SCIENTIFIC AMERICAN

JULY 1990
$2.95

TRENDS IN AEROSPACE: *Can space be commercialized?*

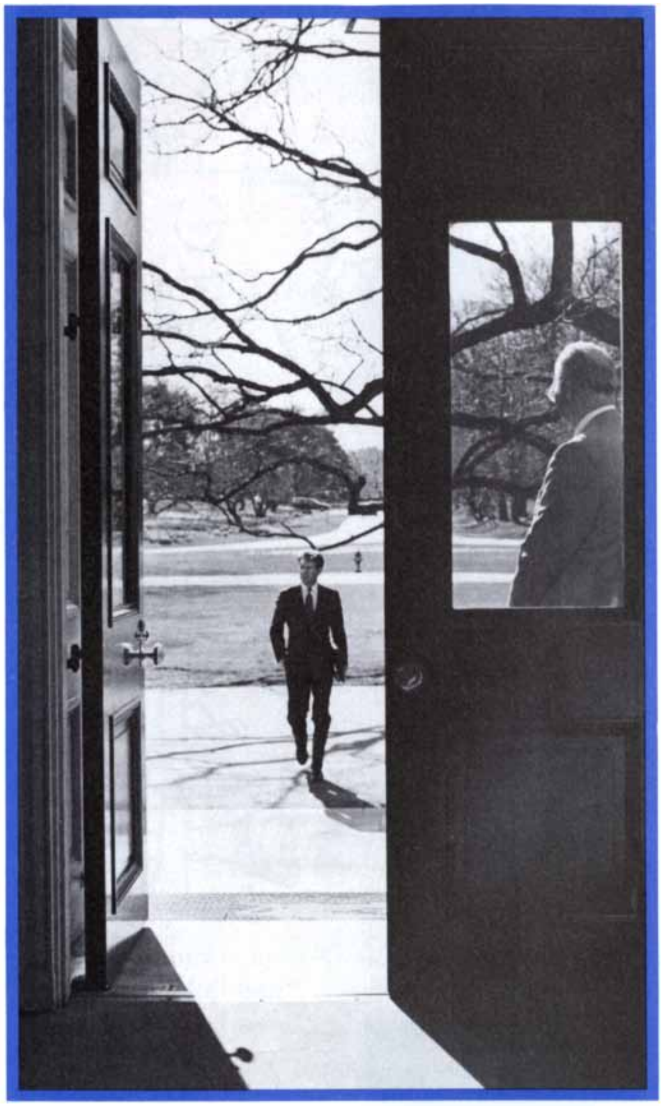
The immune assault that causes diabetes.

Why we should take immediate steps to slow global warming.

***Fireworks** have marked celebrations for centuries.*
Behind the flashes and bangs lies the science of pyrotechnics.

It's not that we're cramped for space at ROLM. Very simply, we believe that every moment our field force spends in our office is a moment that they could be spending with you. Learning your business. Asking questions. Finding out what they need to know to make your telecommunications system, and your company's operations, as effective as possible. From combining voice and data facilities into one integrated system and explaining to each employee exactly how to use it. To improving your call center productivity through new applications like CallPath.™

You see, at ROLM we make it our business to know yours. And that's something we can't do sitting behind a desk. Or standing around the coffee machine. Which means as a ROLM customer, you'll be seeing a lot more of our people than we do. And that's fine with us. Because we'll always know just where to find them. And so will you. For more information, please call us at 1-800-624-8999 extension 235.



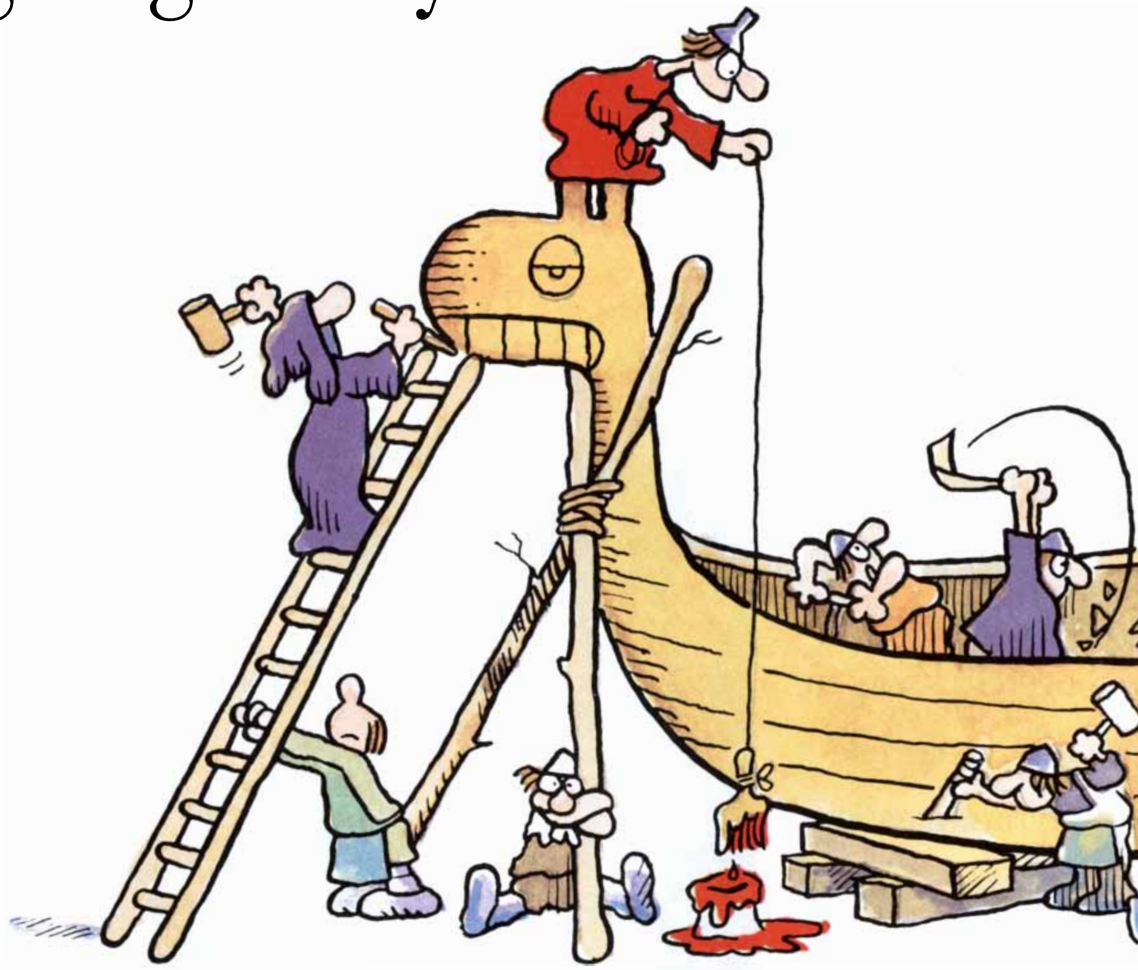
8:56 a.m.: the ROLM representative reporting to work at READER'S DIGEST World Headquarters.

ROLM **WE ASK BETTER QUESTIONS. YOU GET BETTER ANSWERS.**

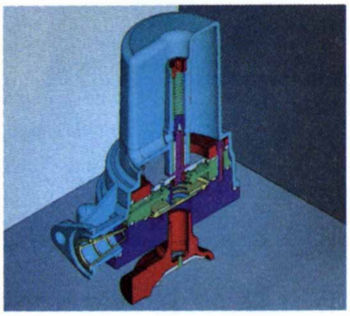
An IBM and Siemens Company

The IBM RISC System/

Designing on any other workstation



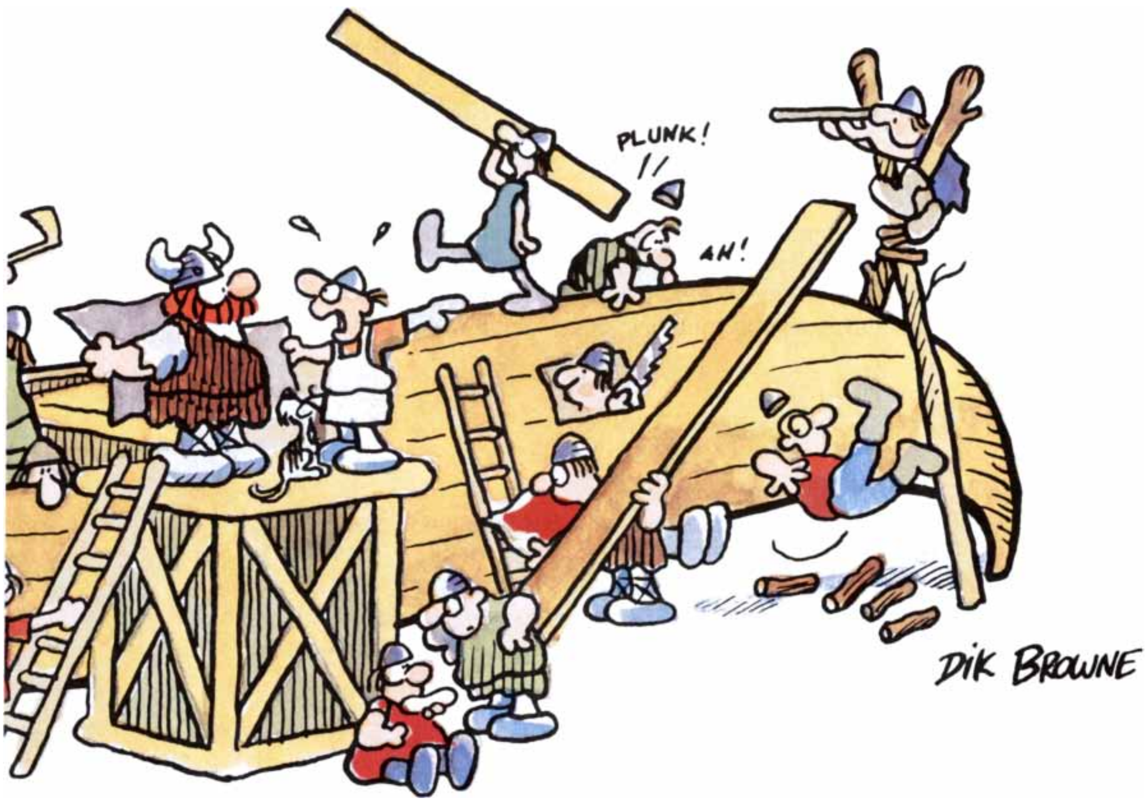
Whatever you're creating, you'll sail into a whole new age with any of the four POWERstations in the RISC System/6000 family. Because POWER (Performance Optimization With Enhanced RISC) processing can give you performance you've probably only dreamed about:



up to four instructions per machine cycle, 42 MIPS and 13 MFLOPS. Suddenly, complex designs don't take eons anymore.

The four RISC System/6000 POWERstations feature a range of graphics processors from grayscale to Supergraphics to satisfy any graphics demand. Great news for Power Seekers working on animation, scientific visualization, medical imaging and engineering solutions like CADAM,™ CAEDS™ and CATIA.™ And for electrical design automation, there's IBM's all new CBDS™ and an arsenal of over 60 EDA appli-

6000™ family.
will seem downright primitive.



cations from more than a dozen vendors.

With every POWERstation, you can get an almost unimaginable palette of 16 million colors, which gives you 3D images so realistic, they fairly leap off the screen, with super sharp resolution of 1,280x1,024 pixels. And when it's time to call in the heavy artillery, the POWERstation 730 draws nearly one million 3D vectors per second. Like all POWERstations, it can come complete with its own graphics processor, freeing the POWER processor to rapidly create and analyze your designs. All at prices that won't sink anybody's budget.

So if you're tired of paddling upstream with yesterday's performance, call your IBM marketing representative or Business Partner to find out more about the RISC System/6000 family. For literature, call 1 800 IBM-6676, ext. 991.

Civilization never looked so good.

For the Power Seeker.

IBM®

SCIENTIFIC AMERICAN

July 1990 Volume 263 Number 1

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 rates: one year $36, two years $64, three years $88 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. only 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51593.

DEPARTMENTS

"TAKE A LOOK INSIDE OUR
THAT'S WHERE THE BIG CHA

You told us you'd consider buying a GM car or truck if we changed more than just the looks. And that's exactly what we've done.

We've listened to what you said about quality. We've heard your demands for reliable, durable cars and trucks. And we've responded with new vehicles, with new technology, with new production methods in new and refurbished plants.

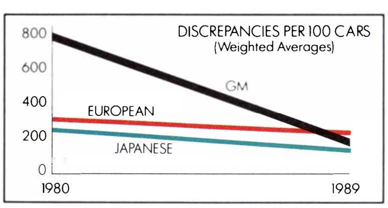
THE HEART OF GM QUALITY

We've introduced 77 new models since 1985. Almost all the cars and half the trucks we make. All re-designed and engineered to dramatically improve quality.

The looks are new. But that's not all. The big improvements are in the engines, transmissions, and electrical systems—the heart of any car or truck. And we're continuing to improve.

MEASURABLY BETTER

GM builds cars you can count on for the long term. No U.S. carmaker has done that job better in the last five years. In fact one measure of dependability, the J. D. Power dependability study of 1985 models, ranks GM highest in vehicle dependability among all American manufacturers.[1]



Since we built those cars, independent measures have confirmed that our quality has improved substantially. The differences today between GM and the best imports are very small indeed.

According to one quality study by Harbour & Associates, we're better than the average European make, and the difference between GM and the average Japanese import is less than one-half of one discrepancy per car. Our own tracking confirms GM's significant improvement.

GREATER DURABILITY

According to R. L. Polk Registrations, a higher percentage of 10-year-old GM cars are still on the road than '79s from Ford, Chrysler, Mazda, Nissan, Toyota, or Honda.

RELIABLE ENGINES

GM's engines are more dependable than those of all other domestic carmakers. The 2.3-liter Quad 4 is as problem-free as 2-liter engines from Toyota or Honda.

