortunately not well understood; anything that is said must remain somewhat speculative. Iron reduction must have been both labor-intensive and expensive in terms of raw materials. Labor power was needed not only to collect the wood but also to gather the ore, prepare the clay, build the furnace, make the tuyeres and operate the bellows. The coordination of labor and the control of its product suggest an efficient organization and perhaps also centralized control. It is not entirely clear, however, by whom that control was exercised during the Early Iron Age.

The social and cultural aspects of technology in the Iron Age are much more difficult to specify than the technical ones, because the evidence is less substantial. It can be said with confidence, however, that the smelter must have been a person of importance in the interlacustrine region during the Early Iron Age. He was the guardian of a complex ritual knowledge and technical skills that were of great significance for his society. Although we cannot yet fully understand the social relations that surrounded his work, on the basis of recent discoveries we have begun to understand the technology of which he was the master.

FURTHER READING

THE EARLY IRON AGE IN THE INTERLACUSTRINE REGION: THE DIFFUSION OF IRON TECHNOLOGY. Francis Van Noten in *Azania*, Vol. 14, pages 61-80; 1979.

THE ARCHAEOLOGY OF CENTRAL AFRICA. Francis Van Noten. Akademische Druck- und Verlagsanstalt, 1982.

EARLY IRON FURNACES WITH 'BRICKS' IN RWANDA: COMPLEMENTARY EVIDENCE FROM MUTWARUBONA. J. Raymaekers and F. Van Noten in *Azania*, Vol. 21, pages 65-84; 1986.

SCIENCE AND BUSINESS

Making Mice
Biotechnology: surviving the adolescent years

Biotechnology's teens
TRON takes Japan
Superconductors
Benchmarks for sale

When Harvard University investigators Mark Ptashne and Thomas Maniatis decided to form a biotechnology company in 1980, they did not bother to write a business plan. Nor did they focus on a specific product or technology. According to Ptashne, they thought the fledgling Genetics Institute would benefit from "complete flexibility."

Times have changed for the biotechnology industry. Almost a decade later, companies making therapeutic drugs with recombinant-DNA and monoclonal-antibody technology have grown in business savvy. Many are aiming for small, well-defined market segments. They also are looking at a wider range of financing options. Meanwhile the extent of patent protection remains an issue critical to companies' long-term strength.

As a result of these changes, business as much as science is guiding the industry. "Companies will fall out not because they don't have good ideas but because they aren't run well," observes Philip J. Whitcome, director of strategic planning at Amgen in Thousand Oaks, Calif.

Venture capitalists applaud the increased emphasis on specific market niches. Brook Byers, a San Francisco venture capitalist whose firm helped to launch Genentech, says he now typically looks for new investments that target specific clinical problems rather than building broad technology platforms. "I start with the disease and work backward," he says.

A case in point was the founding of Athena Neurosciences in San Carlos, Calif. About a year and a half ago "there were a bunch of venture capitalists looking into neurology," Byers says. After several months Byers and four others converged on the same researcher—Dennis J. Selkoe at Brigham & Women's Hospital in Boston—and jointly created Athena.

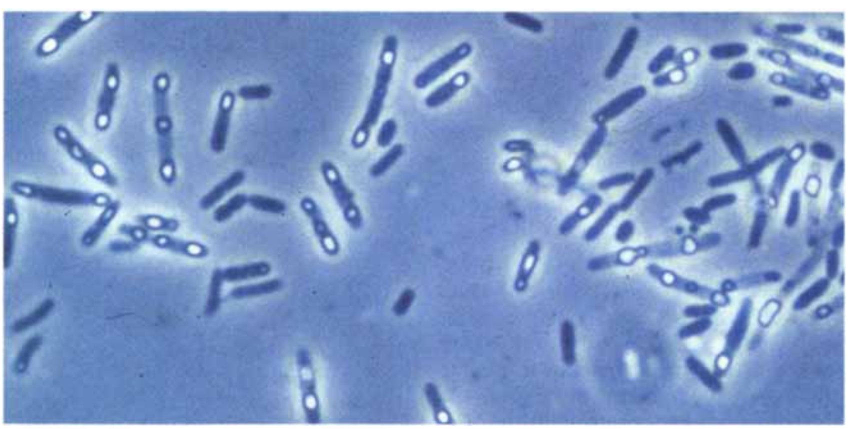
Exploiting a specific market segment has worked for the more established companies. Genentech has become the clear industry leader simply because it is the only newcomer selling its own genetically engineered therapeutic products—namely human growth hormone and tissue plasminogen activator, or TPA. (TPA breaks up blood clots in patients having heart attacks.) All told, there are only nine biotechnology therapeutic drugs on the market; seven are sold by established pharmaceutical companies.

Then there are a handful of strong biotechnology companies, including Amgen, Centocor, Cetus and Genetics Institute, that rely heavily on contracted research for revenues. Accumulated funds and interest also help to maintain viability. Most of these companies are testing therapeutic products; Amgen's erythropoietin (EPO), a protein that stimulates the growth of red blood cells and has been shown to counter anemia in kidney-dialysis patients, may be approved by the Food and Drug Administration later this year. Genetics Institute, which adopted a business plan long ago, is also developing a version of EPO.

Companies that are just starting up or are still in the early stages of research and development cannot hope to have a therapeutic product on the market before the mid-1990's. Some 275 such firms must spend much time looking for additional financing. A relatively unknown company such as BioPolymers in Farmington, Conn., found a patient investor in a Middle East businessman. Others, such as Seragen in Hopkinton, Mass., raised seed money easily but now find venture capitalists "more discriminating" when investing in a "mezzanine," or third, funding round.

As a company grows hungrier for funds, it is likely to turn to the public—specifically the over-the-counter market—for capital. (This market has lower capital requirements than either the New York or the American Stock Exchange.) Since Genentech went public in 1980, more than 180 companies have followed suit. The October market crash dampened but did not eliminate this source of financing.

Apart from raising capital, there are products or services a company can sell before its therapeutic products reach the market. BioPolymers, for example, has not yet put its protein-based adhesive for cells or tissue into clinical trials but sells it to laborato-



GENETICALLY ENGINEERED organisms include Escherichia coli *that synthesize the immune-system mediator interleukin-2* (bubbles) *and the first animal to be patented: a mouse predisposed to get cancer. The bacteria were developed at the Cetus Corporation and the mouse at the Harvard Medical School.*

ries. Almost 40 percent of Amgen's resources are given over to animal health care, diagnostics and specialty chemicals.

An alliance with a large pharmaceutical company is another option. The ultimate alliance is acquisition by the larger company. For their part, most pharmaceutical companies expect to enter the biotechnology field through acquisitions—a path charted by Eli Lilly and Bristol-Myers in 1985 and 1986 when they bought Hybritech and Genetic Systems respectively.

Biotechnology companies are also looking abroad for financing and markets. Linda I. Miller, an analyst with Paine Webber in New York, says that about 20 percent of the funding for U.S. biotechnology companies comes from abroad, and the proportion is growing. Japanese interest in biotechnology research is particularly keen.

Although surviving the research stage may be scientifically crucial, testing and marketing a product are also much costlier and equally challenging. Jumping the F.D.A.'s regulatory hurdles takes several years. Even putting a product on the market does not guarantee that it will be bought.

A Louis Harris & Associates poll conducted in late 1986 found that the public was concerned about genetic engineering in the abstract but approved of nearly every specific therapeutic application proposed. Recent events bear out that conclusion. The approval of TPA was greeted enthusiastically by doctors and patients; the main controversy has been over its high price. Yet a patent awarded in April to Harvard University for a genetically engineered mouse has rekindled concern in Congress and other groups about the ethics of patenting higher life-forms.

Of equal concern to biotechnology entrepreneurs is the security of their patents. The patent that covers the mouse is one of almost 900 that have been awarded for diagnostic and therapeutic biotechnology inventions. The courts have yet to resolve the strength of seemingly overlapping biotechnology patents. Often a protein, such as EPO, can be made with significantly different techniques. Will a patent for EPO cover all variations of the product, regardless of how it is made? Amgen and Genetics Institute are currently embroiled in just such a dispute. Genentech has not yet received a patent for TPA; when it does, the company may sue several other manufacturers working on similar versions of the drug. That "when" remains another critical issue. A backlog of 7,000 biotechnology applications is stacked up in the Patent Office; the review process takes about two and a half years.

Nevertheless, since the early days of the industry both scientists and investors have remained patient with the long development cycle of biotechnology products, "because they believed the products would be blockbusters," according to Ronald E. Cape, cofounder of Cetus. Although the halcyon funding days of 1980, 1983 and 1985, during which "any two men and a mouse could raise money without a business plan," may be over, Cape continues to look for a prolific and profitable future. *—Elizabeth Corcoran*

TRON Takes Hold

A Japanese computer project gains momentum

In Japan, TRON is the talk of the town. TRON, a somewhat strained acronym for "the real-time operating system nucleus," is a homegrown approach to computer architecture that is rapidly gaining favor with major manufacturers; even U.S. companies are taking a close look at it.

In essence TRON is a collection of computer specifications that aim to make it possible to connect distinct microprocessor-based systems as one network. The specifications describe a range of computer hardware and software—from microprocessors and an operating system to a new computer keyboard. It was proposed in 1984 by Ken Sakamura, an associate professor at the University of Tokyo.

The TRON specifications begin with very-large-scale integrated-circuit microprocessors. A personal computer version is called BTRON; industrial applications, such as computer-automated assembly lines, will use ITRON; mainframes or superminicomputers will take the form of CTRON, and MTRON will connect all the computer systems together. The systems are designed to be compatible.

Sakamura contends that different computers should be as easy to use as different automobiles. His long-term vision of a world of computers is more grandiose: almost everything, including factories, buildings and even furniture, could incorporate microprocessors connected in a massive intelligent network, he says.

U.S. manufacturers monitoring the progress of TRON are more pragmatic. "We want to know what TRON means to Japan and how to make our products more acceptable there," according to John R. Barr, manager of computer technology at Motorola's Computer X subsidiary in Schaumberg, Ill. "It's far too early to tell how effective TRON will be," adds James Ready, executive vice-president of Ready Systems in Palo Alto, Calif. Companies trying to sell computers in Japan may, however, have to offer products that are compatible with TRON, he says.

TRON is steadily moving into real systems in Japan. Some 100 companies belong to the TRON Association, chaired by the president of Fujitsu. Earlier this year Hitachi, Fujitsu, and Mitsubishi Electric introduced a jointly developed 32-bit TRON-based microprocessor. The first version of an ITRON operating system has been written by NEC, along with the three TRON chip manufacturers; a second version is expected to be available late this year. Matsushita, Oki and Toshiba are hoping to introduce micro-BTRON's late this year, and BTRON workstations are to follow as early as 1989.

Non-Japanese producers are also exploring the project. Siemens has posted a representative in Tokyo to monitor the program. Computer X is working with Sakamura to see if it can fit a CTRON interface to its real-time distributed computer system. Japanese subsidiaries of at least seven other major U.S. manufacturers belong to the TRON Association.

Outside Japan, however, TRON may face a harder time finding acceptance. Much computer software is based on either the MS/DOS or the UNIX operating system; such software would have to be rewritten to run on a TRON-based system. "It would be impossible to reproduce everything you find in UNIX," Barr says, adding that so far TRON systems are not significantly faster or more efficient than other systems.

Sakamura remains optimistic about TRON's future. An eight-part television series on TRON was broadcast early this year and received "very high ratings" for an educational or technical program, he says. Moreover, research for a "TRON house" and even a leisure resort is under way. *—E.C.*

Cold Start

Industry edges cautiously toward new superconductors

American industry finds itself being nudged into commercializing high-temperature superconductors—whether it wants to or not. Although most researchers argue that applications of such superconductors are far in the future, the Government has been encouraging companies to pursue the technology as well as the science.