Our 3800 V-6 is at the top in engine quality among engines from all makers, foreign or domestic. That's the finding of the most comprehensive customer-based

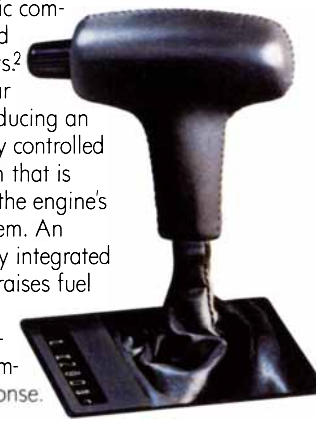
survey in the auto industry. And Cadillac owners report fewer engine problems the first year than Toyota or Mercedes-Benz owners—a tribute to the 4.5-liter V-8 in Cadillacs.

BETTER-BUILT AIR CONDITIONING

The reliability of major mechanical systems is important to our customers. GM's new cars have better-built air conditioning systems than cars from Honda, Nissan, or Mazda.

PROBLEM-FREE TRANSMISSIONS

Any car or truck is only as reliable as its transmission. And our automatic transmissions are more problem-free than those of any domestic competitor—and most imports.[2]



This year GM is introducing an electronically controlled transmission that is linked with the engine's control system. An electronically integrated powertrain raises fuel efficiency, lowers emissions, and improves response.

TECHNOLOGY THAT LASTS

GM is the only domestic manufacturer to design and build multivalve engines. And we build them to last. Our 16-valve Quad 4 has already run the equivalent of 100,000 miles at 100 mph without stopping under test conditions.

On the endurance track, a full-size Chevrolet sport truck equipped with GM's

All of GM going all out for you.

NEW CARS AND TRUCKS.
NGES IN GM QUALITY ARE."

350 cu. in. V-8 captured the Hulman Trophy for sustaining 100 mph for 24 hours at Indianapolis.

MOST TROUBLE-FREE U.S. CAR LINE

A study of 1990 models done by an independent market research company ranks Buick as the top domestic nameplate, based on problems per car in the first 90 days of ownership. LeSabre was the top U.S. model, with Electra and Skylark in the top ten.

Cadillac Eldorado and the Oldsmobile 88 and Cutlass Ciera also made the list, giving GM six of the study's ten best-built American cars.

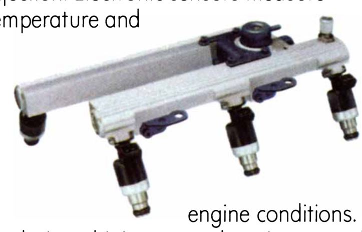
CORROSION PROTECTION

We offer the highest level of exterior-galvanized corrosion protection in the world. In fact, we're the only high-volume carmaker extensively using two-sided galvanized materials to protect against rust.

That's how we can stretch our limited warranty's corrosion coverage to six years or 100,000 miles minimum.[3]

99.9% CERTAIN STARTS

Virtually every engine in every GM car or light truck features electronic fuel injection. Electronic sensors measure temperature and engine conditions.



Re-designed injectors and engine control computers meter the right amount of fuel for sure starts.

These GM cars start the first time, hot or cold, in any weather, regardless of where you live in the United States.

HIGHLY SATISFIED CUSTOMERS

GM's Cadillac division has been the highest-ranked domestic nameplate in customer satisfaction for the past three years.[4]

The customers of every other GM division are highly satisfied as well. After six months of ownership, at least 95% of all Chevrolet, Pontiac, Oldsmobile, Buick, Cadillac, or GMC Truck owners would recommend a vehicle from that GM division to a friend.

ENDURING VALUE

Over the last 10 years, cars built by General Motors have kept more of their original value, on the average, than cars made by any other U.S. manufacturer. Chevy and GMC trucks retain more of their original value than trucks sold in the U.S. by any other manufacturer in the world—foreign or domestic.[5]

WE COVER EVERY PART

We back our vehicles the way we build them. Our Bumper-to-Bumper Plus limited warranty covers every part of every GM car or light truck. For three years or 50,000 miles.[6]

GM
BUMPER TO BUMPER PLUS
3 YEAR/50,000 MILE LIMITED WARRANTY

WE CARE ABOUT YOU

We care what you think about GM quality. People throughout General Motors—the GM Quality Network—have taken a long, hard look at the cars and trucks we build. And the way we build them.

Today there is a new pride at GM. A new commitment to quality. A clear focus on our customers and on their needs. A dedication to continually improving our vehicles. Year after year. Until every model is as good as the best in the world.

We've accomplished much already. We invite you to take a good, close look at the results: our new GM cars and trucks, and the new level of General Motors quality.

1: J. D. Power and Associates Vehicle Dependability Index Study℠. In a ranking of the three domestic manufacturers, based on things gone wrong to 4-to-5-year-old 1985 model vehicles in the past 12 months.

2: GM owner-satisfaction survey for first quarter of 1990.

3: See your GM dealer for details of this limited warranty covering corrosion. A deductible applies.

4: J. D. Power and Associates 1987-1989 Car Customer Satisfaction with Product Quality and Dealer Service℠ studies. After one year of ownership.

5: GM corporate study of 3-to-5-year-old used vehicles resold between 1979 and 1990.

6: Tires on GM cars are warranted by their manufacturers. Cadillac's Gold Key Bumper-to-Bumper limited warranty offers coverage for 4 years/50,000 miles, and Cadillac's Allante is backed for 7 years/100,000 miles. See your GM dealer for details of the limited warranty. A deductible applies.

PUTTING QUALITY ON THE ROAD

GM

MARK OF EXCELLENCE

Chevrolet Pontiac Oldsmobile
Buick Cadillac GMC Truck

THE COVER photograph shows a distinctive red, white and blue fireworks display. The science of pyrotechnics has lifted the veil of secrecy surrounding this ancient craft, revealing the rapid reactions that produce the colors, shapes and sounds. In this case, fragile compounds of strontium and copper, created briefly in the explosion, emit red and blue light. Aluminum powder explodes in a brilliant flash; larger particles form white sparks (see "Pyrotechnics," by John A. Conkling, page 96).

THE ILLUSTRATIONS

Cover photograph by © Joe Viesti, Viesti Associates, Inc.

Page	Source
37	AP/Wide World Photos
38	Bernard G. Mendosa, Geophysical Monitoring for Climate Change Division, National Oceanic and Atmospheric Administration
39–43	Johnny Johnson
47–49	Patricia J. Wynne
50	Edward Bell
51	Patricia J. Wynne
52	Guillermo Oliver (*left*), Anders Molven, University of Oregon (*right*)
55	George Retseck
56	George Retseck (*top*), Johnny Johnson (*bottom*)
57–59	George Retseck
60–61	European Laboratory for Particle Physics (CERN)
62	Gabor Kiss
63	William Riley, University of Florida
66	Gabor Kiss
67	Mark A. Atkinson
68	Gabor Kiss
69	Gabor Kiss (*left and right*), Don C. Wiley, Harvard University (*middle*)
70	Dana Burns-Pizer
72–73	Bernard Paris
74–77	Joe Lertola
80–82	National Space Development Agency of Japan
84	Glavkosmos
96–97	Albert E. Zwiazek, Superstock
98	Andrew Christie
99	Tri-State Manufacturing (*top left*), Zambelli Internationale (*top right*), John Gillmoure, The Stock Market (*bottom*)
100	Andrew Christie
101	© Estate of Harold E. Edgerton, courtesy of Palm Press, Inc.
102	The Metropolitan Museum of Art, the Edward W. C. Arnold Collection of New York Prints, Maps and Pictures, bequest of Edward W. C. Arnold, 1954 (*top*), Fireworks by Grucci (*bottom*)
107	Larry Brewer
108	Tom Prentiss
109	Gabor Kiss (*top right*), Tom Prentiss (*bottom left and right*)
110	Joseph R. Newhouse (*top and bottom left*), Gabor Kiss (*bottom right*)
111	Joseph R. Newhouse
112	Johnny Johnson
113	Gabor Kiss
114–117	Johnny Johnson
118	Andrew Christie
119	Edward Bell
120	Russell Ruthen
121	Edward Bell

SCIENTIFIC AMERICAN®

Established 1845

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue
New York, NY 10017
(212) 754-0550

"KEEP IT CLEAN."

Madeleine Marchese
General Manager, Marketing
Texaco Syngas Inc.

"For five years, Texaco's gasification technology lighted 100,000 homes with clean energy from coal. We keep it clean with a gasification process that we invented and perfected. It's a very workable solution to a difficult environmental problem."

Madeleine Marchese is General Manager, Marketing for Texaco Syngas Inc.

"Our coal gasification process has produced over 2.5 billion kilowatt hours of electricity for California residents. That's a record no other coal gasification process has even come close to.

"The Texaco process has not only been proven economically, it exceeds the clean air standards proposed by the Bush Administration. That proposal is for the year 2000. Texaco is ready *now.*

"As a technology, coal gasification is a powerful tool in the fight against acid rain. As an alternate energy source, it makes us more energy self-sufficient. We have a 300 year supply of coal at current consumption rates.

"Texaco people have even devised ingenious methods to expand the technology for the gasification of municipal sludge. Our process is designed not only to get rid of sludge but to *transform* it into usable energy."

Clearly, there is enormous potential for America and its environment. And for the future.

Star of the American Road

TEXACO–WE'VE GOT THE ENERGY.

How to Choose a Scotch to Impress Your Father-in-Law.

You actually like your father-in-law.

When you come over, he's pleased to see you, and not just because you've brought his daughter. He's interested in your work, your garden, even your opinion. He shows you his latest project and solicits your advice (although he ignores it later).

And he pours you a nice glass of Scotch without waiting for you to ask. His brand? Ballantine's Finest. A sophisticated blend of 42 superb single malt Scotch whiskies and the most popular Scotch in Europe.

A real gentleman, your father-in-law.

So the next time you visit, make the right impression by employing the Golden Rule. Along with his daughter, bring him a bottle of Ballantine's Finest.

When you touch glasses, murmuring an appreciative toast, savor the moment contemplating the woman whose life you share. Then lift the golden spirit and drink in the smooth, mellow flavor with a hint of peat and a breath of smoke.

You'll both be glad that you married into the clan.

Ballantine's

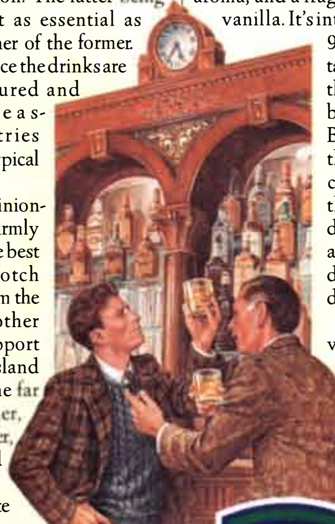
THE TRUE TASTE OF SCOTCH™

How to Choose a Scotch to Impress Your Taste Buds.

In the homes, pubs and hotel bars of Scotland there is a time-honored ritual, handed down from father to son, for the enjoyment of Scotch whisky. It involves clean glasses, moderate amounts of the amber nectar, and a penchant for spirited conversation. The latter being just as essential as either of the former. Once the drinks are poured and pleasantries exchanged, the typical debate begins.

Ballantine's Finest: Good taste installed and maintained.

One rather opinionated participant firmly maintains that the best single malt Scotch whisky comes from the Highlands. Another is adamant in support of the Orkney Island whiskies from the far north. For another, nothing but a softer, sweeter Lowland malt will do, while a staunch advocate of the bold, peaty flavor of Islay insists on being heard. Being Scots, each remains convinced that his opinion is correct, in perpetuity. Of course, when all is said and done, many prefer a blended Scotch like Ballantine's Finest. Because Ballantine's blends 42 of the top single malts from all over Scotland, the taste of the Finest tends to resolve disputes in a most diplomatic manner.



The Royal Scottish Debating Society. Learn from the experts.

If your favorite establishment doesn't yet serve Ballantine's, drop us a line.

But how is the well-intentioned American, who is not born in a land of peat and heather, to form an opinion about Scotch? A good place to start is right in your own home. First, free yourself from distractions. Now pour a healthy shot of Ballantine's and let your senses take over.

Can you detect the subtle fragrance of heather?

Swirl it around, feeling the heft of the glass as the golden liquid shifts from side to side. Notice the malty aroma, and a fragrance reminiscent of vanilla. It's interesting to note that 90% of your sense of taste is centered not in the tongue or mouth, but in the nose. In fact, Ballantine's blenders, the legendary men who create and preserve the Finest, test hundreds of single malts and blends on a daily basis. And they do it with their noses.

Now add a little water and hold it up to the light. It's a beautiful sight as the clear water and golden Scotch mingle together in a shimmering waltz of higher chemistry. Your anticipation builds. Lift the glass and take a sip. The first thing you may notice is a cool, slightly sweet taste. Let the smooth liquid float on your tongue for a moment before swallowing. Then a mellow whiff of peat makes its presence known, and finally, a soft brush of smoke in the finish. It's as if all the virtues of all the wonderful, diverse, eccentric Scotch whiskies landed in your glass at once.

Even though your nose might be the final authority, we'll understand if your taste buds are also impressed.

Please write. We welcome all correspondence.

Ballantine's

THE TRUE TASTE OF SCOTCH™

Write to: Ballantine's P.O. Box 8925 Universal City, CA 91608.

LETTERS

To the Editors:

Your report concerning the discovery by Murray Goodman and his colleagues that "sweetness is related to the shape and size of a molecule" ["Sweet and Sour," "Science and the Citizen," SCIENTIFIC AMERICAN, November, 1989] put me in mind of a strikingly similar conjecture made some 2,000 years ago. In *De rerum natura,* Lucretius theorized that things tasted bitter or acrid because they were composed of sharp-pointed atoms "hurtful to the senses." For instance, he maintained that seawater's tang came from rough atoms of salt intermixed with the smooth atoms of fresh water.

THOMAS A. REISNER

Department of Literatures
Laval University
Quebec, Canada

To the Editors:

The relationship between taste and the shape of substances was investigated microscopically by Antony van Leeuwenhoek, who first published his findings in 1673. Indeed, it was during van Leeuwenhoek's investigations of the saporous properties of pepper water that he recorded the first observations of microscopic organisms.