The most recent effort to promote applications is a project headed by President Reagan's former science adviser George A. Keyworth II. Keyworth chairs the Council on Superconductivity for American Competitiveness, a nonprofit group that boasts the Rensselaer Polytechnic Institute's Roland W. Schmitt and Texas Instruments' Robert Stratton among its 19 directors. The council is sketching a business plan for what it calls the Superchip Corporation.

"The name says a lot," Keyworth remarks, adding that it must produce tangible benefits for whatever groups become participants. Superchip will aggressively push research and development of such thin-film applications as microprocessors for high- and low-temperature superconductors.

A plan to move rapidly into applications is likely to meet with skepticism from industry. Few companies are following the path suggested by Reagan last year, when he urged the private sector to move toward commercializing the scientific breakthroughs. "It's so early to be making estimations on products of [high-temperature] superconductors that I have to laugh at some of these prospectuses that I read, especially from some small company that says 'We are going into production,'" remarks General Electric's senior vice-president for research and development, Walter L. Robb.

"From the computer industry's point of view," adds Paul M. Horn, IBM's acting director for physical science at the Thomas J. Watson Research Center, "the leverage of superconductors is long-term. We're talking about more than 10 years." More than 80 percent of IBM's research in this area is devoted to understanding the mechanisms of high-temperature superconductors and looking for new ones.

Researchers in U.S. industry maintain that their focus and pace are appropriate. "Since we don't understand the mechanism underlying high-temperature superconductivity, there's no real direction for looking for other superconductors," says Robert C. Dynes, director of the AT&T Bell Laboratories chemical-physics research laboratory. Certainly the challenges of processing are formidable: high-temperature superconducting materials carry low current densities and work poorly near strong magnetic fields. There is also "no substrate of choice for high-speed, high-current-density thin films," Dynes observes.

Even basic research has lagged somewhat, observers say. Although a small group of companies led by IBM and AT&T are vigorously supporting work in high-temperature superconductivity, most others are "nowhere near where they should be" in such work, according to Brian B. Schwartz, education officer for the American Physical Society in New York. Even large companies that have a strong commitment to research, such as General Motors, say that since high-temperature superconductors do not currently fit into corporate business plans, any research projects remain small. For its part, "middle-size industry very much has a 'wait and see' attitude," Dynes says.

In Japan, as business-school professors are fond of saying, things are different. More than 80 companies, including banks and steel manufacturers, are supporting a superconductivity-research consortium; half of the participants will contribute $800,000 each to the effort. The consortium aims to push both basic research and applications.

The Administration continues to nudge companies into looking for applications by allocating some funds for processing and fabricating superconducting components such as wires and thin films that can be scaled up for manufacturing. In fiscal 1988 the Defense Advanced Research Projects Agency will spend $15 million on 20 industrial projects. —*E.C.*

Bottled Water?

Don't bother waiting for the year-end sale

Two hundred milliliters of simulated rainwater for $189 might not sound like much of a bargain at first. A glass slide with a few thousand 9.94-micrometer polystyrene spheres sandwiched inside for $86 has elicited few calls.

That suits the National Bureau of Standards just fine. Since 1906 the NBS has run a tidy business selling standardized materials, from anisic acid to zirconium. This year its gross sales are likely to reach $6.3 million, according to Stanley D. Rasberry, chief of the office of standard reference materials.

Whereas most producers would love to monopolize a business, the NBS would not mind more competition. Rasberry says it receives requests to measure the precise chemical and physical properties or compositions of more than 200 new items per year but can manage to handle only about 50 of them. After analyzing a material the NBS bottles it and attaches a certificate describing the composition or dimensions. Those who buy the simulated rainwater can be sure it contains 2.69 milligrams of sulfate per liter, .205 milligram of potassium per liter and a known quantity of other substances. The product serves as a benchmark. Knowing the precise composition of the water, a user can check the accuracy of laboratory equipment for measuring acid rain.

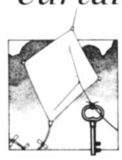
Standard reference materials are essential to every industry, and so they are a coveted commodity at any price. Iron foundries were among the first to make the discovery. In the early 1900's the wheels of railroad cars often fractured because of faulty composition of the metal, sometimes derailing trains and killing passengers. The foundrymen knew they could correct the problem by consistently making high-quality iron, but their laboratory instruments did not give them accurate results. They asked the young NBS to produce a benchmark material and some 14 months later the bureau handed over four bottles, each containing 150 grams of cast-iron chips that had a slightly different specified composition.

The NBS still sells the bottles of iron chips (at about $90 a bottle), having recertified the composition of new batches of the material in the 1970's. All told, the NBS offers some 1,000 commodities, each of which was specifically requested. In spite of booming sales of about 44,000 items a year, the NBS does not make a profit. The price an item fetches covers only a little more than 90 percent of its research and production cost. The Government foots the rest of the bill.

The NBS added the first products made in space—the tiny polystyrene spheres, which serve as a measurement standard—to its wares in 1985. It first packaged them 33 million to a vial and priced the vials at $304 each. Supply and demand is taking a toll: the price of a vial of space spheres has almost doubled since the first sale. "We have only half of the material left," Rasberry says, "and we can't get any more." The NBS decided to extend its supply (and lower the price) by putting only a few thousand on a slide. Although Rasberry had thought universities would be eager to buy them, sales have lagged. "And they're such a bargain," he muses. —*E.C.*

THE AMATEUR SCIENTIST

Does convection or the Bernoulli principle make the shower curtain flutter inward?



by Jearl Walker

Those of us who take showers are often bothered by the shower curtain's habit of moving inward, sometimes to the extent that it interferes with uninhibited lathering, splashing, rinsing and singing. Its motion can be countered in various ways. If the shower is in a bathtub, small magnets along the bottom will hold the curtain to the tub; so will simply wetting the outside of the curtain. A second, decorative curtain can be hung outside the tub, where it seems to block a flow of air that would otherwise push the working curtain inward. If the shower is in a stall rather than a tub, the only remedy I know of is to add small weights along the bottom of the curtain.

Curing the problem is easier than explaining it. Is it caused by a flow of air under the curtain and into the shower compartment, as one might expect when the shower is hot and the room air is cool? Or does the curtain move because the pressure of the water or the air in the compartment is reduced by a principle of fluid flow called the Bernoulli principle? When I investigated the matter, I discovered that neither explanation appears to be correct. The Bernoulli principle seems to have no role in the curtain's motion and the airflow under the curtain is usually outward, even during a hot shower.

I begin with what is perhaps the most popular explanation: that the curtain moves because of the principle named for the 18th-century Swiss mathematician Daniel Bernoulli. The explanation has it that the water in the shower is at a pressure lower than atmospheric pressure because it is moving; the water's lower pressure reduces the air pressure in the shower compartment and so, because the air outside the curtain is at atmospheric pressure, the curtain is forced inward. Proponents of this explanation point out that the faster the water flows, the more the curtain moves in. They reason that the faster flow means more reduction in the water and air pressure within the compartment.

The Bernoulli principle involves the energy of a fluid in motion. If the flow is constrained by walls, as it is in a pipe, the energy must remain constant (provided that friction from the walls is not important). Suppose the flow passes through a section of pipe that narrows. When the water enters the narrow section, the water's speed increases; so does the energy—kinetic energy—associated with speed. The increase in the kinetic energy is at the expense of the fluid's pressure, which is a form of potential energy. The Bernoulli principle states that the exchange leaves the total energy unchanged. In a pipe the regions of faster flow have lower pressure.

The shower jet is different because it is not constrained. When the water emerges from the shower head, its pressure matches that of the air around it, which is at atmospheric pressure. As the water falls it certainly increases in speed (as do most things that fall), but the increase in kinetic energy is provided by gravity; no energy is removed from the water pressure, which continues to match atmospheric pressure. A simple application of the Bernoulli principle, then, fails to explain the curtain's motion; it also overlooks the fact that a shower is often a spray rather than a jet.

Sometimes advocates of the Bernoulli principle consider the movement imparted to the air by the jet or spray. Some of the air is entrained by the water and thus made to move downward. Since air is a fluid, might its motion be covered by the Bernoulli principle? In other words, does the air gain kinetic energy at the expense of its pressure? No, it gains kinetic energy because it is forced to move by the water.

Another popular explanation for the shower curtain's movement involves a chimneylike convection. When the shower water is hotter than the air in the room, it heats the air within the shower compartment, which rises and flows out over the curtain and into the room. Cooler air from the room should flow under the curtain and into the compartment, replacing the air lost over the curtain; the influx should blow the curtain inward. The explanation is bolstered by two observations. The hotter the shower is, the more the curtain moves inward, presumably because of stronger convection. Also, if the influx is blocked by a second, exterior curtain, the interior curtain may be stationary.

For a long time I assumed that chimneylike convection was responsible for the curtain's motion. Then, when I was leading a workshop at a hotel a few years ago, one of the teachers in the audience tested the idea. Returning to her room, she turned on just the cold water in the shower. If the heating-and-convection explanation was correct, the curtain should not move inward—but the experimenter quickly returned with the news that the curtain did move inward even in the absence of heating.

What then does account for the curtain's movement? I decided to investigate two shower arrangements. One was a large shower stall enclosed by tile walls and a glass door (except for a space above the door). The other shower was in a bathtub, bounded by three tile walls and with a single plastic curtain. In both compartments the water struck the middle of the floor. The impact region where the water hit the floor took up about half of the width of the stall and the full width of the bottom of the bathtub. I began my work with the stall shower, because the stall was roomy enough for me to explore the air circulation without having to stand under the shower.

I monitored the airflow within the compartments with a lighted candle. The flame's orientation provided a crude measure of flow direction and speed—particularly in the case of horizontal flows, which tilted the flame from the vertical. Upward flows of air lengthened the flame and downward flows shortened it. Turbulent air set the flame dancing wildly.

It was winter, and so I was concerned that variations in the house heating might influence convection in the two bathrooms. To reduce that possibility, I closed the doors to the rooms and kept the thermostat controlling the house heat at a constant setting; after two hours the air in the rooms settled to a temperature of about 20 degrees Celsius. Running a hot or a cold shower would, of course, tend to warm or cool the bathroom, and so I worked quickly enough to prevent the room temperature from changing by more than a few degrees in the course of a trial.

In each trial I recorded the directions of the airflow on photocopied sketches of the shower compartments on which I had labeled the three walls. The "back wall" was the wall opposite the door or curtain, the "head wall" was the one with the shower head and the "foot wall" was opposite the shower head.

My candle flame proved to be surprisingly durable when I probed the air near the falling water, but of course it was often extinguished. To relight the candle I kept a second one burning in the room. Whenever my probe candle was doused by the shower stream, I reached out to hold the wet wick in the flame of the second candle; the flame evaporated the water in a series of sputters and then relighted the wick. (If you try these experiments, take care not to burn yourself or to let the unattended second candle start a fire.)

The door into the large shower stall was near the head wall. With the door open, I ran a shower with only cold water, which was at five degrees C. Although the water was painful to my bare feet (I was otherwise clothed), I managed to map the airflow throughout the compartment and in the doorway [*see illustration below*]. Air flowed vigorously outward along the entire width of the bottom of the opening. Along the top of the opening the air flowed outward on the foot-wall side but inward on the head-wall side. At all levels inside the stall the air circulated horizontally around the falling water, moving from the head wall along the back wall and then along the foot wall and finally escaping through the doorway.