STEVEN R. SPILATRO

Department of Biology
Marietta College
Marietta, Ohio

To the Editors:

Anthony Robbins's assertion ["How to Control U.S. Health Costs," SCIENTIFIC AMERICAN, December, 1989] that more resources should be directed toward prevention and primary care of diseases omitted comments on a critical point: What happens when the beneficiaries of his improved care eventually get sick? Do they conveniently drop dead?

It is my suspicion that any savings realized by the prevention of some cases of nonterminal but costly diseases would be more than counterbalanced by other expenses, such as the additional costs of longer payout periods for pension funds and Social Security annuities. It is also my impression that most of the medical costs he laments are associated with terminal illnesses and can, at best, only be delayed.

MIKE RETHMAN

Columbia, Md.

To the Editors:

I believe that Dr. Robbins's effort to change long-term health patterns among Americans is a laudable one. It is unrealistic, however, to believe that this change will alter the costs of health delivery in the U.S. in the foreseeable future, if ever. Health costs are spiraling upward in part because of the increasing availability of technology to treat degenerative conditions, such as heart disease, and to prolong the lives of the elderly and low-birth-weight infants.

The problem lies in the public's attitude that decisions about how much health care to make available should never be made on the basis of cost. As long as this attitude persists, there will be no cost containment.

Can one really justify a liver transplant with all its attendant expenses? How much care should be given to a nursing-home resident with Alzheimer's disease? Who should really have heart surgery? Unless these kinds of issues are faced, medical costs will continue to increase irrespective of how much preventive care is provided.

ROBERT SHAPIRO

Cardinal Cushing General Hospital
Brockton, Mass.

The author responds:

Drs. Rethman and Shapiro have correctly detected the weak spot in any scheme that proposes to save money by preventing disease: everyone dies eventually, and most costs for medical care are concentrated at the end of life. When economists look at this problem, they usually conclude that programs that let people die quickly are more cost-effective. I described the twin strategies of prevention and primary care as ways "to improve life and simultaneously hold off unneeded expenditures." I hope that Americans, when deciding how to redistribute the $300-billion U.S. defense budget, will invest more heavily in prevention and primary care, because those are certainly the least costly ways to improve the nation's health.

ANTHONY ROBBINS

Boston University
School of Public Health
Boston, Mass.

To the Editors:

As the author of the article "Quasicrystals" [SCIENTIFIC AMERICAN, August, 1986], I was quite interested in John Horgan's piece on the subject in "Science and the Citizen" this past January. Although Horgan does a nice job summarizing the current debate about entropy versus matching rules in determining the structure of icosahedral crystals, he overstates my own views. While I do find the entropy arguments convincing, I certainly do not regard a search for perfect Penrose tilings as a "waste of time."

DAVID R. NELSON

Department of Physics
Harvard University
Cambridge, Mass.

To the Editors:

Concerning the article "New Radioactivities," by Walter Greiner and Aurel Sandulescu [SCIENTIFIC AMERICAN, March], I hope I'm not the first to point out that Otto Hahn and Fritz Strassmann were physical chemists, not physicists; that the discovery of nuclear fission, which was reported in January, 1939, was the result of experiments they conducted between 1935 and 1938 with the physicist Lise Meitner; and that Niels H. Bohr's explanation of the mechanism of fission, which he worked out jointly with John A. Wheeler, was published in 1939. As I discussed in *The Making of the Atomic Bomb,* it was the liquid-drop model, which Bohr developed in the late 1930's, that allowed Meitner and Otto R. Frisch, her nephew, in 1938 to see how fission might be possible.

RICHARD RHODES

Cambridge, Mass.

We and our authors thank you for sharing your thoughts with us. The number of letters that we receive makes it impossible for us to answer more than a fraction of them.

Paramount Pictures

Presents

THE STARSHIP ENTERPRISE™

The first pewter re-creation ever authorized by Paramount Pictures of the most famous starship of all time.

The Starship *Enterprise*.™ Symbol of the eternal quest to explore new worlds.

Now, the starship that carried Kirk, Spock and Uhura to the far reaches of the galaxy can be yours. Announcing the first fully authorized pewter re-creation of the starship that has inspired millions.

AUTHORIZED BY PARAMOUNT PICTURES.

Precisely reproduced with assistance from STAR TREK® creator Gene Roddenberry.

Sculptured and handcrafted in fine hand-finished pewter. The main sensor and navigational deflector glow with 24-karat gold electroplate. The top propulsion units blaze with fiery red crystal cabochons.

Available only on planet Earth, and only from The Franklin Mint.

Priced at $195.

AUTHENTICATED BY GENE RODDENBERRY.

ORDER FORM

The Franklin Mint
Franklin Center, Pennsylvania 19091

Yes, I want to take advantage of the unprecedented opportunity from Paramount Pictures to acquire the only authorized and fully authenticated pewter replica of the Starship *Enterprise*.™ The reproduction is hand-cast in fine imported pewter, hand-chased and hand-rubbed to a classic patina, and accented with crystals and 24-karat gold.

I need send no money now. I will be billed for my deposit of $39.* prior to shipment and then in 4 equal monthly installments of $39.* each, after shipment.

*Plus my state sales tax and a total of $3. for shipping and handling

Paramount®

A Paramount Communications Company

Please mail by July 31, 1990.

Signature ____________________
ALL ORDERS ARE SUBJECT TO ACCEPTANCE.

Mr./Mrs./Miss ____________________
PLEASE PRINT CLEARLY.

Address ____________________

City, State, ____________________

12135-109

Each replica of the Starship *Enterprise* is accompanied by a Certificate of Authenticity with Gene Roddenberry's signature.

50 AND 100 YEARS AGO

JULY, 1940: "Eyes of the scientific world widened the first week in May at the startling news that work of a 28-year-old physicist at the University of Minnesota in isolating the explosive uranium isotope, U-235, had been corroborated. The physicist is Dr. Alfred O. C. Nier, who modestly expressed dismay at extravagant reports which blossomed on front pages everywhere, to the effect that the incalculable forces of the atom were on the verge of being pressed into the service of mankind. There were fabulous predictions that this discovery made feasible bombs of unheard-of strength, that five or ten pounds of U-235 would propel a vessel around the world for an indefinite time without refueling. Rumors said that the discovery had set all available researchers in Germany at work in the Kaiser Wilhelm Institute at Berlin to seek a means of economically extracting U-235."

"None of the World War II warplanes rank so high on *all counts* as do our new pursuit-interceptors, two of which are some 70 miles faster than any Nazi ship known to be in service, and four of which will climb to 10,000 feet in far less time than Britain's 'Spitfire,' topmost World War II climber."

"The ability of any kind of microscope to show clearly minute objects is sharply limited by the wavelength of the illumination it uses. With high speed, short-wave electrons, it has become possible to observe objects and details of objects smaller than 100 angstroms in diameter—that is, about 1/2,500,000 of an inch. It may well be asked why X-rays or gamma rays, with wavelengths as small or smaller than electrons', could not be used for the same purpose. The answer is that no one has ever been able to devise a lens that would focus them."

"Artificial autumn can be brought to rose bushes, causing them to shed their leaves in a few days, by locking them up in the same room with apples, it has been discovered at Oregon State College by J. A. Milbrath, Elmer Hansen, and Prof. Henry Hartman. Ordinarily such defoliation would be undesirable, but when large numbers of field-grown rose bushes are being prepared for shipment to market it is necessary to rid them of their leaves, to cut down water loss through evaporation. The apples produce this effect because they give off small quantities of ethylene, which is also a common constituent of natural gas."

SCIENTIFIC AMERICAN

JULY, 1890: "The stock of fire crackers in this country at the present time is said to be from twenty-five to thirty per cent less than is usual at this season. This shortage is due in part to labor strikes in China, where all the small crackers and most of the large or cannon crackers are made, and also to the imposition by the Chinese government of the *lekin,* or tax. No evidence exists of the use of gunpowder as an agent of warfare until the middle of the twelfth century, nor did a knowledge of its propulsive effects come to the Chinese until the reign of Yunglop in the fifteenth century—a thousand years after its first employment in fire crackers."

"The subject of securing to the originators and owners of new varieties of fruits and vegetables the sole right by law to propagate and sell them engaged the attention of the recent meeting of American nurserymen in New York. The proposed law to that effect is certainly a praiseworthy attempt to secure to originators of varieties compensation for their labor, but it strikes us as involved in difficulties. Patents for machinery may be accurately described, and each kept entirely distinct from all others; but the differences between varieties of plants are so indefinite that a description in words would fail to point them out. The only practicable remedy appears to be in the use of names as trade marks."

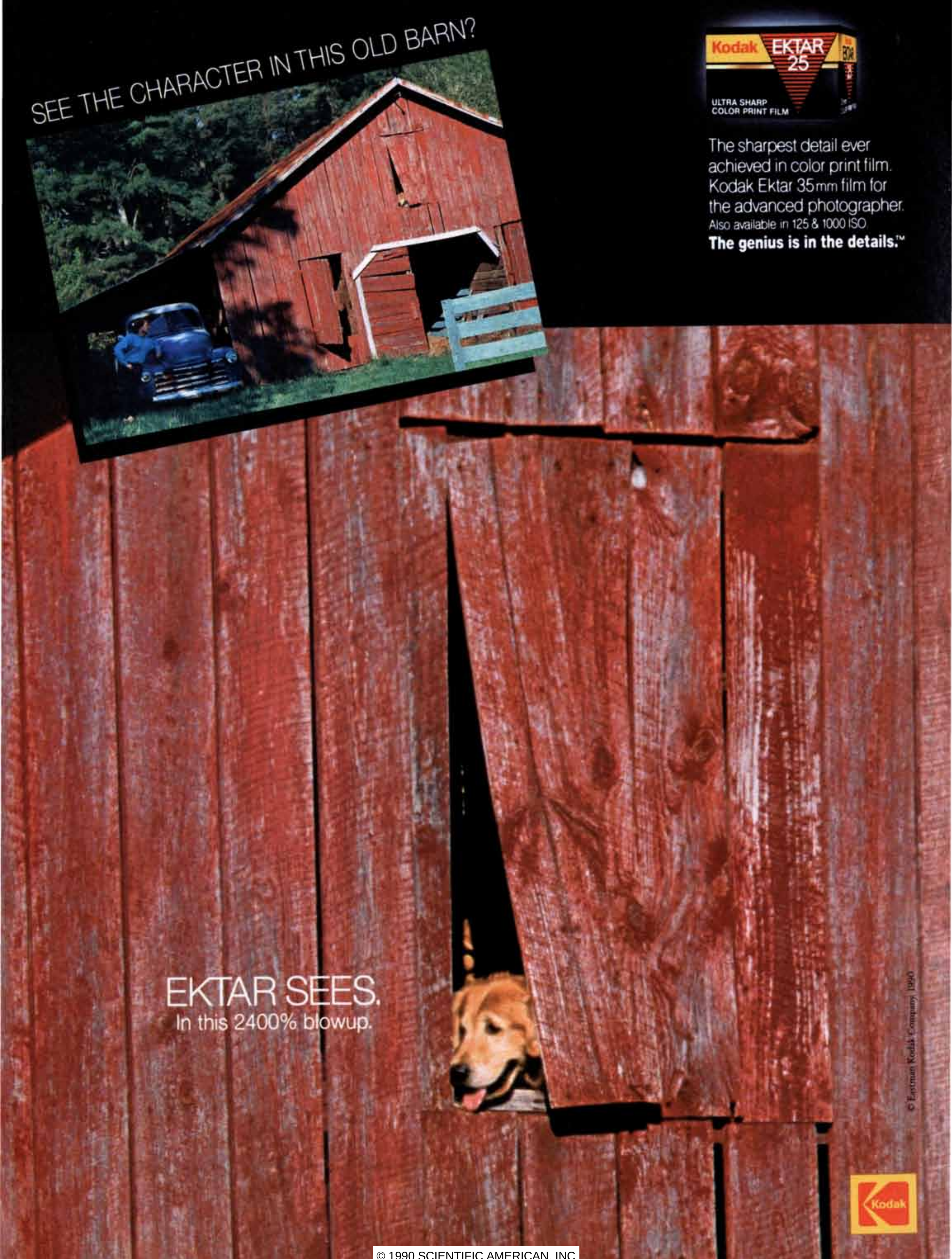
The electric trap

"A novel telephone station is being introduced in Connecticut. The instrument cannot be used unless a fee is paid. If five cents is dropped in the slot, it strikes a bell of a high note, once. A quarter strikes a bell of a lower note, once. A half dollar strikes that bell twice, while a silver dollar strikes a very low tone 'cathedral gong.' "

"It is estimated that from fifteen to twenty persons out of every hundred bitten by mad dogs or cats develop hydrophobia, but in 2,164 persons treated at the Pasteur Institute to January, 1887, there was a mortality of only 1.4 per cent, while in 1887 the mortality was reduced to 1.3 per cent, and in 1888 to 1.16 per cent. As touching this point, Sir Henry Roscoe says: 'Pasteur's treatment is really a race between a strong and an attenuated virus. In cases in which the bite occurs near a nerve center, the fatal malady may outstrip the treatment in the race between life and death. If the weakened virus can act in time, it means life; if the strong virus acts first, prevention comes too late, and it means death.' "

"The new cruiser Philadelphia, built for the government by Messrs. Cramp & Sons, at Philadelphia, has been so far completed as to be able to make her four hours' trial under steam, as required by the contract. By the terms of the latter the vessel was to be capable of making a mean speed of 19 knots per hour during a four hours' run at sea. If she made less, then the contractors were to forfeit $50,000 for each quarter knot below the standard. If she exceeded 19 knots, they were to receive a premium of $50,000 for each quarter knot in excess of the standard. The trial took place on the 25th of June, off the southeasterly end of Long Island. It is believed the ship made an average of 19 1/2 knots per hour, and earned a premium of $100,000 for her builders."

"This electric trap forms the subject of an American patent recently issued to Mr. F. Scherer, a resident of Paris, France. Any suitable bait is located within the cage, behind a grid composed of metal wires, arranged side by side to form the positive and negative wires of the circuit. When the rat or other foredoomed victim, seeking the bait, comes in contact with the wires, the circuit is thereby closed."


SEE THE CHARACTER IN THIS OLD BARN?
Kodak EKTAR 25
ULTRA SHARP COLOR PRINT FILM
The sharpest detail ever achieved in color print film. Kodak Ektar 35mm film for the advanced photographer.
Also available in 125 & 1000 ISO.
The genius is in the details.™
EKTAR SEES.
In this 2400% blowup.
© Eastman Kodak Company, 1990
Kodak

An infrared seeker, designed to function under severe environmental and operational conditions, is being prepared for flight evaluation tests. The brassboard model is part of the Strategic Defense Initiative's high endoatmospheric defense interceptor (HEDI) program. The system will locate an incoming reentry vehicle target, compute an aimpoint, and generate appropriate missile guidance commands. The Hughes Aircraft Company-built mercury-cadmium-telluride seeker, which will be cooled by liquid nitrogen, will operate in the mid-infrared band, and generate megabits of data every second. Output from the sensor will be processed by three Intel 80386-based signal processors.

The first optical fiber durable enough to meet military specifications consists of a unique metal-coated fiber that can be soldered to provide a hermetic seal. The fiber, called a "pigtail", is used to connect an optical fiber cable to a package containing a laser or sensor and associated electronics. Typically, optical fibers are coated with plastic for protection. The plastic is later removed and the fiber is vacuum metalized to enable soldering, however, this leaves the fiber weak. Because a hermetic coating is applied as the fiber is drawn, the Hughes pigtail retains its initial high strength. The Hughes metal-coated pigtails can be used in fire detection systems, radiation environments, undersea cables, high power laser transmission systems, and other environmentally demanding applications.

An automated planning system helps eliminate errors in assembly instructions for workers building radar modules. The Integrated Manufacturing Process Information System (IMPIS), developed by Hughes, features an integrated manufacturing database which stores standardized graphics and texts. IMPIS reduces the time to prepare planning documents by 50 percent over manual methods and creates a more accurate and consistent planning package. The detailed planning documents, referred to as picture books, created with the system are used by factory workers to build specific parts or assemblies. IMPIS has been upgraded to generate standard labor hours required to complete each part or assembly.

An automobile manufacturer helps improve the reliability and quality of its cars with thermal imaging systems. The Hughes-built PROBEYE® Thermal Video Systems (TVS) is being used by General Motors for a variety of applications. Included are tests to verify the placement and heat dissipation of electronic components used for dashboard and other electrical subassemblies. Thermal testing is also used to improve the reliability of voltage regulators, to monitor glass thickness and uniformity assuring higher quality windows and windshields, and to evaluate automotive air conditioning and heating systems and engine cooling systems. The TVS consists of a PROBEYE® infrared imager, an image processor, and a high resolution color monitor.

At Hughes, we're applying our mission analysis expertise to a wide range of critical DoD mission areas and societal issues. These important studies involve such areas as air combat, air defense, C^3, land combat, space defense, transportation safety, law enforcement and air quality control. To continue these vital efforts, we currently have opportunities for Mission Analysts with a proven record of performing defense analysis or systems effectiveness studies and a minimum of 5-10 years experience. Send your resume to: D. Robinson, Hughes Aircraft Company, Dept. S2, P.O. Box 92426, Los Angeles, CA 90009. U.S. Citizenship may be required. Equal Opportunity Employer.

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

HUGHES

Subsidiary of GM Hughes Electronics

SCIENCE AND THE CITIZEN

Combining AIDS drugs,
Amphibian warfare,
Hubba-Hubble,
Roald Hoffmann: poet

The Death Penalty

Most Americans favor it, but what purpose does it serve?

Portugal abolished it in 1976; Denmark in 1978; Nicaragua and Norway in 1979; France in 1981; the Netherlands in 1982; Australia in 1985; the Philippines and East Germany in 1987; Cambodia, New Zealand and Romania in 1989; and Namibia this year. Worldwide, the trend is clear: a growing number of nations have decided that capital punishment is wrong.

The U.S.—alone among Western democracies—is defying the trend. Indeed, since the Supreme Court ended a 10-year moratorium on executions in 1976, Americans have embraced the death penalty with a vengeance. Recent polls show that about 75 percent of the public supports capital punishment, up from 57 percent in 1972.

Little wonder, then, that politicians are competing with each other to demonstrate their enthusiasm for capital punishment. Thirty-six states already have death-penalty statutes, and pressure is building in at least two more—New York and Alaska—to jump on the bandwagon. From 1976 through late May of this year, states had electrocuted, gassed, injected or shot 128 prisoners; more than 2,300 others are on death row, awaiting similar fates. Lawmakers at the state and federal level have proposed a welter of bills to expedite executions and to extend their reach: drug "kingpins" are the most popular target.

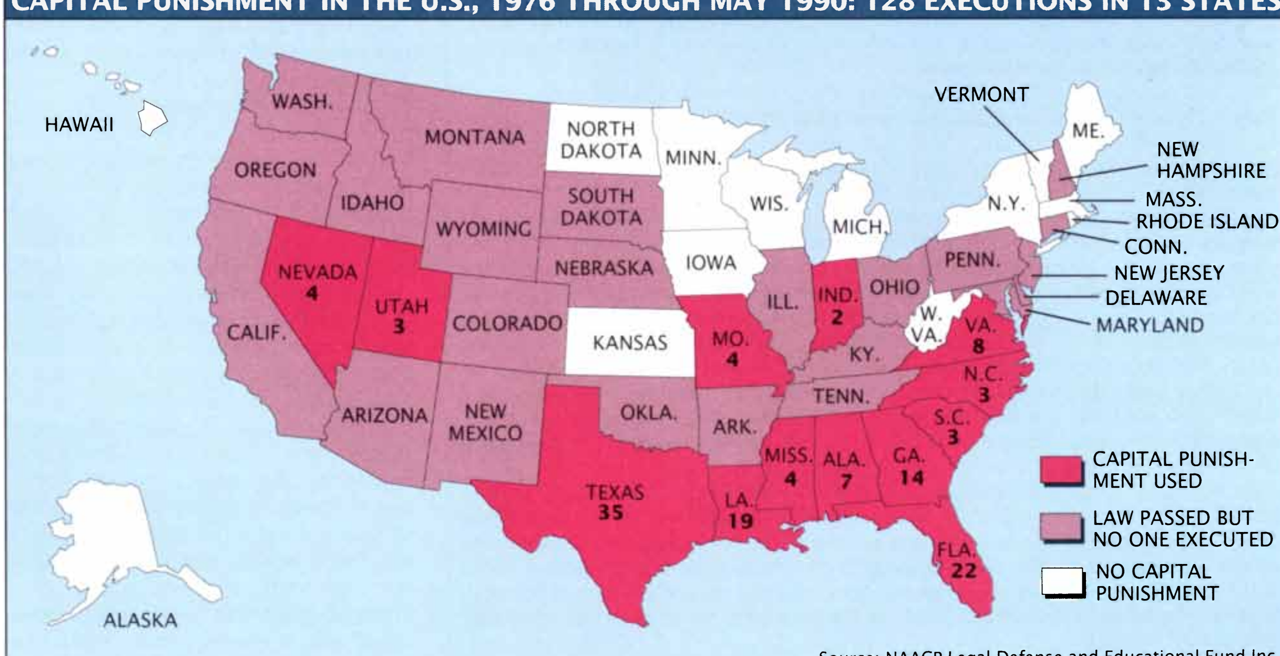
What purpose do executions serve? None, according to Amnesty International, the American Civil Liberties Union, the NAACP Legal Defense and Educational Fund and other organizations opposed to capital punishment. "People are understandably frightened and angry about violent crime in this country," says Henry Schwarzschild, who heads the ACLU's capital punishment project. "But the death penalty is not the solution."

More than a century of research in the U.S. and other countries, Schwarzschild points out, has produced no evidence that capital punishment reduces the rate of murder or other violent crimes. In fact, a study by William J. Bowers, a criminologist at Northeastern University, suggests that executions may have the opposite effect. When he analyzed the murder rate in New York State from 1907 to 1963, Bowers found that the number of murders rose by an average of slightly more than two in the month following an execution. He theorizes that executions—far from deterring acts of violence—foment them by "brutalizing" the public.

Of course, executing a murderer deters him from killing again—permanently. The publicity invariably generated by any murderer who repeats his crime after being released from prison has created the impression that such incidents are common, observes Hugo Adam Bedau, a philosopher at Tufts University. Actually, Bedau says, murderers have the lowest rate of recidivism of any class of felon: of 2,646 murderers released in 12 states from 1900 through 1976, 16 were convicted of a subsequent murder. People who served time for other offenses committed murder at a much higher rate.

If society decides that any recidivism is unacceptable, Bedau points out, it can always tighten parole requirements or even eliminate parole entirely in

CAPITAL PUNISHMENT IN THE U.S., 1976 THROUGH MAY 1990: 128 EXECUTIONS IN 13 STATES



Source: NAACP Legal Defense and Educational Fund Inc.

more murder cases. Contrary to popular belief, sentencing convicts to life in prison rather than death would actually save states money, according to Robert L. Spangenberg, an attorney who directs the Boston Legal Assistance Project. Studies done in Kansas, New York and Florida, Spangenberg says, show that states spend anywhere from $1.6 to $3.2 million to obtain and carry out a capital sentence; states could incarcerate someone for 100 years or more for less money.

The chief factor driving up the costs of capital cases is the appeals process, which has taken an average of eight years for those executed since 1976. Lawmakers are determined to compress the process. The Supreme Court has limited appeals of capital sentences at the state level, and in May, Chief Justice William H. Rehnquist called for restrictions on federal appeals as well. Congress is considering two bills that would impose such restrictions.

Yet even the most drastic proposals for curtailing appeals would reduce the period from conviction to execution only by two years at most, Spangenberg predicts. He asserts that the proposals could actually boost the costs of execution, since they stipulate that states set aside more money for the initial defense of capital cases to reduce the chance that a verdict or sentence will be reversed on appeal.

Limiting appeals may also increase the likelihood of the ultimate miscarriage of justice: the execution of an innocent person. In the past 18 years, at least 27 people condemned to death have later been found innocent by a higher court. Some of these reversals came about through sheer serendipity. The innocence of Randall Dale Adams, released last year after spending 12 years on death row in Texas for murdering a police officer, came to light only because a filmmaker happened to take an interest in the case. Others have not been so lucky. From 1900 to 1985, at least 23 Americans were executed for crimes they did not commit, according to a 1987 report in the *Stanford Law Review.*

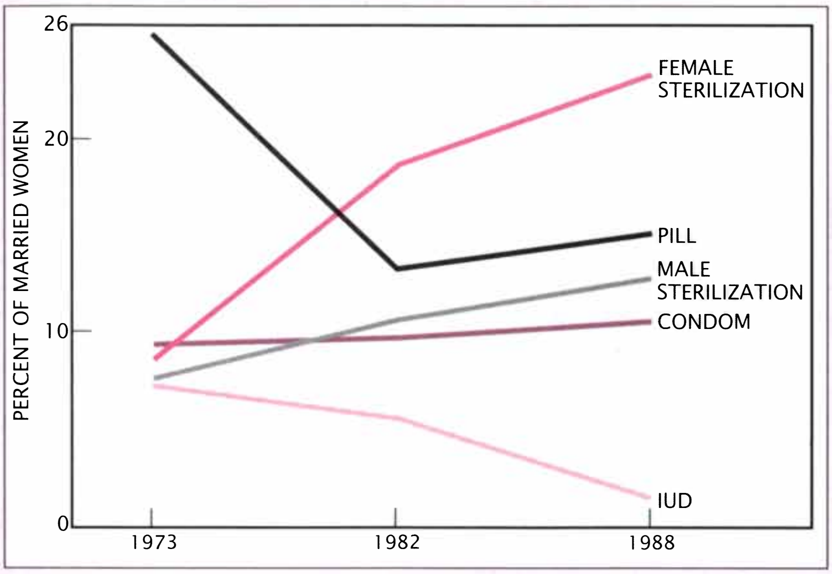
Opponents of the death penalty hope that such facts can turn the tide of public opinion. They try to take heart in the fact that when pollsters give respondents a choice between execution and life imprisonment without parole, the percentage favoring the death penalty drops from 75 percent to less than 50 percent. The percentage falls further—below 30 percent—if the inmate is required to provide monetary compensation to the family of his victim.

But opponents acknowledge that the pace of executions is likely to increase sharply in the next few years, because many inmates have run out of appeals. "We have exhausted the major constitutional issues," observes the ACLU's Schwarzschild. "We don't have any more even in the oven."

Evidence of racial bias in sentencing helped to convince the Supreme Court to rule against capital punishment in the late 1960's and early 1970's. New studies show the race of the victim to be the most significant factor: 84 percent of those executed since 1976 murdered a white person, although half of the murder victims in the U.S. are black. No white has been executed for killing a black in that period.

Three years ago the racism issue was raised anew when the Supreme Court considered the death sentence of Warren McCleskey, a black man convicted of killing a white police officer in Georgia. David Baldus, a professor of law at the University of Iowa, presented a study showing that in Georgia murderers of whites were 4.3 times more likely to receive a death sentence than

As birth-control options in the U.S. dwindle, more married women turn to sterilization



Sterilization—particularly of women—has become the birth-control method of choice for married Americans. According to a study by the National Center for Health Statistics, nearly one in four married women between the ages of 15 and 44 has opted for sterilization.

That conclusion is based on surveys conducted in 1973, 1982 and 1988 of married women of childbearing age; some 30 million women fall in this category. By 1988, 23.4 percent of this group had been sterilized, almost three times the percentage in 1973. The percentage of husbands who received vasectomies also increased in that period but much less dramatically.

Women, obviously, still bear most of the responsibility for birth control. But why are so many turning to sterilization? William D. Mosher, who coauthored the study, notes that three quarters of American women have all the children they want by the time they reach 30. After that, he says, women's contraceptive options become increasingly limited. Many physicians do not recommend the pill for women older than 35, for smokers or for those with certain medical conditions. Safety concerns have also sharply curtailed use of the intrauterine device (IUD).