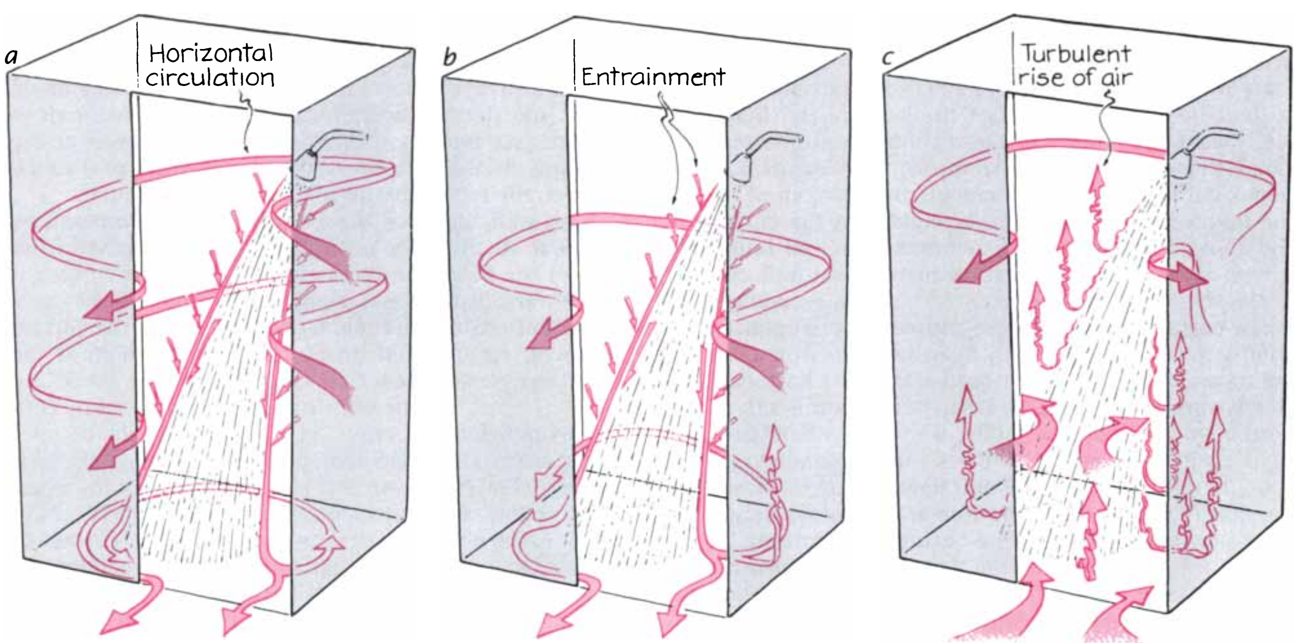
When I held the candle near the falling water, the flame was pulled toward the water, revealing that the air was being strongly entrained. The layer of entrainment was about one centimeter thick on the head-wall side of the spray and twice as thick on the opposite side. Except near the floor, air was entrained along the entire length of the falling water. Along the floor near where the water hit, air flowed strongly away from the impact region and then either joined the horizontal circulation or escaped directly through the doorway. Since the water cooled the air within the compartment, some part of the airflow pattern I recorded was certainly due to convection.

In order to eliminate convection, I next opened the hot-water valve until the shower water was within a degree of the room temperature. (I also adjusted the cold- and hot-water valves to maintain the original flow rate.) After allowing 30 minutes for the compartment to adjust to the new water temperature, I began probing the airflow with my candle.

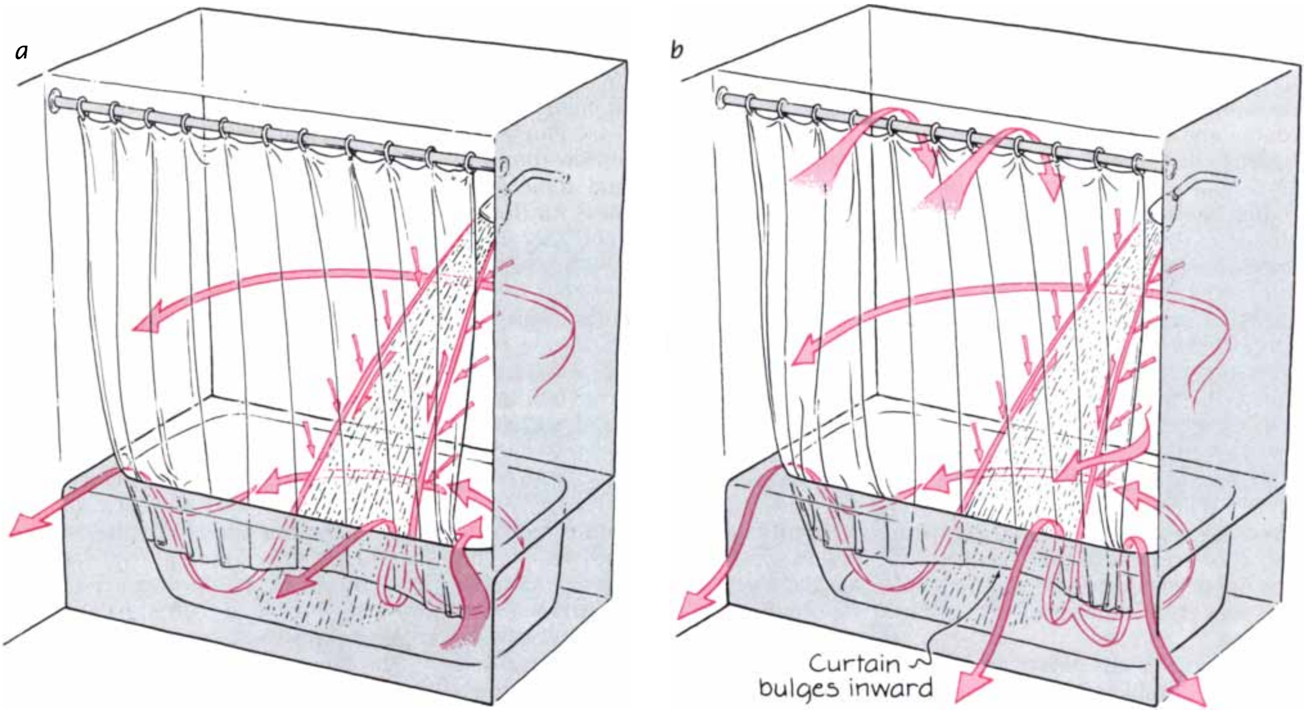
The horizontal circulation around the falling water was again clearly present near the floor and at intermediate heights, but now it was weak near the top of the compartment. Turbulent air flowed upward near the head wall. On the head-wall side of the falling water the entrainment of air was similar to what I had seen earlier, but on the foot-wall side the entrained air veered off toward the back wall to join in the horizontal circulation. As before, air flowed strongly outward at the bottom of the doorway, but now the inflow through the top half of the opening was weak.

The most interesting aspect of the flow pattern was along the floor of the compartment. Air that was released from entrainment flowed away from the impact region. Some of it escaped immediately through the doorway; the rest was eventually forced upward by a wall and then into the horizontal circulation around the falling water.

Next I cut off the cold water and increased the hot-water supply, again attempting to maintain the same flow rate. The water temperature rose to 51 degrees C. but then gradually dropped to 45 degrees as my water heater struggled to keep pace. The airflow



The air circulation in a shower stall when the water is cold (a)***, at room temperature*** (b) ***and hot*** (c)



The air circulation when a tub shower is at room temperature (a) ***and cold*** (b)

within the compartment was remarkably different from that in the previous trials. Along the bottom of the doorway air flowed inward vigorously. The horizontal circulation around the falling water was missing except near the top of the compartment. Everywhere along the bottom of the compartment the candle revealed strong upward flow. Even the entrainment layers seemed to be modified: apparently the entrained air heated so rapidly that it broke free of the water and actually rose.

I moved from the shower stall to the bathtub shower. My findings in the stall had taught me that convection can greatly alter the circulation system established by entrainment. In the bathtub I therefore first brought the shower water to room temperature in order to explore what happens in the absence of convection.

The curtain moved into the shower compartment, most strongly near the falling water and near the head wall [*see illustration above*]. Air circulated horizontally around the falling water as it had in the previous trials. Air near the falling water was entrained, brought down to the tub bottom and then sent outward along the bottom until it was forced up by the side of the tub; where the curtain bulged inward, this upward flow sent the air out into the room. The only strong flow of air into the compartment that I detected was in the lower part of the small space between the head wall and the end of the curtain.

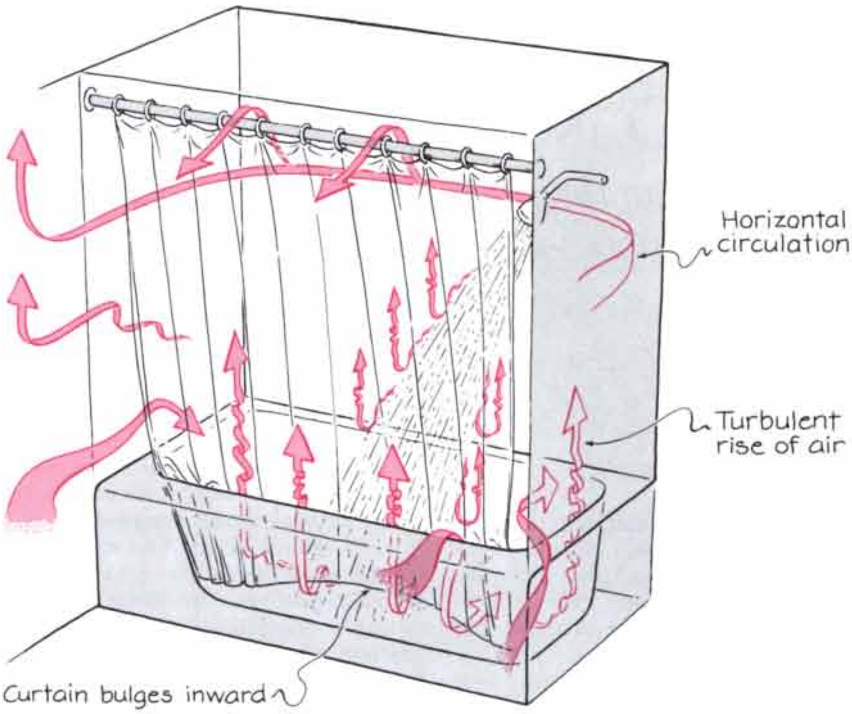
When I turned on only cold water, the airflow pattern changed in two respects. Air poured out through the lower part of the space between the curtain and the foot wall and air flowed into the compartment over the top of the curtain. Otherwise the flow pattern was like the cold-water pattern in the shower-stall trials.

When I turned on only hot water, most of the curtain bulged into the compartment; air flowed vigorously outward through the space provided by the bulge, except near the head wall, where it flowed inward [*see top illustration on opposite page*]. At the bottom of the space between the end of the curtain and the head wall, air flowed into the compartment. At the foot-wall end of the curtain the flow was chaotic but generally inward. The horizontal circulation had shifted to the top of the compartment; air in the bottom of the compartment rose turbulently.

When the trials were completed, I finally understood why a curtain moves inward when I take a shower. Both entrainment and convection are important, but entrainment is the primary factor [*see bottom illustration on opposite page*]. Entrainment removes the air that is next to the falling water, except near the bottom. On the head and foot sides, additional air flows directly toward the falling water to replace the removed air, only to be itself entrained and removed. In the region between the curtain and the falling water the replacement is slow, because the air must flow first along the curtain and then toward the water. When the shower begins, the air pressure in that region drops as a result of that slow replacement. Since the air outside the curtain is at atmospheric pressure, the curtain billows inward. Although the bottom of the curtain is near where entrained air is released, the edge is pulled in by the higher parts of the curtain. The bulge along the bottom opens an avenue for air to flow into the room. The flow is nearly always outward, regardless of the relative air and water temperatures.