Other alternatives—including condoms, diaphragms, foams and sponges and the rhythm method—are safer but much less reliable. At the same time, surgical sterilization of women has become less risky and easier to perform, Mosher says; most women can go home on the day of the operation.

The trend toward sterilization will probably continue. One reason, Mosher points out, is that the aging of the "baby boom" generation—born in a population surge beginning after World War II and ending in the mid-1960's—is still pushing up the average age of men and women in their childbearing years. Another factor is that—as the National Academy of Sciences reported recently—various political, legal and regulatory factors are blocking the development of better birth-control alternatives in the U.S. —*J.H.*

those who killed blacks under similar circumstances. The court did not dispute Baldus's findings of systematic bias, but it ruled that the defense had to prove bias in each particular case to have a sentence overturned.

The Supreme Court has also permitted executions of retarded people and those who were younger than 18 when they committed the crime. More than 75 death-row inmates fall into one or both of these categories, and at least four have been executed in the past decade. The latest was Dalton Prejean, a retarded black man electrocuted in Louisiana on May 18 for shooting a policeman when he was 17. Only four other nations—Bangladesh, Barbados, Iran and Iraq—have executed juvenile offenders in the past decade, according to Amnesty International.

Besides Prejean, four other prisoners were put to death in May. One was Jesse Joseph Tafero, electrocuted on May 4 for killing two policemen in Florida 14 years ago. When a prison official threw the switch, a sponge delivering electricity to Tafero's skull burst into flames. The executioner cut the current for six seconds—during which Tafero, smoke pouring from his hooded head, moved and breathed, according to eyewitnesses—and then shocked him again. After a total of six minutes and three jolts, Tafero was pronounced dead. The electric chair was introduced in the U.S. 100 years ago as a humane alternative to hanging. —*John Horgan*

Strength in Numbers

Researchers begin combining compounds to fight AIDS

A decade after AIDS first seized the public's attention, no cure is in sight, and all the drugs identified as potential treatments have significant limitations. Increasingly, researchers say their best hope for combating the disease is combining those drugs—both simultaneously and in sequence—to minimize their drawbacks and amplify their strengths.

"Combination therapy is not as sexy as a single magic bullet," says Robert Yarchoan, who is conducting trials involving multiple drugs at the National Cancer Institute. "But we hope that the sum of some treatments may be more than the parts." Combination therapy, Yarchoan points out, has already proved effective against some forms of another extraordinarily tenacious disease: cancer.

The common denominator in almost all combination therapies being considered for AIDS is AZT (azidothymidine, also called zidovudine). AZT, which is still the only AIDS treatment approved by the Food and Drug Administration, slows the progress of symptoms and reduces the mortality rate in people with AIDS. The drug's chief drawback is that it kills bone-marrow cells; it can cause anemia so severe that blood transfusions are needed. Making doses smaller reduces the side effects, but researchers still have not agreed on what the minimum dose should be.

Another limitation of AZT—and potentially of all treatments that attack the AIDS virus directly—has come to light only in the past year. Studies by Douglas D. Richman of the University of California School of Medicine at San Diego and others show that the human immunodeficiency virus (HIV) develops resistance to AZT after about six months. Resistance develops more rapidly in those in whom the disease is more advanced, Richman says.

These findings have spurred the already intense effort to find drugs to replace or supplement AZT. Two candidates—which have shown anti-HIV activity in test-tube studies and small clinical trials but have not yet been proved effective—are ddI (dideoxyinosine) and ddC (dideoxycytidine). Like AZT, ddI and ddC block the reproduction of the AIDS virus by interfering with its genetic machinery. They also have troublesome side effects. Both drugs cause painful inflammation of peripheral nerves, and ddI has been linked to a potentially fatal inflammation of the pancreas.

Other candidates for combination therapy include two drugs produced by biotechnology: soluble CD4, which consists of a cell receptor that can prevent HIV from binding to cells, and alpha interferon, which combats both HIV and Kaposi's sarcoma, an AIDS-related cancer. Researchers hope these agents can complement the effects of AZT and its analogues, ddI and ddC, because they hinder the activity of the AIDS virus at different stages of its cycle. If the virus evades one roadblock, it still has others to face.

Candidates for Combination

AZT The only approved antiviral treatment for AIDS. Causes anemia. Chronic use leads to viral resistance.

ddI Shows antiviral activity but not yet proved effective. Causes inflammation of peripheral nerves and pancreas.

ddC Not yet proved effective in clinical trials. Inflammation of peripheral nerves has limited wider testing.

ALPHA INTERFERON Shows antiviral activity in small clinical trials. Retards spread of Kaposi's sarcoma.

SOLUBLE CD4 Protects cells from AIDS virus in test-tube studies. Some promising data from clinical trials.

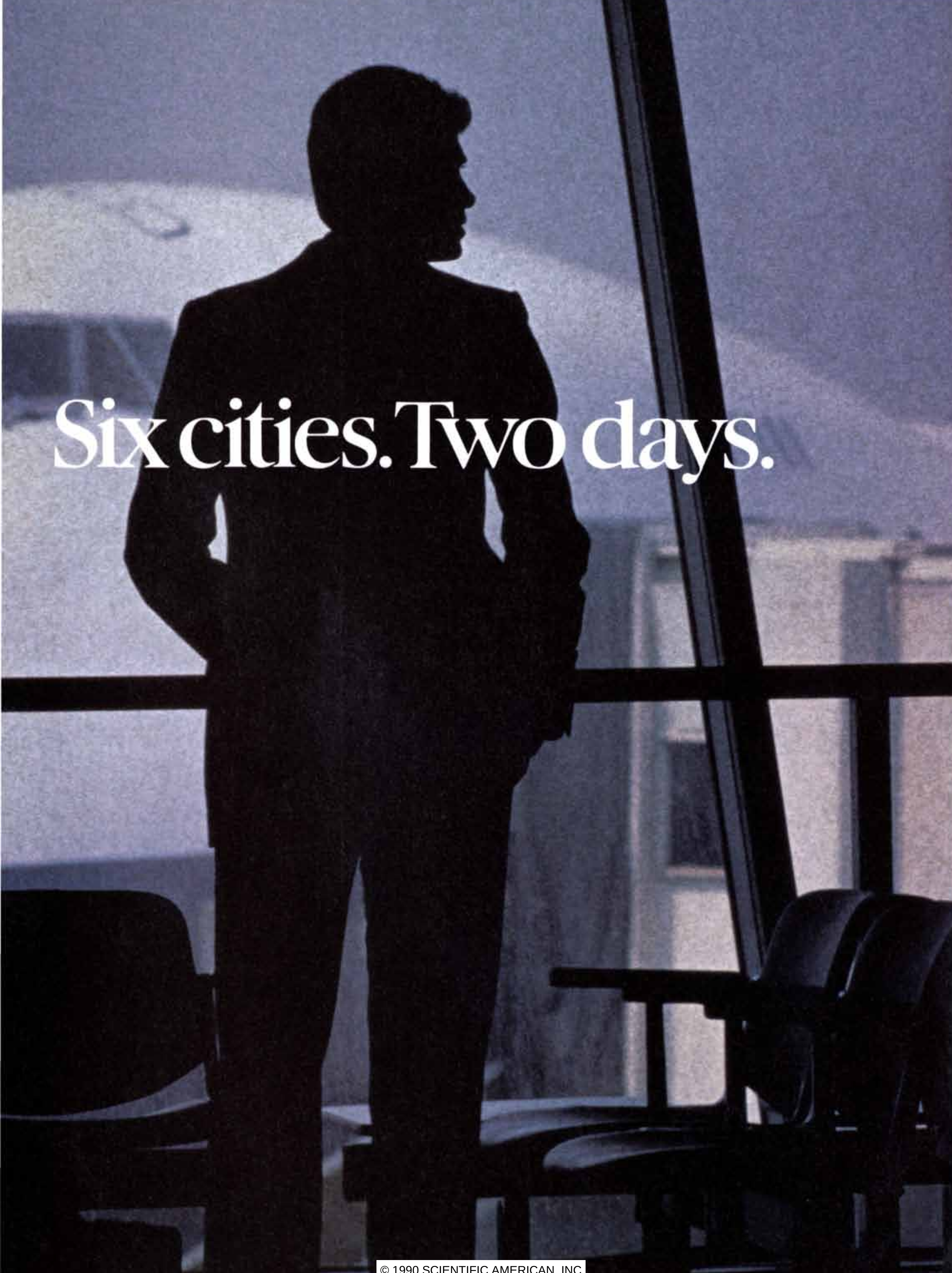
The agents that attack the AIDS virus are also being tested in combination with drugs that combat the opportunistic infections that actually kill people with AIDS. Two of the most common examples are acyclovir, an antiherpes agent, and pentamidine, which has been approved for warding off AIDS-induced pneumonia.

Various combinations of these drugs are being tested in cell cultures and in small clinical trials. A group at the Massachusetts General Hospital is doing in vitro tests of AZT in combination with CD4 and alpha interferon. Clinical trials involving alpha interferon and AZT are under way at the Memorial Sloan-Kettering Cancer Center. And researchers at the National Cancer Institute are giving 14 patients AZT and acyclovir for one week, ddI for another week and ddC for a third week before repeating the cycle. Preliminary data from these and other tests are encouraging, according to Anthony S. Fauci, director of the National Institute of Allergy and Infectious Diseases. "Combination therapy is certainly the strategy for the future," he says.

Precisely assessing such therapies, however, will be extraordinarily difficult. David A. Katzenstein, who is testing AIDS therapies at the Stanford University School of Medicine, explains that some of the variables involved in evaluating even a single drug increase exponentially as other drugs are added. He notes, for example, that researchers typically test a single drug at three different doses in three sets of patients. To evaluate how that drug will interact with another compound—also given at three different dose levels—the researchers would need nine sets of patients. Cancer researchers overcame these difficulties in developing an arsenal of combination chemotherapies, Katzenstein says, "but what people don't want to hear is that it took them 20 years."

Mathilde Krim, director of the American Foundation for AIDS Research, offers a more optimistic view. She notes that enrolling more people in clinical trials should accelerate the evaluation process. This is already beginning to occur, she says, at scores of community-based centers for AIDS research all across the country. —*J.H.*


BOMBAY SAPPHIRE. POUR SOMETHING PRICELESS.
PRODUCT OF ENGLAND
BOMBAY
SAPPHIRE
Distilled
LONDON
DRY GIN
Distilled from 100% Grain Neutral Spirit from a 1761 recipe.
1 Litre 94° Proof
Bombay® Sapphire™ Gin. 47% alc/vol (94 Proof). 100% grain neutral spirits. © 1989 Carillon Importers, Ltd., Teaneck, N.J.


Six cities. Two days.

RAIN FOREST in Serra dos Carajás, Pará, Brazil, was one region considered by scientists establishing conservation priorities. Photograph by Luiz Claudio Marigo, Peter Arnold, Inc.

Mapping the Immeasurable

Amazon researchers face up to the vastness of their job

With his khaki shirt unbuttoned to expose grizzled hair and a necklace, Guilherme M. de La Penha might be a prospector, a rancher or a poacher. All are common here in the Amazon, drawn by the promise of wealth from the planet's most abundant wilderness.

As it is, de La Penha belongs to another group. The director of the Goeldi Museum in Belém, one of Brazil's two main Amazon research centers, he heads a team of about 100 scientists, whose work on everything from archaeology to zoogeography is as varied as the region itself. But de La Penha is frustrated. "The problem," he says in brusque tones, "is too much of the Amazon has to be examined. Too much is said without scientific base."

In January some 100 ornithologists, entomologists, botanists, primate zoologists and other scientists took a step toward resolving this problem by creating a conservation map. During a 10-day meeting in Manaus, the attendees ranked areas with respect to levels of endemism (which reflects the number of species unique to a locality), diversity and presence of endangered species. Overall, 53 percent of the entire Amazon basin was deemed worthy of one of five grades requiring priority consideration; the rest was either unknown to scientists or unsuitable.

Although a coarse approximation, the conservation map "will be a big help," according to Eneus Salati, the new director of the National Research Institute for Amazonia (INPA), the area's biggest research center with some 280 scientists. A climatologist with a previous stint as INPA director in the early 1980's, Salati thinks 75 to 80 percent of the forest should remain covered by canopy. His goal is to devise a plan for sustainable development. He intends to augment the biogeography map with studies on soil, climate and local populations. The results are scheduled to be presented to other Brazilian agencies in November.

Providing a more thorough assessment of the basin's biota, however, is a staggering task. Scientists estimate that no fewer than three million species occur in tropical or subtropical regions, and the total could be 10 times greater, according to a 1989 report to the U.S. National Science Board. A single small lake in the Amazon may be home to more than 100 species of fish, which makes finding samples of particular species "a little bit complicated," says Ulrich Saint-Paul, a biologist from the Max Planck Institute.

The human diversity is vast as well. Of some 170 indigenous languages in Brazil, fewer than 20 are fully documented, according to Goeldi Museum data. Linguistic research could aid taxonomy: an analysis of the language of the Kayapos, a tribe in the Xingu Valley, has already helped scientists "discover" several new species of bees.

Support for Amazon research is increasing, particularly from European countries, such as Great Britain and Germany, and among private groups. James L. Edwards of the U.S. National Science Foundation says the agency is exploring a wide-ranging biodiversity program for the tropical forest. The plans call for improving microbial systematics and ecology, which are now poorly understood despite the biotechnology potential.

Another NSF goal is training more researchers in such fields as conservation biology and restoration ecology. Together with the U.S. Agency for International Development, the NSF is scheduled to announce the first year's multimillion-dollar awards in July.

Brazil's contribution, which is imperative, is also expected to increase. The new government, installed in March, seems to be more concerned than its predecessors about preserving the Amazon basin. Indeed, Ghillean T. Prance of the Royal Botanic Gardens in Kew, England, says the gap between rampant destruction of the forest (which is about 90 percent of its original size) and conservation efforts finally seems to be narrowing.

Both Salati and de La Penha emphasize, however, that conservation plans must take into account the social concerns of the Brazilian Amazon's 1.4 million people, including its 220,000 Indians. One way to minimize conflict is to create buffer zones—such as national parks and extractive reserves where local residents could harvest rubber, Brazil nuts or other natural products from the forest—between highly restricted biological reserves and more populated areas. "If the social problem is not put first," de La Penha says, "we have no solution for the Amazon." *— Pat Janowski*

Of Mice and MS

Antibodies cure a crippling disorder—but not in people

"We've received many heartrending letters from people with multiple sclerosis based on early reports of our work," laments Leroy E. Hood of the California Institute of Technology, adding that "we have not developed a cure for multiple sclerosis or any other human autoimmune disease. Our technique is at least 10 years away from human application." He is, quite rightly, trying to cool imaginations that may be overheated by a new report in the *Journal of Experimental Medicine,* in which he and his colleagues describe an apparent cure for experimental allergic encephalomyelitis (EAE) in mice.

The pathology of EAE is nearly identical to that of multiple sclerosis (MS), the most common neurological disease in people between the ages of 20 and 40. Both disorders are relapsing, inflammatory conditions that can cause weakness, body tremors and, in extreme cases, paralysis. Both are autoimmune disorders in which the immune system attacks the protective myelin sheath surrounding peripheral nerve cells.

There are nonetheless crucial differences between MS and EAE. MS is a naturally occurring human illness caused by as yet unidentified genetic and environmental factors. EAE is an artificial disorder that researchers induce in laboratory animals by injecting them with a major protein constituent of myelin.

In EAE the injected protein stimulates certain *T* lymphocytes in the immune system to attack the myelin protein in the mouse's nervous system. MS wages a similar immunologic campaign against myelin, but the biomolecular details are still a mystery. Today the only way to impede the erratic progression of MS over the long term is with general immunosuppressive agents, such as steroids, which leave the body vulnerable to infection.

Hood, Dennis M. Zaller and Gamal Osman of Cal Tech and Osami Kanagawa of Washington University have successfully developed a more selective way to knock out destructive autoimmune reactions. Their technique uses two antibodies against a pair of peptide molecules, Vβ8.2 and Vβ13, one or the other of which appears on the surface of every *T* cell involved in the EAE autoimmune reaction. The injected antibodies turn the immune system against the traitorous *T* cells to destroy them. Because most other *T* cells in the body are not affected, the immune system can continue to fight infections.

Whereas 14 out of 17 untreated mice in Hood's experiments developed EAE, only one out of 20 treated with the two antibodies became sick, and its symptoms were relatively mild. When the double antibody treatment was given to mice that already had EAE, three out of five mice with paralyzed hind legs regained their normal movement. The fourth mouse, which was almost completely paralyzed, recovered except for some immobility in the tail.

The combination of Vβ8.2 and Vβ13 antibodies was much more effective than either antibody alone. Indeed, the antibodies against Vβ13 seemed to have no effect by themselves. Hood and other investigators had previously found that antibodies against the Vβ8 family of peptides could lower the incidence of EAE to about 25 percent.

Encouraging as the results are, Zaller echoes Hood's caution. He emphasizes that before their technique can be tried on people the antigens and *T* cells responsible for MS, lupus erythematosus and other human autoimmune diseases must be identified and studied. "Five to 10 years would be the minimal estimate," he concurs.

The identification process is already under way—at least in the case of MS. Shortly after Hood's group announced its findings, researchers led by Lawrence Steinman of Stanford University reported in *Nature* that they had identified the key *T* cells involved in multiple sclerosis. *—John Rennie*

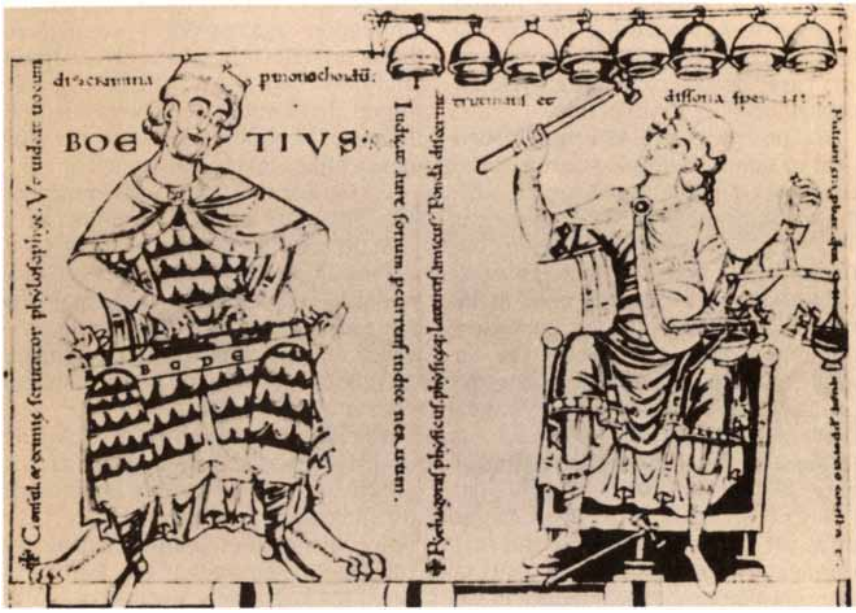
Pythagoras's Bells

Was the Greek mathematician the first expert in vibration?

The Greeks are often described as pure thinkers, disdaining gritty reality for the loftier realm of intellect. That was certainly true of Plato (429–347 B.C.) and, perhaps to a lesser extent, of his younger contemporary Aristotle (384–322 B.C.). But it was not so of a much earlier giant, Pythagoras (570–497 B.C.).

Best known for his work in mathematics (in particular the Pythagorean theorem, which states that the square of the hypotenuse of a right triangle equals the sum of the squares of the other two sides), he was also an avid experimentalist. He established science's first known laboratory, and in it he elucidated principles of vibration generally credited to Galileo.

At least, that is the contention of Andrew D. Dimarogonas, a professor of mechanical engineering at Washington University. Dimarogonas bases his views, which he sets forth in the *Journal of Sound and Vibration,* primarily on the writings of the Roman author Boethius (A.D. 480–524). Although he lived a millennium after Pythagoras, Boethius may have had access to contemporaneous accounts of Pythagoras that have now been lost, Dimarogonas says. Boethius's descriptions of Pythagoras are also partially corroborated by other ancient authors, such as Vitruvius (first century B.C.), Theon of Smyrna (second century A.D.) and Proclus Lycios (fifth century A.D.).



PYTHAGORAS* (right) *experiments with bells and hammers (which he is weighing in the scale in his left hand) in an illustration from a fifth-century history by the Roman scholar Boethius. Boethius pictured himself* (left) *holding a monochord, which Pythagoras may have used in acoustics experiments.

According to these and other writers, Pythagoras was born and raised on the Greek island of Samos but spent most of his career in Croton, a town in southern Italy. There he established a school where "rich people sent their kids," Dimarogonas says. It was also there that, as Boethius tells it, Pythagoras had an experience that led him to investigate the nature of vibration.

He was walking past a metal-working shop when he began to wonder why the hammers of different smiths rang at different pitches. He speculated that the pitch depended on the force with which the hammer was swung and so on the strength of the individual worker. To test his hypothesis, Pythagoras asked the workers to switch hammers. To his surprise, the hammers rang at the same pitch even when swung by different workers.

Pythagoras then went home, Boethius wrote, and conducted further tests on bells, reeds, strings and clay vessels. He also built a monochord, a single-stringed instrument that provided a benchmark for studies of pitch. These activities, Dimarogonas asserts, qualify Pythagoras's home as a laboratory.

Eventually, according to Boethius, Pythagoras surmised that the pitch at which an object vibrates—and so its frequency—depends not on the force with which it is struck but on the object's intrinsic properties: its composition, size, weight, length and (in the case of a string) tension. Galileo, working with a pendulum and other instruments, rediscovered these same principles some 2,000 years later and gained most of the credit for them, according to Dimarogonas.

Indeed, Dimarogonas says his research has taught him that almost every scientific "discovery" has some precedent. He notes, for example, that although Pythagoras was apparently the first to prove the theorem bearing his name, the Babylonians had already been applying the mathematical rule for centuries. *—J.H.*

Shocking Genes

Electromagnetic fields stimulate genetic activity

Whether the electromagnetic fields from electrical appliances and power-distribution equipment might cause cancer or birth defects has long been debated. It is accepted that the fields are too low in frequency and too weak in intensity to cause heating or genetic damage in cells. Without another obvious mechanism, researchers could not make a convincing case that such radiation could be harmful.

Now Reba M. Goodman of Columbia University and Ann Henderson of the City University of New York think they have found another way electromagnetic fields might influence cells: by indirectly stimulating genes and thereby altering the proteins cells produce. The workers appear to have demonstrated effects from fields like those found near domestic a.c. electrical devices.

The investigators first exposed "immortal" cells derived from a leukemia patient to the pulsed electromagnetic fields from devices used to promote the healing of bone fractures. After exposure to those fields, several genes increased the rate at which they produced messenger RNA, the substance that conveys genetic information from the nucleus into the body of a cell.

After further investigation, the researchers found that a 60-hertz field, no stronger than that produced near many domestic appliances, was an even stronger promoter of gene transcription than the bone-growth stimulators. A frequency of 45 hertz has an even greater effect.

Although not all genes are affected, those that are react in a characteristic way. Oddly, the effect is strongest after some 20 minutes of exposure; then it diminishes. More powerful fields, such as those found near high-voltage equipment, seem to promote transcription less well. "These results are highly reproducible," Goodman says.

The researchers suspect that very low frequency fields stimulate sensitive receptor molecules on cell surfaces. Such receptors are known to influence gene expression through a complicated sequence that alters the concentration of calcium ions in the cell.

The two investigators presented their data at a recent meeting of the Federation of American Societies for Experimental Biology in Washington, D.C., and since then other researchers have taken them more seriously. "I tend to believe their results," says Mays L. Swicord of the Food and Drug Administration, who already confirmed a similar effect on gene expression in preliminary experiments.

Nor is the Electric Power Research Institute, a utility-funded research organization that is helping support Goodman and Henderson's work, ignoring the new findings. According to the institute's Charles N. Rafferty, project manager for radiation studies: "The data are sufficiently strong that we feel it's important to carry out a major research program to see if the effects are real or not."

Further support for the idea that exposure to low-frequency electromagnetic fields might affect human cells comes from a quite different source. A draft study carried out by the Environmental Protection Agency finds evidence for a statistical association between high exposure to such fields and childhood leukemia. *—Tim Beardsley*


AMERICAN EXPRESS
ROB REINER

FIRE SALAMANDER irritated by a blade of grass squirts a neurotoxin from glands in its back in a photograph from the University of Texas at Arlington.

A Nasty Little Squirt

Europe's fire salamander lives up to its legend

The fire salamander, which haunts forests throughout Europe, has inspired a rich body of folklore. Medieval naturalists believed that the six-to-eight-inch-long amphibian could survive and even thrive in a fire (whence its name); that its skin, when strapped to a woman's thigh, kept her from becoming pregnant; that it poisoned the fruit of any tree it touched; and that it "casteth forth a white mattery liquor" to repel its enemies (as one 17th-century expert on serpents put it).

The French biologist Marie Phisalix-Picot pooh-poohed this last belief in 1900. She asserted that while glands on the salamander's back do secrete a powerful neurotoxin—which causes a burning sensation in mucous membranes and can be fatal if ingested—the animal cannot spray the liquid. That seemed to settle the matter.

Then a few years ago Edmund D. Brodie, Jr., a herpetologist at the University of Texas at Arlington, was observing a fire salamander in his laboratory when suddenly he felt an excruciating pain in his eyes. He was blinded for about 20 minutes, but his alarm and discomfort were allayed somewhat by his instant realization that perhaps Phisalix-Picot was wrong and the ancients were right.

In subsequent studies, Brodie and fellow biologist Neal J. Smatresk, both safely begoggled, confirmed that suspicion. By poking fire salamanders with blades of grass, they induced them to spray their "mattery liquor" over distances of at least seven feet. The amphibians shifted their bodies to direct the spray—sometimes toward the hand of the "attacker" but also frequently toward the face and eyes.

In *Herpetologica,* Brodie and Smatresk note that their finding puts the fire salamander in a rather exclusive group of animals—including skunks, certain types of cobras and gekko lizards and an assortment of insects—which can repel predators by spraying them with noxious fluids.

Could there be any truth to the other myths about the fire salamander? Brodie doubts it. In particular, he does not recommend that women test its contraceptive powers. —*J.H.*

Gothic Mystery

Pixelated probe pries prose from palimpsests

Much lore has come down to us because medieval scribes, in recycling the parchment of old books, failed to scrape away every vestige of their original contents. Such inadvertent entombments of ancient texts, called palimpsests (in Greek, "scraped again"), have strained the eyes of generations of scholars who tried to see beneath their surface. Some texts have still not been deciphered, even with the aid of ultraviolet scanners that highlight ink by causing it to fluoresce.

Now James W. Marchand, a philologist at the University of Illinois at Urbana-Champaign, has coupled ultraviolet scanning with computerized image-enhancement techniques developed for the space program. He employs the system, which he put together in his basement, for the reconstruction of the literary remains of the Goths, who ruled Italy after the fall of Rome and spoke a Germanic language that has no modern descendants.

Marchand uses a digitalizing camera to convert his ultraviolet photographs into a million pixels having 256 levels of brightness. Then he runs the data through computer programs that find edges, remove backgrounds and strip away unwanted intensity levels associated with later ink. "You wouldn't believe what sometimes comes out," he said. "We read things we could not read well before and get printable results from pictures we could not previously print."

Printing is key because Marchand wants to make scarce texts available to students in a form as close as possible to the scribe's original production. Marchand custom-designs a typeface based on the scribal hand in which the original was made. This allows him to create a facsimile that in many respects is superior to the original.

Such niceties are important because they retain the clues that lead one scholar to correct another's misinterpretation of a letter or word. Marchand has already reconstituted a page from the *Codex Argenteus,* a Gothic New Testament, and printed it for scholarly and not-so-scholarly purposes. "I made myself a T-shirt with a page from the *Codex* on it," he says.