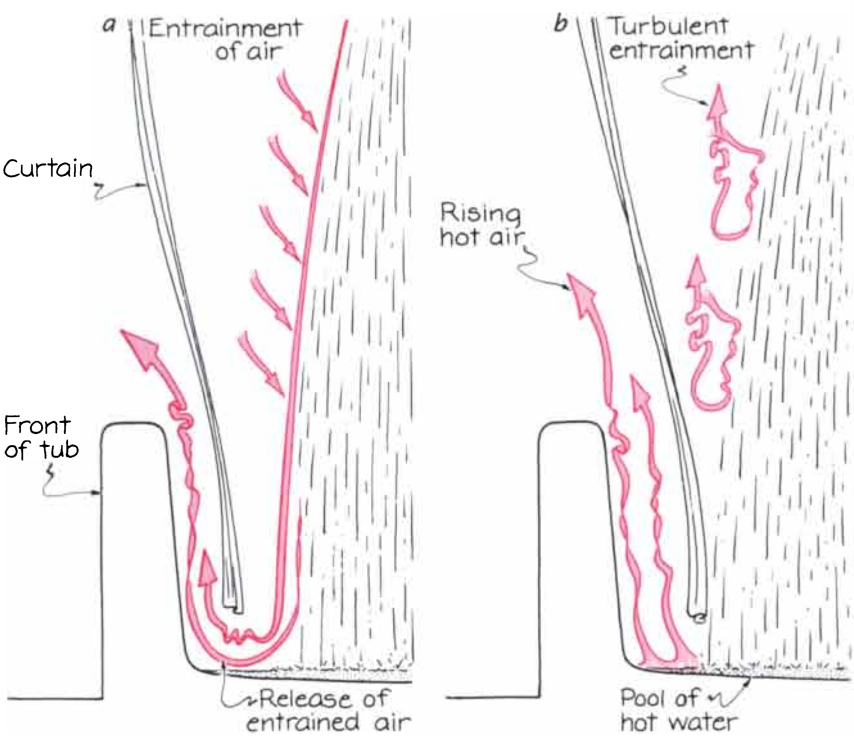
If the water is at room temperature, the outward flow is air released from entrainment just below the curtain; it flows along the bottom of the compartment, upward outside the curtain and finally out into the room. If the water is colder than the room air, the outward flow is enhanced by convection as the chilled air escapes into the room and warmer air flows over the curtain and into the compartment. The enhanced flow under the curtain may move the curtain in more than entrainment alone does.

If instead the water is hotter than the room air, there may be little release of air under the curtain because the air entrained by the falling water quickly heats up and breaks off from



The air circulation when a tub shower is hot

the fall. The bulge in the curtain does, however, expose a pool of hot water on the floor of the stall (or the bottom of the tub). The pool heats the air above it, sending the air upward against the outside of the curtain and from there out into the room. The resulting extra push on the curtain is the reason the curtain may move inward more during a hot shower than during a cooler one. To sum it up, entrainment initiates the inward movement of the curtain and convection can enhance the effect.



Entrainment by a room-temperature shower (a) *and by a hot shower* (b)

SCIENTIFIC AMERICAN

CORRESPONDENCE

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

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the flat fee of $1.25 per copy of each article or any part thereof. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 21 Congress Street, Salem, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/88. $1.25 + 0.00.

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

Advertising correspondence should be addressed to Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017.

Address subscription correspondence to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 953, Farmingdale, N.Y. 11737. The date of the last issue on subscriptions appears at the right-hand corner of each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (if convenient, on a mailing label of a recent issue) as well as the new one.

Name

New Address

Street

City

State and ZIP

Old Address

Street

City

State and ZIP

COMPUTER RECREATIONS

Imagination meets geometry in the crystalline realm of latticeworks

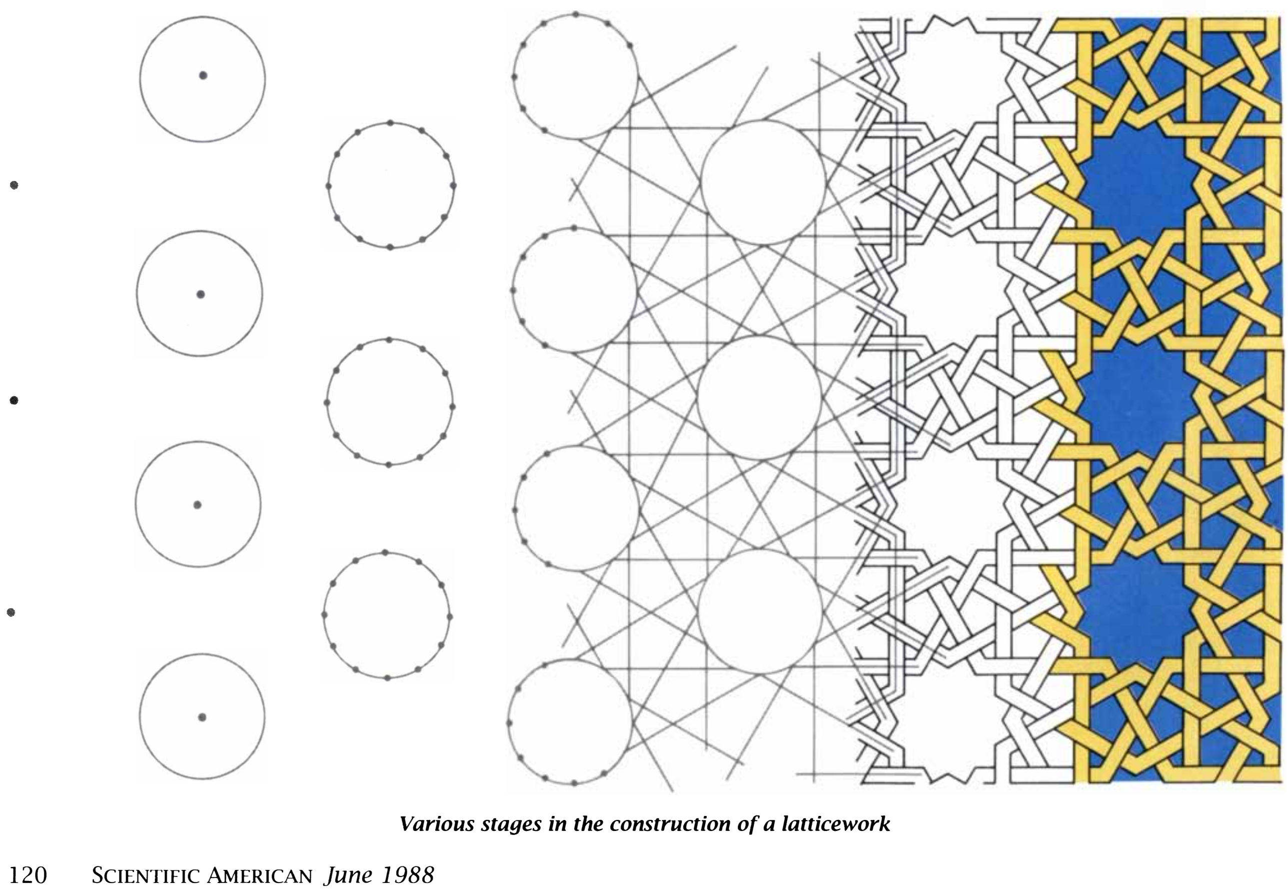
by A. K. Dewdney

Although computers are creeping tone by tone and pixel by pixel into the arts, vast areas of endeavor still remain open to human beings. The designs I call latticeworks are not exhibited on computer screens but rather on ordinary sheets of paper. The latticeworks are not produced in a few seconds by a computer program; they emerge slowly from a ruler and compass guided by the human hand.

The designs begin when imagination meets a grid of circles. Anyone can play and possibly produce an amazing interlocking network of lines that confuses and delights the mind. The designs are similar to those found on ancient tombs, mosques and palaces from Samarkand to Seville. Rectilinear ornaments from the Islamic world of medieval times hint at infinity within a finite space. Did the artisans use the methods I shall describe? Experts are uncertain what method was employed, but the mathematical reverberations of the patterns echo from geometry to topology.

There are many ways to view a latticework, but perhaps a good starting place is to look for symmetry. The finished version of the specimen below, for example, has a high degree of symmetry: one can rotate it by 120 degrees about certain points and end up with an identical pattern. One can also reflect it across certain lines and get an almost identical pattern. The chief elements of the pattern are gold bands that weave like small highways of thought through a potentially infinite landscape. Throughout the design circles are spaced according to the underlying symmetry of the pattern. The bands travel from one circle to another, converging in 12's and then abruptly bending away. The angle of reflection equals the angle of incidence. The bands themselves have shapes, in this case either hexagonal or zigzag. Everywhere one looks, overpasses alternate with underpasses. By what means was this arranged?

Latticeworks belong to a wide class of infinite patterns that have at least two independent symmetries of translation: if one translates, or slides, the pattern over a copy of itself in either of the two directions, eventually the pattern and its copy will line up with each other. No matter how cleverly one lays down an infinite pattern, as long as it has two independent translations it will inevitably belong to one of the 17 possible crystallographic groups first classified by the Russian scientist Evgraf S. Fedorov. Considering the enormous wealth of lattice-



Various stages in the construction of a latticework

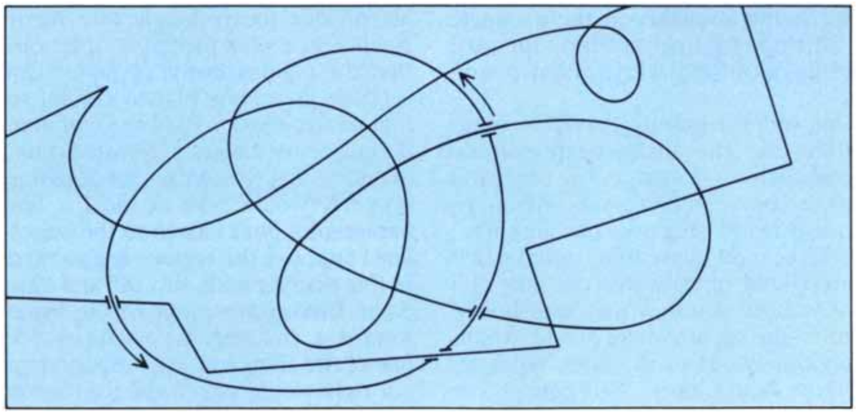
works inherited from the past, it is not surprising that examples of all 17 groups can be found, whether peeping shyly through a palace window grille or proclaiming the sultan's grandeur in the audience hall.

Translations are just one of four possible symmetry operations that can be done on a plane imprinted with a pattern. A reflection flips the plane like some vast door about a line that lies within the plane itself. This is essentially the looking-glass operation, in which letters get reversed and faces look slightly different. The third symmetry operation is the rotation, and in it the plane is rotated through a specific angle about a single point. The fourth and final symmetry is called a glide reflection. It consists of a translation followed by a reflection about a line that is parallel to the translation.

Symmetry operations (of whatever type) can be combined simply by performing them in some sequence. An algebraic structure called a group emerges from all of this. In a group the product of any two operations is an operation, and every operation has an inverse operation that in effect undoes it. In addition a group has a so-called identity element (the do-nothing operation). A group must also satisfy the associative law: when three successive operations are carried out, it does not matter whether the third operation follows the first two or whether the last two follow the first one. Put this way the law sounds mildly idiotic, but only because the associative law happens to be transparently true in the case of symmetry operations on the plane. The symmetry group of the latticework in my example is called *p*6 in the international crystallographic language. It is characterized by having rotations of 60 degrees about one set of centers and rotations of 120 degrees about another set. Readers will readily find the centers in the illustration.

I produced the design a few years ago by the method I shall outline here. Later I discovered the same latticework in a book about Islamic art of the medieval period. My deflation at not being first was more than compensated for by the discovery itself; the method seemed confirmed. Since then it has been my lot to "rediscover" other designs.

The method requires the would-be artisan to set up a grid of points. The grid is limited to one of four types: triangular, square, rectangular or hexagonal. Such grids are easy to lay out by ruler or compass: draw a base line, then use the compass to mark off evenly spaced points. Square and rectangular grids employ a right-angle construction to add new points above and below the base line. Triangular and hexagonal grids require equilateral triangles.



The "over-under" rule in progress

The design under discussion began life as a triangular grid. A circle is then drawn at each of the grid points. Here intuition makes its first entrance, because the size of the circle happens to be critical. In a moment I shall explain the role intuition may play in selecting the size.

Once the circles are all drawn, the amateur artisan selects points evenly spaced around the circle. These points will anchor the lineal elements of the design. The position and number of points must reflect the symmetry of the grid itself. In other words, the points should preserve symmetry with respect to reflections and rotations. Any of the intended symmetries should carry points on one circle onto points on another circle or on the same circle. In the example under construction the number of points on each circle must be a multiple of 3. I chose 12 to give body to the lattice. Since the underlying pattern was to have reflectional symmetry, only two positions for the points on the circles were possible. I chose the position in which six of the points were closest to the surrounding circles. As a general rule, whenever there is a choice, the best decisions are those that harmonize with a symmetry already present.

In the next stage of construction one joins the points on each circle to points on other circles. Here intuition makes a second entrance. The possibilities appear to be so numerous that only intuition would seem to serve. In fact, the combinatorial possibilities are greatly limited once more by conditions of symmetry: if I join a certain point on circle *A* to another on circle *B*, the symmetries of the pattern carry that connection onto other points on circle *A*. Before one has drawn more than two lines it may be necessary to erase the experiment and try another connection.

A kind of feedback loop binds this design phase with the earlier choice of circle size. Once a seemingly satisfactory and consistent scheme of interconnection has emerged, the results may look unauthentic, not to mention ugly. The lines do not harmonize with the symmetry of the pattern. In such a case it is usually obvious whether shrinking or expanding the circles will produce connections that parallel the major symmetries of the pattern. Here intuition may provide the leap of insight, a kind of artistic "Aha!" experience. In the mind's eye one suddenly sees the lines generated by the new circles.

At this point in apprenticeship a certain excitement causes the compass and ruler to quiver slightly. An amazement that is perhaps half artistic and half mathematical grips the holder of these instruments. Should one take credit for an intuition that was merely acceding to geometry?

In any event, it is now time to pave the highways of thought by giving them some width. No sophisticated system of roads should suffer traffic lights, and so how can the angular freeways be interlaced? In true weaving, after all, overs and unders alternate. Can the artisan be forced into some kind of logical cul-de-sac where a road faces two consecutive underpasses? Topology saves the day.

The following experiment in scribbling shows how. Draw a large rectangle on a sheet of paper. Then scribble inside the rectangle, abiding by only two rules:

1. If a line begins or ends, it must do

so on the boundary of the rectangle.

2. No more than two lines (or parts of the same) may cross at any point.

One can, for example, scribble something like the abstract-expressionist composition shown on the preceding page. Lines can be curved or bent, repeatedly crossing over one another.

To convert the scribble into the cleverest knot imaginable, it is now only necessary to follow the "over-under" rule: starting anywhere one likes, simply follow one of the lines, repeating "Over, under, over..." Of course, one must make an overpass for the line being followed when "Over" is said. By the same token, "Under" calls for an underpass. Eventually one either returns to where one started or the boundary of the rectangle is reached. En route the line may have crossed itself. Amazingly, whenever one arrives at a previously processed crossing, it already has the required structure. In other words, one never finds that an overpass is called for at a crossing that has already been designated an underpass. Eventually all ridges have been built and the scribble takes on an appearance that is almost intelligent.

The over-under rule works because in a sense it must. The simplest demonstration at a public level invokes a pleasant thought excursion. The scribble divides the rectangle into many small regions, or pieces. It turns out that the regions in the rectangle can each be given one of two colors, so that no two regions sharing a common boundary are assigned the same color. (A convincing elementary proof of this property would take at most a few paragraphs, but I hasten to the punch line.) Suppose the regions are painted in this manner with, say, red and blue paint. Driving along one of the roads toward a crossing, we would notice one of two things: either the region on our right would be red and the one on our left would be blue or vice versa.

The recipe for crossings lurks in this simple observation. Traveling along a given road, the mental construction crew will know whether to build an overpass or underpass when they arrive at a crossing. If the region to the right is red, build an overpass; if it is blue, build an underpass. The result is exactly the same as (or possibly the "negative image" of) the result if the crew had followed the over-under rule stated above.

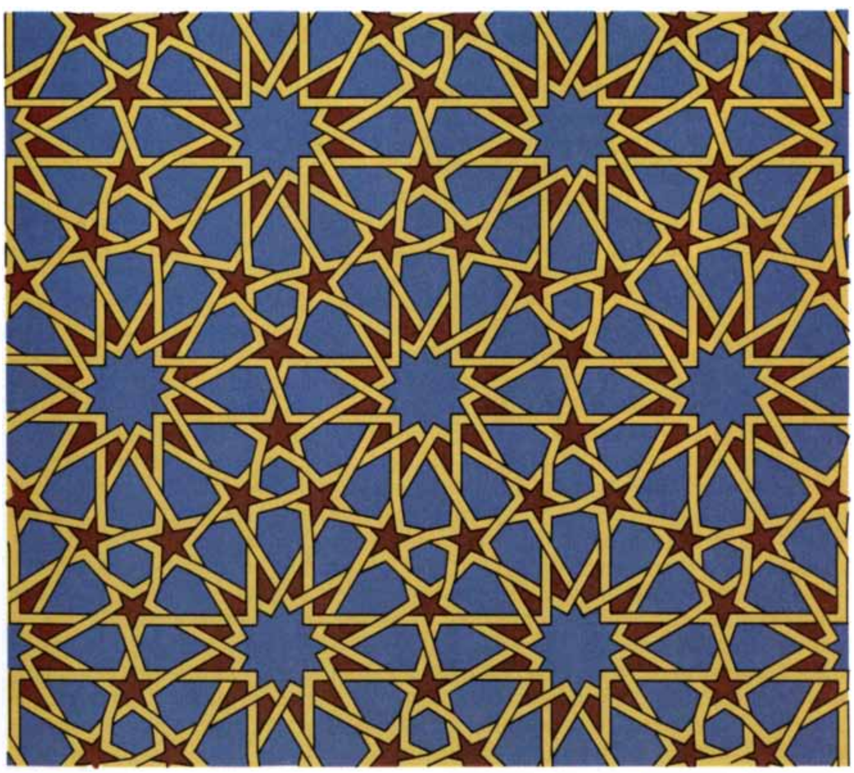
With a pencil and a ruler the designing reader can easily thicken each line of the latticework by drawing a parallel line on each side of it. At bends there is some fiddling with the meetings of these lines, but the project proceeds more or less mechanically. The thickening procedure may ignore the crossings until they are complete. Equipped with an eraser, the artisan now attacks the crossings, invoking the over-under rule as he or she goes. To create an overpass, one must erase the two segments of road edge that cross the road one travels. As soon as this is done the overpass springs into existence. The road that was erased seems to pass under the road one is on. Underpasses are created by the opposite procedure. It is interesting to note that the interlacing procedure destroys all reflectional symmetries of the pattern; the new pattern may look the same but underpasses and overpasses have been swapped.

The final stages in the creation of a design involve inking and coloring. A good pen will follow the ruler and produce an even line with occasional interruptions at bends and crossings. A design of moderate complexity may take an hour or more to ink, but what is an hour in the timeless world of the artisan? There is time to think of other things during this phase. It is perhaps a loose form of meditation.

When the design has been inked and all the pencil marks have been erased, it can be colored directly. Because coloring does not always turn out as well as one might like, it may be preferable to copy the inked original and color the copy. In this way the original can produce many offspring, each more beautiful than the one before. I suggest using tempera paint. It goes on evenly, produces minimal wrinkling of paper and is available in virtually all colors, including gold and silver. Moreover, water-based paints such as tempera seem to be repelled by xerographic inks. This is fortunate because unless one has a very steady hand, overpainting of ink lines seems inevitable. One wants all colored areas to be finely edged with black in order to enhance contrast.

The colors one selects are of course a matter of personal choice. Authentic latticeworks often employ dark and muted primaries for the regions between the bands. Such treatment results in a retreating background that points up the latticework itself all the more prominently. I have included another design that uses such a color scheme [*see illustration at left*].

The second design was produced by methods similar to the first. Advanced latticeworks have not only what might be called primary circles of inflection but also secondary ones. Each five-pointed star in the second design arises from such a secondary circle. I have included two additional charming examples of latticework in merely skel-



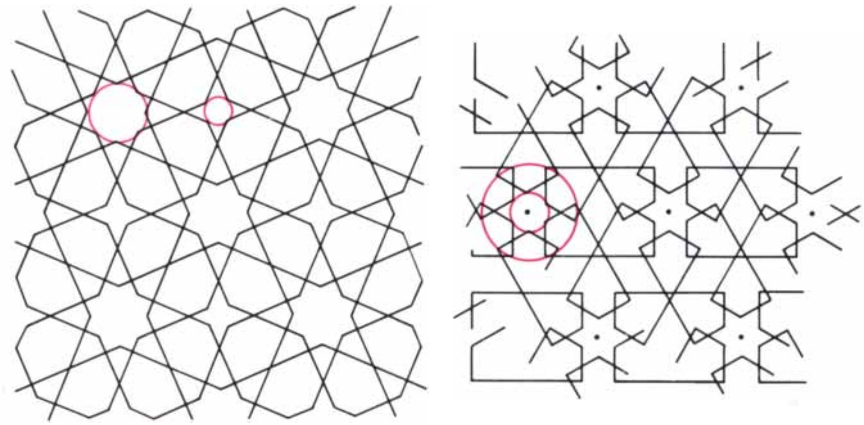
Islamic latticework

etal form. In one of them the primary and secondary circles are situated on different centers of rotational symmetry; in the other they are concentric [*see illustration at right*]. Both latticeworks can be found in traditional settings.

Latticeworks are closely related to two-dimensional tessellations: patterns produced by laying down multiple copies of a single shape without overlap. Each latticework can be produced in the form of a single, finite shape, more specifically a tile. On the tile appears a small part of the overall design. If one has available a great many copies of the tile, one can reproduce the latticework to any extent by laying down enough tiles in the proper way. An examination of any of the latticeworks shown here will reveal a small area in each example that could act as a tile. Here, in any event, lies the prospect for a larger view of any such pattern. Make a tile of sorts on paper and make many copies of it. When the tiles are colored, they lend themselves to a host of decorating ideas.