Marchand dreams of expanding his basement scrivener's den into a national archive for the storage and analysis of digitized manuscripts. Such an archive would include holographic representations of runes and other monumental inscriptions and would perhaps transmit its treasures via the computer networks of the University of Illinois, a national supercomputer center. So useful and accessible a resource not only would illuminate the study of the Dark Ages, but it also would put pressure on those nonpublishing scholars and uncooperative archivists who monopolize a large share of the world's manuscripts. —*Philip E. Ross*


3-D SPREADSHEET
TIME/EXPENSE MANAGER
BOX JOCKEY GAME CARD
SCIENTIFIC CALCULATOR CARD
32K/64K MEMORY EXPANSION CARDS
8-LANGUAGE TRANSLATOR
NEW WIZARD SOFTWARE LETS YOU DO MORE THAN IS HUMANLY POSSIBLE.
Like your PC, The Wizard electronic organizer from Sharp never forgets a name. Never misses a date. And makes short work of jobs that it's only human to postpone. But unlike your PC, The Wizard slips into your coat pocket. So it goes places most PCs can't.
But with an optional PC Link, your IBM® PC-compatible or Macintosh® personal computer will never be out of reach.
And with a growing library of optional software cards, The Wizard can up- and download numbers into Lotus 1-2-3® from its own 3-D Spreadsheet. Store up to 64K of data on a single card. Track your expenses. Tackle scientific equations. Translate eight languages. Even play games. Yet despite all this power, it's humanly possible to master Wizard software quickly. Because these new programs are as easy to use as The Wizard itself. To find out where to see Wizard software in action, call 1-800-321-8877.
Links With PCs
TO DO LIST
BY DEADLINE
BY MANAGER
BY PRIORITY
BY PROJECT
WIZARD
SHARP EXPANDS THE POWER OF THE WIZARD.
SHARP
FROM SHARP MINDS COME SHARP PRODUCTS™
© 1990 Sharp Electronics Corporation

Orbiting Eye

Hubble's first photo promises better to come

On May 20, the U.S. opened its eye in the sky. After a frustrating month of technical glitches, a camera shutter in the *Hubble Space Telescope* slid aside and the $1.5-billion instrument recorded its first image. That "first light" view was a 30-second exposure of an unremarkable star cluster in the constellation Carina.

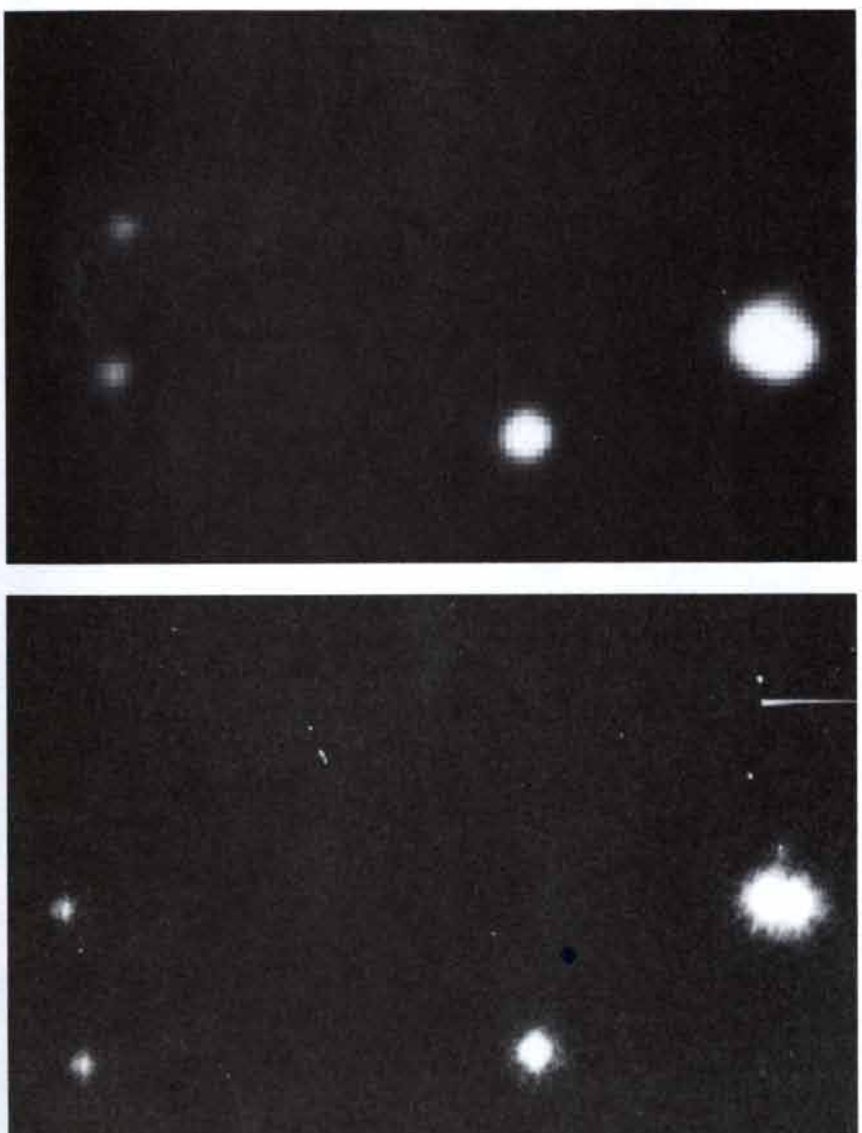
But the photograph, relayed to the earth shortly afterward, makes its point. Even though much of the delicate task of focusing *Hubble* still lies ahead, the first image has a resolution of about .7 arc second, twice as good as expected and considerably better than the 1.1-arc-second resolution achieved by a 100-inch ground-based telescope in Chile. A star that appears as a misshapen blob from the ground can be recognized as a double star in the *Hubble* image.

Since it was cast adrift from the space shuttle *Discovery* on April 25, the initial checks and adjustment of the telescope have been anything but smooth sailing. Problems started early, when one of the solar panels at first refused to unroll during deployment. Later, one of the two high-gain antennas needed to communicate with ground controllers at the Space Telescope Operations Control Center in Greenbelt, Md., kept pressing against its feed cable, sending the high-strung telescope into a panic-stricken "safe mode." Controllers solved that problem by rewriting software to restrict the range of movement of both antennas. The fix will limit the times that data can be transmitted and received, but controllers say that by careful scheduling they can avoid any significant loss of observing time.

At one point, efforts to calibrate the telescope's pointing systems were held up because star data sent to orient the telescope proved to be out-of-date. The data were based on a star catalogue produced decades ago, and one star had apparently changed in brightness—a phenomenon that could recur.

By late May, technicians were only part way through exhaustive checks on the telescope's systems. The telescope's tendency to roll slightly and vibrate whenever it passed in or out of the earth's shadow was a puzzle. Initial system checks, running about 10 days behind schedule, were expected to be completed during June. The checks will be followed by four or five months of calibration of the telescope's five instruments.

Nevertheless, hopes are high that *Hubble* will meet its goal of .1-arc-second resolution for most observations, which should start before the end of the year. Says Ray Villard of the Space Telescope Science Institute in Baltimore, Md., of the first images: "You ain't seen nothing yet." —*T.M.B.*



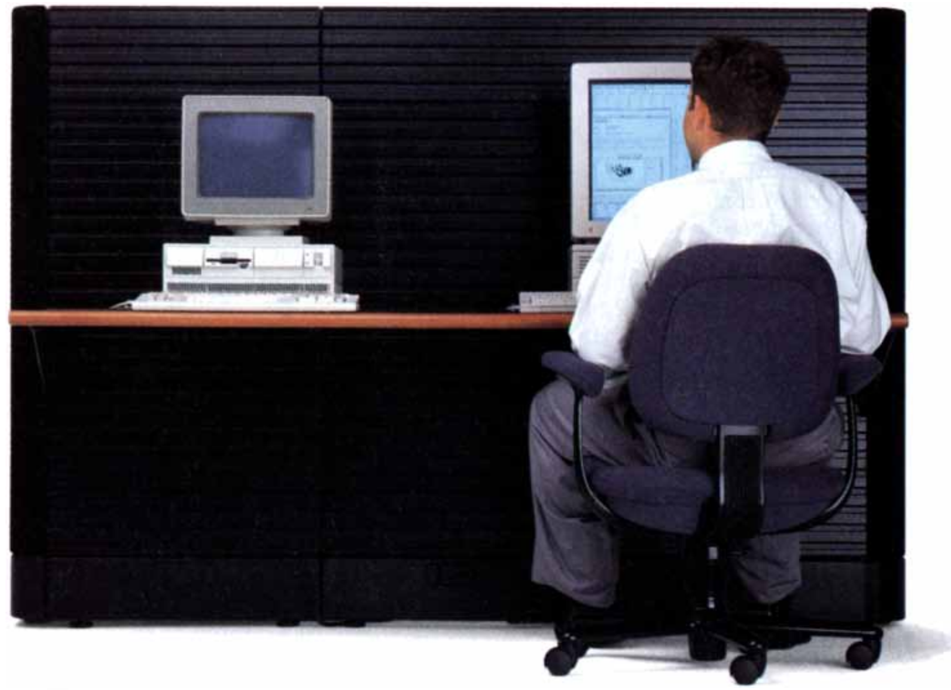
***FIRST PHOTOGRAPH taken by the* Hubble Space Telescope (bottom) *shows more detail than images from a ground-based telescope* (top).**

Lost in the Clouds

Astronomers hunt for 100 trillion missing comets

Those who recall trying to glimpse Comet Kohoutek in 1973, or the disappointing Comet Austin two months ago, know that spotting even a bright comet is no easy task. But astronomers continue to search for these objects because they are well-preserved samples of the material from which the sun and planets formed, and their distribution may be a relic of the solar system's early days. If it were possible to observe comets that originally orbited other stars, astronomers could learn much about the composition and dynamics of other stellar systems.

Every comet observed so far appears to have originated from an orbit bound to the sun. These comets are thought to come from a huge, tenuous swarm, the Oort cloud, at 1,000 to 100,000 times the earth's distance from the sun. Current theories hold that comets formed much closer in, alongside the giant planets (Jupiter, Saturn, Uranus and Neptune) in the outer solar system. The planets' gravity would have ejected



some of the icy surrounding material into distant orbits, forming the Oort cloud.

If so, many comets undoubtedly exist outside the cloud. Dynamical studies by several groups, including Martin Duncan of the University of Toronto and colleagues, suggest that most of the comets broke free from the sun's hold and sailed into interstellar space. This process is still occurring (at a very slow rate) today. Our solar system alone may have shed 100 trillion comets. If ours is typical, the Milky Way should contain hundreds of billions of comets per cubic light-year.

These interstellar comets fascinate S. Alan Stern, an astronomer at the University of Colorado at Boulder, who calls them "the smoking gun of planetary formation." If one of them were to approach the sun, it could easily be identified because it would follow a hyperbolic path rather than the parabolic paths of comets from the Oort cloud.

The density of interstellar comets should reflect the fraction of stars that have planetary systems to eject the comets. So far, however, not a single comet has been seen to follow a hyperbolic path. Stern hopes a multiyear, dedicated search for faint hyperbolic comets will settle the question of whether planetary systems are common features or a lucky fluke. Paul R. Weissman of the Jet Propulsion Laboratory is concerned that "a deliberate search is beyond anything now practical." Finding a hyperbolic comet will just take patience—perhaps 50 to 100 years of waiting, he thinks.

A faster way to locate interstellar comets might be to search directly for Oort clouds around other stars. Comets that approach their star will evaporate and shed dust that would emit infrared radiation. The National Aeronautics and Space Administration's *Infrared Astronomical Satellite* failed to find such emission; the upcoming *Space Infrared Telescope Facility* might be sufficiently sensitive to do so.

In *Nature,* Stern's group suggests that red giant stars might be bright enough to heat their Oort clouds to visibility. The star's energy would cause water molecules and hydroxyl radicals to emit radiation. This signal could be detectable because stars are too hot to radiate at these wavelengths.

Astronomers are understandably excited by the prospect of finding comets from other stellar systems. Stern muses that the only other way to determine the prevalence of planetary systems throughout the galaxy is to "talk to an extraterrestrial visitor"—of the noncometary kind. —*Corey S. Powell*

PROFILE: Modest Maverick

Hoffmann's world of chemistry, poetry and pedagogy

Nobel laureate Roald Hoffmann caresses a model of a molecule whose geometry he calls "seductive and beautiful." To me, it's just yellow triangular panels with green spheres at the corners. He traces his finger from a sphere—an atom—along the edge of a panel—a molecular bond. Then I notice that the panels form a series of tetrahedrons stacked one on top of the other. Touching three atoms at a time, Hoffmann twists his hand around the model, revealing a triple helix. I begin to decipher the intricate structure of this "tetrahelix." But is it art?

Chemists certainly commend the beauty of Hoffmann's work. He was awarded the 1981 Nobel prize in chemistry and the 1990 Priestley Medal, the American Chemical Society's highest honor. Now Hoffmann would like all scientists and humanists to appreciate the aesthetics of chemists' work. The tetrahelix, he explains, is beautiful not just because of its shape but also because of its novelty, its dynamics, its utility, its richness.

In his 32nd year as a theoretical chemist, Hoffmann now devotes much of his time to communicating the beauty of molecules. As a professor at Cornell University, he awakens first-year students to the fundamentals of chemistry. As a poet and writer, he exposes the connections that unite chemistry, literature and art. As a television-show host, he will introduce viewers to *The World of Chemistry*—a series of 26 half-hour programs scheduled to air on public television in September.

It is a wonder that Hoffmann can find any beauty in his life after its ugly beginnings. In June, 1941, a month before his fourth birthday, German troops marched into his hometown of Złoczów, Poland (now Zolochev in Soviet Ukraine). Nazi officers forced him, his father, Hillel Safran, and his mother, Clara, to move to a Jewish ghetto. A few months later they were deported to a labor camp called Lackie.

In January, 1943, Safran arranged to smuggle his wife and Roald out of the camp. They were received by a Ukrainian teacher who hid them in the attic of a schoolhouse. In this dark, cramped room, Clara began to teach her son geography and reading.