But what of the original latticework designs? How were they produced and what, for that matter, did they mean? Lisa Golombek has studied the designs for many years, both on site and in her office at the department of Middle East and Islamic studies at the University of Toronto. She thinks some kind of geometric underpinning in the form of grids and circles is the likely method. Undoubtedly individual craftsmen employed variations on the basic technique. As for meaning, Golombek takes issue with the usual explanation that geometric forms were used because of the Islamic prohibition on figurate (human and animal) forms. In her opinion the latticeworks represent a cosmic order that is a hallmark of Islam. For one thing, latticeworks appear in the remains of private residences along with figurative works. In view of the prohibition ignored, the cosmic order obviously appealed to the residents.

It was not altogether correct of me to say that computers have no role to play in the production of latticeworks. Although it may be difficult to program a computer to make the kind of intuitive choices that lead to beautiful patterns, it would be less difficult to write what might be called a computer-aided design program. Such a program would lay out a grid of points according to human choice. The human user might also select the size of the circles, the number of points and so on. When the user of such a program indicates what connections to try, the entire screen would fill with the implication of his or her choice. In short, much of the tedium could be removed. The latticework would emerge, one hopes, from the printer.



Two latticeworks in skeletal form

In February the blind watchmaker presided over a column about biomorphs, the creation of Richard Dawkins, well-known author of *The Selfish Gene*. Strictly speaking, biomorphs are created not by Dawkins but by a program written by him. The screen of a Macintosh computer is divided into nine squares. The center square contains a treelike structure and the eight flanking squares contain variations of it. The computer operator selects one of the variants, which then takes center stage and spawns variants of its own.

Evolution directed by humans results not in trees but in forms that are sometimes bizarre and that may resemble anything from an insect to an airplane. In his book *The Blind Watchmaker* Dawkins employs the program to make the point that even if evolution proceeds in small steps, it nonetheless can produce remarkable creatures. Still, it bothers him that the evolution in his program is not exactly blind. How, he wondered in the book, can form emerge from the interaction of a species with its environment?

A conference on the subject of artificial life held at the Los Alamos National Laboratory last September brought a number of scientists and amateurs claiming emergent evolutionary features for their organisms. None satisfied Dawkins, who was at the conference. More recent potential claimants were brought to my attention by John Mitterer of Brock University in Ontario. They are David C. Plaut of the department of psychology at the University of Rochester and David A. Taylor of San Mateo, Calif. Examined closely, the Plaut-Taylor study bears a generic resemblance to many of the projects reported at Los Alamos: a simple creature endowed with the ability to find food and avoid enemies under the direction of a neural network exhibits certain behavior. Creatures that fail to survive will not pass on the maladaptive brains to their progeny. But the creatures were given the nets to begin with. Dawkins would like to see the networks themselves emerge from some more primitive structure under the pressures of a competitive environment.

David M. Chess, a researcher at the IBM Thomas J. Watson Research Center, feels that simulating the evolution of behavior is nonetheless more rewarding. For example, the bendosaurus I described could eat the current model of spikophyte only if the horizontal stretch of its neck enabled it to reach the succulent leaf cluster. How well, wonders Chess, will the bendosaurus flee predators with such a neck? Will it be prone to high blood pressure, neck injuries and so on?

Another IBM employee, Donald E. Curtis of Kennesaw, Ga., argues that more serious attention should be paid to scientific creationism. The proponents of such a theory accept microevolution: stepwise change within a species. But, contends Curtis, "no experimental or fossil evidence exists for macroevolution: fundamental change from one species to another."

FURTHER READING

SYMMETRY. Hermann Weyl. Princeton University Press, 1952.

ARABIC GEOMETRICAL PATTERN AND DESIGN. J. Bourgoin. Dover Publications, Inc., 1973.

ART OF ISLAM: LANGUAGE AND MEANING. Titus Burckhardt. World of Islam Festival Publishing Company Ltd., 1976.

BOOKS

Radon's daughters, three gentlemen of Treviso, Kudzu cloth, polarizing rainbow and sun dog

by Philip Morrison

Ionizing Radiation Exposure of the Population of the United States, NCRP Report No. 93, National Council on Radiation Protection and Measurements. NCRP Publications, 7910 Woodmont Avenue, Suite 1016, Bethesda, MD., 20814 (paperbound, $15). **Indoor Radon and its Hazards,** edited by David Bodansky, Maurice A. Robkin and David R. Stadler. University of Washington Press ($20; paperbound, $9.95). **Radon: A Homeowner's Guide to Detection and Control,** by Bernard L. Cohen, with others. Consumers Union of United States, Inc., Mount Vernon, N.Y. 10553 (paperbound, $10).

The elimination of all motor-vehicle accidents would add just one year to the three score and 11 that actuaries allot to an average American male. Twenty years of participation in a daring profession, say that of an auto racer or a performer on the high wire, is considerably less risky than everyday automobile traffic: those arresting but transient performances cost in the mean only three or four months. The usually long-delayed death from lung cancer, whatever its cause, exacts from us all a shared public cost in expectancy of about 10 months. Smoking a daily pack or two of cigarettes demands a payment of some seven long years in life's coin.)

Long exposure to ionizing radiation is a peculiarly hidden risk to all living things. The imperceptible hazard, as old as life, was first recognized in this century. Its cellular consequences set an unceasing gamble for each of us. One loses that gamble, often after 30 years' delay, mainly to the chance initiation of an eventually fatal cancer that began with the alteration of one or a few multiplying cells. What we know from bitter experience are the dreadful statistics of small groups of people exposed to an intense dosage of radiation: radium pioneers, young women who delicately painted radium-luminous dials, using mouth-tipped brushes, and the survivors of the two A-bomb attacks. Among such individuals a death rate of 50 percent after 10 years of exposure is not unexpected. What we want to know are the effects of a much diluted dosage administered to an entire population. That is the life-and-death task of radiation epidemiology.

Exposure can to some degree be controlled; although we cannot modify the atomic event, we can reduce our risks. The National Council on Radiation Protection, whose report is the first book listed, was formally chartered by Congress almost 25 years ago. Dozens of its reports have expertly treated specific topics in the discipline, from dental-X-ray protection to radiation alarms. This compact study issued in the fall of 1987 is the latest comprehensive evaluation of "exposure of the U.S. population from all sources," a summary of technical contributions by groups including some 50 specialists.

Everyone knows something of the radiation we receive. The star-sent cosmic rays and gamma rays from rocks contribute equally. The normal internal radioactive content of the body, mostly potassium, adds somewhat more. Manmade radiation comes in the main from diagnostic X rays. These amount to about half of the contribution from the three natural sources named. A long table lists consumer products that produce radiation, those involving many people as well as those affecting few. The overall contribution is small, a fifth of the amount ascribed to medical procedures. Television receivers irradiate everyone, although only a little, much less than we take from the radioactive minerals within brick and concrete, or from the variable content of our domestic water supplies. Smoke detectors and luminous watches, thoriated gas mantles, spectacle lenses and welding rods are all on the list.

Stray radiation from the entire nuclear-fuel cycle, from the mine to spent fuel storage, when spread over all of us at risk, adds less than 1 percent of the dosage from the normal internal radioactivity of the human body. Occupational exposures and the residual weapons-test fallout amount to several times more than that. Even the major burst of fallout from Chernobyl added an increment of only one extra year of background-level exposure to the amount of radiation that would in any case have been received in 50 years by the 100 or 200 million people in the U.S.S.R. and Europe who may have been in the path of Chernobyl's exhalation.

Nothing unexpected in all of that: a variety of earlier studies gave similar numbers, perhaps less coherently. Something is new, however, plainly visible only within the past decade. More than half of the radiation to which the American public is exposed is now ascribed to the air we breathe at home [see "Controlling Indoor Air Pollution," by Anthony V. Nero, Jr.; May]. The source is natural, the capricious leakage of the radioactive gas radon from the soil, to which it is steadily restored by the geologically slow decay of ubiquitous traces of uranium in rocks of every kind. Of course, radon is out in the open air as well. There, caught by the winds, it is diluted below much notice, tenfold lower than the concentration found in a typical living room. Ever since our species began to dwell in caves, we have taken in the active products of radon decay along with the quiet air of almost any shelter. It was not until 1984 that the NCRP officially signaled the dangers of indoor radon.

The element radon itself is not a direct source. A chemically inert gas that takes days to decay, it enters and leaves the lungs unchanged as we inhale and exhale. Meanwhile it steadily produces radioactive daughter atoms much more short-lived than itself, single ions around which air molecules cluster by the dozen. Most ions affix themselves to good-size dust particles that the lungs can exclude. A fraction remain almost free. Within the air passages, however, deep down in the finer twigs of that tree of life, some of the active products do decay, irradiating and damaging nearby surface cells of the bronchial epithelium. The dose is mostly a sum of such random local events in tiny cavities of the lungs, but it can be high.

The mischief that radon's daughters inflict was first written in the annals of mining. The lodes of the Ore Mountains in southern Germany are rich in a variety of metallic elements, including uranium. A hundred years ago the pathologists realized that the common mine sickness that stopped the breath of as many as a fourth of the veteran miners was not the same as black lung. The newly recognized illness was a cancer. Radon was not generally accepted as the cause until the 1960's. Now it is the hard-rock miners, seeking uranium from Czechoslovakia to Colorado, digging for fluorspar in Newfoundland (where the seeping mine waters are rich with radon) or laboring in Swedish iron mines and Chinese tin mines, whose life and death form the epidemiological basis of our concern.

The work histories of tens of thousands of men among whom many hundreds died from lung cancer allow a complex mapping of radon exposure against its malignant outcome. Deadly mines must be transformed by calculation into benign homes where radon risk is mostly very dilutely spread. The aerosol physics of dusty, smoky pits must be related to the wandering of active atoms within the clean air of living rooms. Just what is the radon content of a house? The clearest answer is that it varies widely. Soil geology, construction, ventilation and leaks can skew exposure so that a percent or two of all houses may well place residents at high risk. Most houses offer a much smaller reduction in life expectancy, close to the month of loss ascribed to drownings or to fire (both hazards against which there is traditionally firm public action).

As yet no clear finding exists of an excess of lung cancers ascribable to indoor radon. What is the highest domestic exposure rate? Hundreds of thousands of houses have been measured in the U.S. (mostly at the homeowner's initiative) and random sampling has been conducted in Sweden and Canada. The highest exposure found is equivalent to that of a classical uranium mine in Saxony: 100 times higher than the present legal limit for U.S. uranium mining.

There is not much doubt of a deadly radon signal too weak to read in the noisy public record of deaths from lung cancer. It is estimated that about 10,000 deaths a year from indoor radon are camouflaged among the 140,000 Americans who die each year from that same disease, mostly related to cigarette smoking. Sixty years ago the same radon was in our rooms, which to be sure may have been somewhat draftier and less radon-filled. In those days, however, fewer men and many fewer women had inhaled cigarette smoke for 20 years or more; the rapidly rising rate of lung cancer was still below today's by a factor of 15. Leukemia, alone among major cancer types, shares with lung cancer an incidence that has risen considerably during the past 50 or 60 years, although less rapidly by a factor of four.

These three brief books on radon differ widely in content and level. The NCRP report is a rather technical summary of a wide span of topics including indoor radon. The University of Washington book is a collaborative effort by seven faculty members, who offer a many-sided introduction to the entire complex question of indoor radon. Their discussion ranges from measurement techniques to models of the lung. It is a first-rate entry to an entire important field of applied biophysics.

The Consumers Union book, by a University of Pittsburgh physicist, is brightly written and is just technical enough to convey information accurately. The approach is suitable to a book aimed at direct personal action: there are long tables of the radon content found in 35,000 houses listed by county and by zip code. Geology has a large-scale effect; your neighbors' dose is correlated with yours, but the variation can be high from door to door even on the same subsoil. "What about *my* house?" If you want to know, only measurement can tell. The book describes and rates the simple radon-detection devices on the market, tells how to measure dose and explains how to safeguard your room air from radon.

Low-radiation hazards evoke responses from governments and the public that are strangely inconsistent. The Environmental Protection Agency intends that a national nuclear-waste repository meet the standard of causing less than a tenth of a fatality per year, but it has taken no positive steps at all against indoor radon. Its remedial recommendations would leave hidden indoor radon 1,000 times more lethal than the visible nuclear waste. State involvement has been more active, impelled by geologic truths. Yet, with laudable exceptions, "neither environmentalists nor the nuclear industry seem to like talking" about radon. The one side can find no one to blame, and it fears diminution of public concern over nuclear power. The other side dislikes publicity about radiation effects "in any form," as one English commentator observes. The problem is worldwide, of course; governments have yet to take widespread measures of survey or remedy. Perhaps the present elaborate extrapolation from mine to house will be found too conservative; perhaps there really are unknown mechanisms of cellular healing that turn out to defend us at low dose levels. We may hope so, for they would save many lives. But certainly—by two orders of magnitude more—we need to keep the nuclear peace.

CAPITALISM AND ARITHMETIC: THE NEW MATH OF THE 15TH CENTURY, by Frank J. Swetz, including the full text of the *Treviso Arithmetic* of 1478, translated by David Eugene Smith. Open Court Publishing Company ($33.95; paperbound, $16.95).

Already mature in this first of all printed books of mathematics in the West are the kind of arithmetic problems that still call forth schoolday groans: "Three men, Tomasso, Domenego, and Nicolo, entered into partnership. Tomasso put in 760 ducats on the first day of January, 1472.... And on the first day of January, 1475, they found that they had gained 3168 ducats, 13 grossi and 1/2. Required is the share of each, so that no one shall be cheated." The solution is worked out neatly—if wrongly, as one of the many problem conditions is misstated. Partner Nicolo is found to gain 1,173 ducats, 22 grossi and 17 pizoli, all checked. (Business was good; they earned some 40 percent per year on capital invested.)

We are at the flourishing of the Renaissance, printing is lively in the cradle. This is the still expansive Venetian Republic. Even the town of Treviso, 15 miles away, about a day's travel from the capital city, supports half a dozen printers. No other book in mathematics has reached print, not even famous Euclid; the academics would have to wait a few years more for the first printed *Elements*. Perhaps the scholars were not as attractive a market as the many young merchant's sons for whom the affairs of imagined Tomasso and his partners were so meticulously set out. Geometry, moreover, demands all those costly labeled diagrams; even the numerical examples were not easy for early printers. Here they made do with the letter i for the numeral 1 to save time or type. (The inexpensive book at hand is not a facsimile, although the edition is both typographically attractive and evocative of the rare original.)

This sprightly piece of scholarship presents in 300 readable pages the text of the anonymous old practica itself, a documented analysis of its operations and techniques in the context of the applied mathematics of the day. A concise and colorful perspective view of the thriving mercantile society that this arithmetic text so artlessly reflects, ducat by golden ducat, emerges from the book. The text, originally written in the Venetian dialect, was put into clear-running English before World War I by David Smith, a well-known Columbia University historian and teacher of mathematics. Smith never published it in full. It was noticed among his papers a few years back and given this delightful form by Dr. Swetz, a scholar at the Harrisburg campus of the Pennsylvania State University.

Untitled, anonymous, the *Arithmetic* was a working textbook for young students of the new commercial arithmetic. It was the first to see print of a large number of books that expounded the new math in many European vernaculars. They were called algorisms, after the Baghdad savant Muhammad al-Khwarizmi, whose ninth-century arithmetic on the new numbers had been among the earliest to gain Latin translation. Also in use were Latin theoretical arithmetics of medieval provenance, based on Roman numeration.

Other books were at this very time in embittered competition with the algorisms; they were also practical in intent but dealt with the venerable techniques of the counting table, how to use its movable tokens and ruled lines. By 1400 the new numbers written in ink on paper had won their way in the south but not in northern commerce. Even the University of Padua had required as late as 1348 that book prices be written out in Roman numerals, "not by figures but by clear letters." It was thought the strange new digits could be more easily falsified.

It is no mere chance that many words of modernity—such terms as bank, discount, net, credit, opera, science, even ghetto—are Italian in origin. All the action was there. Prosperous and up-to-date German families from the 14th century on had sent their sons to northern Italy to learn the new ways of ciphering and the double-entry bookkeeping already in wide use. The apprentice merchants studied the texts of their masters, who ran the special reckoning schools. Such a master was our unnamed author of the Treviso arithmetic; in all, we know of 30 practica like his printed by century's end, about half of them still in Latin.

Those people who would become literate began study in a basic school where the three R's were taught. On that foundation a good student with support might go on to grammar school, where he would become proficient in Latin and the Roman authors. He might then make his way through the university to become a cleric, a lawyer or a physician. One such career would be that of Nicholaus Copernicus, who came from Baltic shores to study at Bologna and Padua; he was still a child when the Treviso arithmetic came out. The strongest educational alternative to grammar school for middle-class youths was the reckoning school. Mercantile education beyond that vocational level meant wide travel and the mastery of foreign languages. The 200-year-old model of the Venetian merchant Messer Marco Polo set the pattern.

This arithmetic is broadly familiar. In our time we still place it with the force of law before all our pupils. The 10 digits that served the merchants of Treviso in 1478 are the ones we use today. There are differences. No number word beyond *million* appears; *billion* and the rest entered a few years later. The algorithms for adding and subtracting are still familiar, although curiously no exercises on adding long columns are set. Multiplication tables up to 10 times 10 appear early in the book, although that powerful operation is clearly less than routine. Long and short division are treated with special care and taste. Several clever schemes are described that lay the work out by the column, the cross, the chessboard. As many as eight algorithms were current. Saving on expensive paper seems one criterion for choice. Checks of computation are regularly taught, including the trick of casting out 9's. Although units of money and weight are much discussed, the tables of equivalents and exchange ratios are not as elaborate as those in other works of the time. The Rule of Three for ratios is a big set piece. Fractions are assumed, but they are not much elaborated. The systematic use of decimals lay a century ahead, although percentage and the utility of place notation are already well appreciated.

As one might expect, the otherwise intensely practical book begins with some neo-Aristotelian philosophy of number (perhaps misunderstood) and closes with a few fancy problems on rates: "If 17 men build 2 houses in 9 days, how many days will it take 20 men to build 5 houses?" These are not in fact relevant to commerce; they are, "in a sense, a boast of the power of mathematics." Students, amaze and puzzle your friends! There is not the least hint of astrology for the young businessman; instead there is a final page or two on calendrical computations that fixed the holidays.

The goods most in trade in these problems are crimson cloth, raw silk and French wool, beeswax for candles, and spicy ginger, pepper, saffron and cinnamon. Relative prices are not grossly discordant compared with those of today—not even the price of sugar, then perhaps grown on Venetian-held Cyprus. On the 10th day of December, 1478, the Trevisan *maestro d'abbaco* closed his book with an affectionate farewell to his dear friends, the diligent students. "I promise you the same gratifying usefulness." That mind and attitude remain familiar 500 years and an ocean away from the Rialto, embodied in our Harvard M.B.A.'s.

A WEAVER'S GARDEN, by Rita Buchanan. Interweave Press, 306 North Washington Avenue, Loveland, Colo. 80537 ($16.95).

"If you have five hundred acres in Texas and want to grow a cotton crop, you can find plenty of information on how to go about it...even...how to recognize the effects of lightning strikes on your fields...but it's more than you need to know if you just want...a little cotton patch in your backyard. So, translated...from volumes of agricultural handbooks, and filtered through my own experience, here is how to grow some Upland cotton." In evidence there is a drawing of the plant, a photograph to show the flowers (which resemble yellow hollyhocks) and images of intricate white cotton cloths woven by this author-craftswoman. Add to cotton 18 or so other plants that yield fibers from seed, stem or leaf; among them are blue-eyed flax, the pesty kudzu vine (woven into a pliant and lustrous golden cloth) and stately yucca.

That is only the beginning of this firsthand, participatory and charming economic botany. Here are complex indigo and other dye plants, aromatic plants such as tansy and lavender that scent textiles and repel insects, and plants that provide weaver's tools such as the fierce burrs of the European fuller's teasel (the spines of the common teasel, genetically distinct, are too weak and straight). Gardeners and weavers will find this a practical guide to a many-faceted technology that is as old as agriculture.

If the book offered no more, it would be of high value to a happy not-so-few. Because the book's breadth of context transcends the empirical, it also informs and delights the general scientific reader, even one who has little energy for peat pot and trowel. Rita Buchanan concisely includes enough of modern plant science, chemistry and economic geography and history to explain and unite what first came from witty and skillful practice.

Cotton in its homeland, India, like the pima cotton of pre-Columbian Arizona, has smooth seeds, from which the fibers are easily detached. A rolling in, and later a roller gin with hand- or watermill-driven wood rollers, was the ancient solution. In our southland, slave labor could tediously separate by hand the tight-sticking lint of green-seeded Upland cotton. What Eli Whitney did was to invent the hook or spike gin, a mechanism able to pull the lint from the seeds of the major cotton variety at hand.

Inert cellulose, a simple polymer, has always been hard to dye. The protein fibers, wool and silk, take colors much more easily: their chains offer plenty of reactive sites. Reasonably fast colors on plant fibers, direct-dyed without prior chemical treatment, come from only a few sources: the pigment of the seeds of the tropical American shrub annatto, the yellows and reds from safflower, and above all that remarkable indigo. Annatto never became a major dyestuff. Yet annatto is common in the supermarket, for its colorant is believed nontoxic and is held well in fats and oils. In fact, most packaged yellow foodstuffs contain annatto (consult the label). Madder roots—the perennial is hardy and easy to grow, although not from seed—are the source of a strong red dye. Called alizarin, the dye has been economically important since medieval times. During the 19th century the prized crop was abandoned worldwide within a few years of the discovery of the process for synthesizing alizarin. Heinrich Caro in Germany and William Perkins in England share honors for the achievement; they independently sought patents for it on two successive days in June, 1869. Coal-tar chemistry had risen to power.

Spinning is the first subtle step from harvest to cloth, and it demands the most hand labor. Dexterous people can spin by rolling the fibers between the palms or between palm and thigh. But you must wind the yarn in a ball or onto a stick. When the stick itself is used to impart twist to the thread, it is a spindle. (Plant fibers come with a built-in twist, its direction constant for a given species.) All over the world since Neolithic days there were spindles; at first they were only smooth, straight twigs. Soon the skillful added a round whorl to increase inertia and a hook or notch to catch the yarn. Some cultures stand or walk as they drop the whirling spindle; some sit and roll the spindle in hands or against a thigh; some use a dish as bearing for the rotating end of the spindle. The lore from India recounts that one ounce of cotton could be spun into five miles of thread through 60 long spinster-days of work. The woven cloth was transparent. We cannot do better today.

The annotated bibliographies supplied here chapter by chapter are a pleasure; they include the grand technical monographs, old and new, the best popular books and the vigorous periodical literature and specialized publications of the world of the crafts. (The last deserve to be more widely known and used by readers who are not yet doers.) One color photograph samples this author's mastery; her pile of wound yarns glows like a heap of jewels. These are "a dye garden harvest: sixty colors of yarn dyed on a September weekend in Virginia."

POLARIZED LIGHT IN NATURE, by G. P. Können. Translated by G. A. Beerling. Cambridge University Press ($37.50).

The silvery mirage on the hot road surface looks just like the shimmering surface of an illusory pool of water, one that forever recedes as we drive toward it. Viewed through a small square of Polaroid sheet, mirages are unchanged in brightness however we turn the filter. A real pool does not act that way; its grazing reflections are much polarized. The gradual upward refractive bend of the rays from a hot layer of air that forms the deceptive mirage cannot polarize the light as abrupt surface reflection nearly always does. Polarization-blind, unaided human eyes miss the telltale difference, but birds and bees can catch it because they are sensitive to the effect. It is such barely hidden sights that appear here in 100 variations, copiously documented by comparative color photographs and explanatory diagrams.

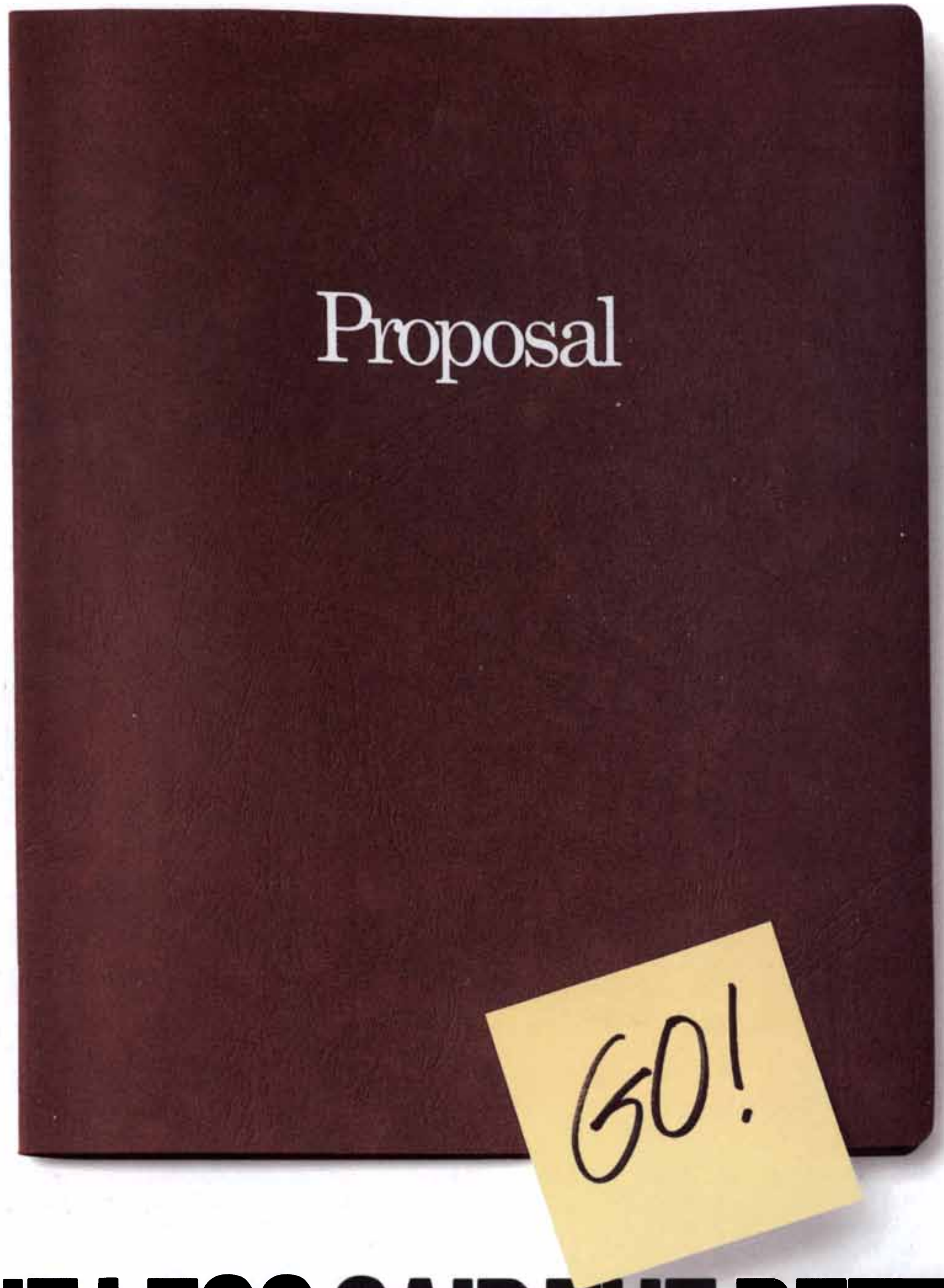
The author is a Netherlands meteorologist (his Dutch uncle was the translator) who loves the look of the outdoors. He has watched through Polaroid the sun and sky, the moon, darkened city streets, the depth of a swimming hole, rainbow and sun dog, tempered windshields and picnic tables, specimens in a mineralogical museum, just about everything. Explicitly extending a beginning made in M. Minnaert's marvelous classic on light and color in the open air, Können has sojourned for years with the single aid of a polarizing sheet to gather these devoted observations and artful photographs, some done by his friends.

The visual landscape depends on sources of light; on clear days they are the predominant sun and the blue sky. The color and contrast that we see depend on what is illuminated, from the vault of the sky to the sunset horizon, the sea spray, a glossy fender or newsprint. The polarized landscape arises in the same way. Like color, it too depends on what the light has illumined. Polarized light sources are rare in nature. But refraction, reflection and scattering all leave their polarizing mark, determined by the physics of substance and the geometry of view. Polarized views reward subtlety: avert your gaze.

Ice crystals and water droplets, leaves and grass, metals and glass and plastics—all have specific polarization properties, easy to analyze qualitatively after a little guidance. Every fluorescent tube generates unpolarized light. Yet each surface of the enclosing glass cylinder refracts the light and marks it with polarization; again the effect is best caught from an oblique view.

There is a fine turn-of-the-century rule for roughened surfaces with little gloss. A Russian experimenter, N. Umov, explained that dark, rough surfaces polarize more strongly than light ones: brick polarizes better than paper, asphalt better than snow. The dark surface does not return much multiply scattered light to the eye; it is the repeated scatterings from the irregularities on a rough surface of light-colored, nonabsorbing material that bring back so much light. That random walk mixes the vibration directions, so that the plentiful reflection has little directional preference. In contrast, light can scatter only once or twice within a roughened but absorbing surface.

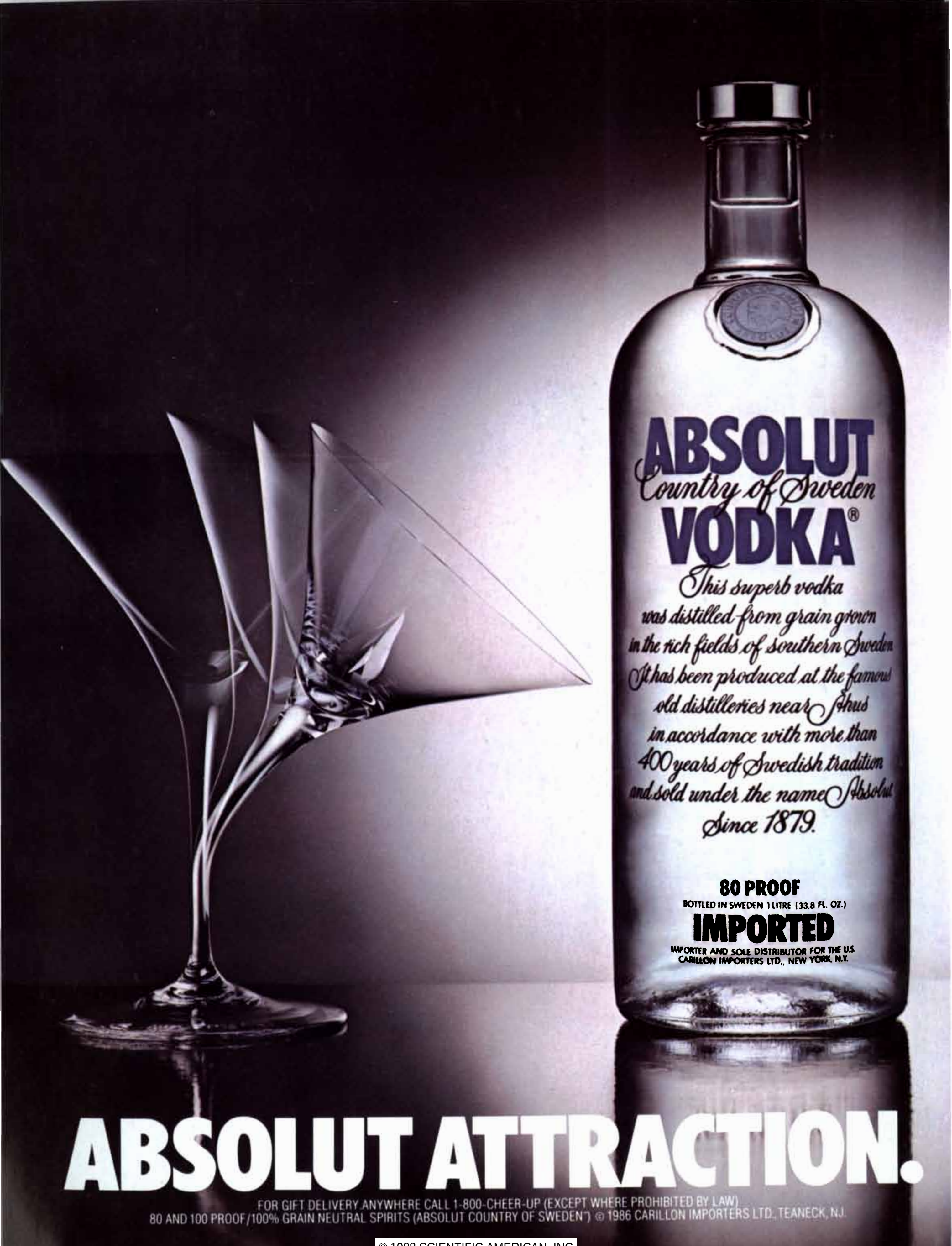
The world of light and color takes on another dimension, disclosed by a couple of square inches of cleverly stretched plastic. Surely that earns this admirable extension a place on the shelf next to Minnaert and the accounts by unpolarizing witnesses of rainbow nature. There is more left to do. The light emitted from glowing matter, say molten metal, is visibly polarized by refraction as it leaves the surface if there is no disturbing film of oxides. A photograph here shows the effect in streaming molten iron, oxide layer unformed. Who will check this out as the lava flows past, luminous from Pelé's kitchen?



THE LESS SAID THE BETTER.

When People Count On You, Count On Post-it™ Notes.

BRAND


ABSOLUT
Country of Sweden
VODKA®
This superb vodka
was distilled from grain grown
in the rich fields of southern Sweden.
It has been produced at the famous
old distilleries near Åhus
in accordance with more than
400 years of Swedish tradition
and sold under the name Absolut
Since 1879.
80 PROOF
BOTTLED IN SWEDEN 1 LITRE (33,8 FL. OZ.)
IMPORTED
IMPORTER AND SOLE DISTRIBUTOR FOR THE U.S.
CARILLON IMPORTERS LTD., NEW YORK, N.Y.
ABSOLUT ATTRACTION.
FOR GIFT DELIVERY ANYWHERE CALL 1-800-CHEER-UP (EXCEPT WHERE PROHIBITED BY LAW)
80 AND 100 PROOF/100% GRAIN NEUTRAL SPIRITS (ABSOLUT COUNTRY OF SWEDEN™) © 1986 CARILLON IMPORTERS LTD., TEANECK, N.J.