Later, in June, his father made plans to escape from Lackie with several other prisoners. The Nazis discovered his

This biconcave bialy platelet
of the erythrocyte, the red
heart of the blood, holds the oxygen
carrier, hemoglobin. Four coiled

polypeptide chains, four subunits
changing pairwise twice in the fetus
to let it soak up placental O_2
steadily. Each chain a globular

protein, juxtaposed twining
of helical segments, predestined
kinks, sequences of amino acids
alike in sperm whale and horse,

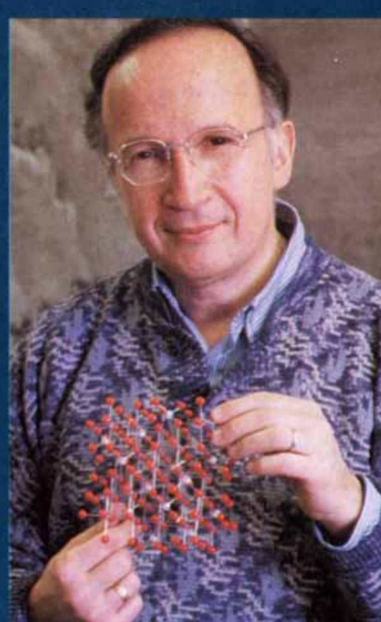
a meander of bonds around
the flat disc that colors all...heme,
the active site, the oxygen binding
site, a porphyrin, iron. Oxygen,

enflamer, winds to a pocket
molded by protein, binds iron, moves
it in consummation, chains
tethering heme tense—a far

subunit feels the first heme's bond
quiver, the chains pull, O_2 binds
easier. Cooperativity, an allosteric
protein. In 1937

not long before the war,
Felix Haurowitz watched crystals
of deoxyhemoglobin
shatter on oxygenation.

from "Jerry-Built Forever,"
Gaps and Verges



Roald Hoffmann

plot and executed him. Hoffmann later wrote in a poem, "...I was five/when the news came to us in the Ukrainian's attic,/and I cried, because my mother cried. That's when/my father became a hero.... The war ended,/80 of 12,000 Jews in our town survived."

After their liberation in June, 1944, Roald and his mother journeyed to Kraków, Poland. There Clara met Paul Hoffmann, whose wife had died in the Holocaust. A year later they were married. Hoffmann remembers his stepfather as a "kind and gentle" man. The Hoffmann family traveled through Czechoslovakia to Austria to West Germany and finally emigrated to the U.S. in 1949.

Hoffmann was introduced to chemistry at an early age through the biographies of Marie Curie and George Washington Carver. "I showed neither precocity nor early interest in chemistry," he commented during a recent interview. In the fall of 1955 Hoffmann entered Columbia University to prepare for a career in medicine. "My mother wanted me to become a doctor; maybe by now she's forgiven me for becoming a chemist," he said with a smile.

At Columbia, Hoffmann studied everything from mathematics to French to chemistry, taking six or seven classes each semester. He found art history most intriguing and nearly abandoned the laboratory in favor of the gallery. He completed his course work in only three years and graduated summa cum laude in chemistry.

Hoffmann's affinity for chemistry, however, mainly developed during the summers of his college career, when he studied the chemistry of cement and hydrocarbons at the National Bureau of Standards. It was this experience that motivated him to attend graduate school at Harvard University. For his doctoral thesis, working under Nobel laureate William N. Lipscomb, Jr., he predicted the structure of polyhedral hydrocarbons and boranes.

After receiving a Ph.D. in chemical physics in 1962, Hoffmann accepted a three-year fellowship at Harvard. It was during this period that he began his two most important collaborations: the first with Harvard professor Robert B. Woodward investigating the theory of organic chemistry; the second with his wife, Eva, raising two children, Hillel Jan and Ingrid Helena.

Woodward and Hoffmann formulated a general rule, which has been regarded as the most important conceptual advance in theoretical organic chemistry. As molecules combine with one another, they sweep through transition states. The reaction will proceed

Programming in FORTRAN no longer requires heavy machinery.

For all too many people, programming in FORTRAN means relying on a department mini or corporate mainframe.

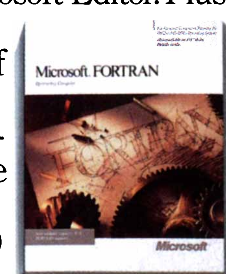
But with Microsoft® FORTRAN version 5.0, this no longer has to be the case.

Because now you can develop as well as maintain applications and subroutines. For the mainframe, the mini, and, of course, the PC.

All on that relatively light piece of equipment residing right on your desk. Otherwise known as your PC.

Now you're probably asking yourself how is this possible? Well, for starters, it supports FORTRAN 77 ANSI-standard syntax. Secondly, it has major VAX® extensions. And finally, it's completely SAA compliant.

And unlike "heavy" equipment, with FORTRAN 5.0 you can expect to find a range of high-quality development tools. Like our CodeView® debugger and Microsoft Editor. Plus, you can use Dynamic Link Libraries to link in portions of your code at runtime.



Giving you time to concentrate on other things. Like your work, perhaps.

To learn more, call (800) 541-1261, Department L88. You'll find that programming in FORTRAN doesn't have to be a real heavy experience.

Microsoft®
Making it all make sense™

with ease if the energy of the transition states is low and with difficulty if it is high. The Woodward-Hoffmann rule is a procedure for relating the energy of the transition states to the number of electrons involved in the reaction and their motions.

During his collaboration with Woodward, Hoffmann developed and refined a distinctive research style. "I don't start with big tasks or the great questions of chemistry," he remarked. "I do many small problems inspired by experimental work. I try to explain the shape of a molecule or some reaction. Everything in the world is connected to everything else. I know I will begin to see connections."

Hoffmann has successfully applied his methods to organic, inorganic and solid-state chemistry. He has devoted roughly 10 years of his life to each, but he confesses that his decisions to switch were not conscious or sudden. "I have always done the next interesting thing."

At times when Hoffmann has just entered a new field, his colleagues have criticized his work for being too simplistic. But propelled by his early successes, Hoffmann has managed to excel. In inorganic chemistry, he developed a technique to predict the structure and reactions of organometallic compounds. In solid-state chemistry, he translated the physicist's concept of densities of electron states into the chemist's idea of molecular orbitals. Indeed, Hoffmann is the only person in the history of the American Chemical Society to be honored for work in both organic and inorganic chemistry.

Hoffmann was only 44 when he shared the Nobel prize with physicist Kenichi Fukui of Kyoto University. Hoffmann received the prize in part for his contribution to the Woodward-Hoffmann rule. Had Woodward not died two years before the Nobel committee's announcement, Hoffmann believes, Woodward would have also been awarded the prize.

In many ways, Hoffmann's style of chemistry pervades his teaching. Even in introductory courses, he feels it is important to communicate what modern chemists find exciting. He prefers to begin a lecture with a discussion of a paper from the current chemical literature. He then highlights the connections between the paper and the curriculum. Nearly every year since starting at Cornell in 1965, he has taught introductory chemistry, which was attended this spring by 740 students.

Hoffmann believes that teaching complements his research efforts. "The desire to teach others, enhanced by being *obliged* to teach others, leads to greater creativity in research," he wrote in a column for the *Boston Globe* last November.

Out of a desire to express the beauty of chemistry and his own emotions, Hoffmann at the age of 40 began to write poetry. He was inspired by Mark Van Doren, a professor of poetry at Columbia, and Pulitzer prize-winning poet Wallace Stevens. "I should have taken a course, but I thought I was too old." Today his source of "valuable criticism and support" comes from a group of Cornell poets, including A. ("Archie") R. Ammons.

Hoffmann recently finished his second volume of poetry, *Gaps and Verges,* published by the University of Central Florida Press. Among his topics are his childhood experiences, romance, natural beauty and, of course, chemistry. Peter Harris wrote of his work in the *Virginia Quarterly Review,* "At times, Hoffmann's treatment of the scientific seems prosy.... But, in general, his poetry is distinguished by the appreciative ease [with] which it moves between scientific and nonscientific understanding."

Branching out from his poetry projects, Hoffmann has written several articles on science policy and many essays for the magazine *American Scientist.* Recently he teamed up with artist Vivian Torrence to produce a book entitled *Chemistry Imagined.* Torrence is creating some 25 collages on chemical themes. To accompany these images, Hoffmann is composing related essays and poems.

In 1986 Hoffmann began taping the television series *The World of Chemistry,* produced by the Educational Film Center in Annandale, Va. The program introduces a variety of chemical principles and theories, which are placed in historical context and are presented through experimental demonstrations and computer animation.

In front of a camera, Hoffmann seems awkward; his voice, which is usually soft and soothing, becomes high-pitched. "Our budget was too small to send me to acting school," he explained. Nevertheless, he has embarked on a new television project called *The Molecular World.* Still in the planning stages, it will include three prime-time programs.

Having made the transition from poetry to television and from organic to solid-state chemistry, Hoffmann, now 53, has earned celebrity status among chemists. In spite of his fame, he remarked that "it has not become any easier to publish a paper, nor, for that matter, a poem." *—Russell Ruthen*

SCIENTIFIC AMERICAN July 1990 Volume 263 Number 1

The Great Climate Debate

Greenhouse effect and the prospect of global warming is the subject of scientific and political controversy. Should we take steps now to avoid consequences we cannot foresee?

by Robert M. White

In the waning years of the 10th century, millions braced themselves for the apocalypse, believing that the approaching year 1000 was the very millennium—the end of the heavens and the earth prophesied in the Bible's Book of Revelation. Not surprisingly, the prospect of the impending Day of Wrath terrified normally sane people into rash and (in retrospect) foolish actions. Some gave away all of their possessions; others hastened to do harsh penance for their deeds.

In this final decade of the 20th century, a different kind of apocalypse causes widespread concern. This time the hand of God has been replaced by more visible agents: belching smokestacks, gasoline-powered automobiles, power-generating stations and the voracious destruction of forests, all of which may be turning up the heat on an overburdened environment. Global climate warming, some claim, threatens the very habitability of the planet. Others hold that the predictions of environmental collapse are not well founded and are goading us into hasty political action. Is our planet the "Endangered Earth," as *Time* magazine would have it in its 1988 year-end cover story? Or is it as *Forbes* magazine put it, "The Global Warming Panic: A Classic Case of Overreaction"?

ROBERT M. WHITE is president of the National Academy of Engineering. He was formerly chief of the U.S. Weather Bureau and administrator of the National Oceanic and Atmospheric Administration. He has also served as president of both the University Corporation for Atmospheric Research and the American Meteorological Society. In 1979 he was chairman of the first World Climate Conference of the World Meteorological Organization.

Debate in the media reflects uncertainty among climatologists and geophysicists. Some of the world's eminent authorities on the atmosphere recently hurled verbal brickbats at one another in the pages of the prestigious journal *Science*. Their charges of "junk science" and "science by consensus" reflect the acrimonious nature of the debate within the scientific community. Some members of the National Academy of Sciences, including one of its former presidents, charge that policymakers are being induced to take unwise actions on the basis of uncertain scientific evidence. Set against this view is the recent statement of the Union of Concerned Scientists urging action by the government. It was signed by 52 Nobel laureates and more than 700 members of the NAS.

In spite of the scientific uncertainty, government and nongovernment groups are rushing to outdo one another in urging drastic action now to "stabilize" the global climate. From Washington to Toronto and The Hague, from Cairo to Moscow, international conferences of experts and political leaders have called for action. Soviet President Mikhail S. Gorbachev, President George Bush, British Prime Minister Margaret Thatcher and French President François Mitterand share similar views on the climate-warming issue.

Back home, debate within the Bush administration on how the U.S. government should act is intense. Caught between the urgings for action from the Environmental Protection Agency and Congress and cautions from his science adviser and chief of staff, President Bush called two major conferences to address the issue of climate warming. The first, held in April, brought together the heads of scientific, economic and environmental agencies of many governments. The second will be an initial meeting of governments, scheduled for early 1991, to begin negotiation of an international convention to stabilize global climate.

While there are still doubts in the White House, Congress has been environmentally hyperactive. Many pieces of legislation have been introduced to address the predicted climate warming. Leading the bipartisan effort have been Senator Timothy E. Wirth of Colorado, Senator Al Gore of Tennessee and Congresswoman Claudine Schneider of Rhode Island. Some of this legislation is comprehensive and far-reaching. It offers suggestions for action on the energy, agriculture and transportation fronts as well as for intensified research.

The actions proposed would radically change the most vital functions of human economies. They could include such diverse actions as using energy more efficiently, shifting the fossil-fuel mix from oil and coal to natural gas, relying more heavily on renewable energy sources and using more nuclear and solar energy. Measures could also include implementing reforestation, phasing out use of chlorofluorocarbons (CFC's) and changing agricultural practices.

Policy initiatives of this kind would

alter the technology and economics of energy. Our use of land and water would also need to change. Economic growth in nations dependent on fossil fuels might be slowed. And the problems of arresting the growth of global population would become even more pressing. How can national and international policy formulation be moving so rapidly to address the specter of climate warming when agreement about the science is lacking and the economic and social costs of action have hardly been tallied? At the root of this thinking is a confluence of diverse scientific, economic and environmental forces.

The idea that the actions of humanity might change the composition of the atmosphere and hence the world's climate has deep historical roots. As early as the 1860's, it was suggested that slight changes in atmospheric composition might bring about major variations in climate. Increases in carbon dioxide (CO_2) and other atmospheric trace gases can contribute to what has been called greenhouse warming because these compounds allow the sun's energy to reach the surface of the earth, thereby warming it, while preventing much of that energy from being reradiated to outer space.

At the end of the 19th century the Swedish scientist Svante A. Arrhenius calculated how changes in carbon dioxide content would affect the temperature at the earth's surface. He estimated that a doubling of carbon dioxide would produce a global warming of about seven to 11 degrees Fahrenheit (four to six degrees Celsius), not too far off modern calculations. Yet it was only with the inception of the International Geophysical Year, a worldwide experiment in 1957 to monitor the global environment, that scientific data validating the increase of carbon dioxide in the atmosphere became available.

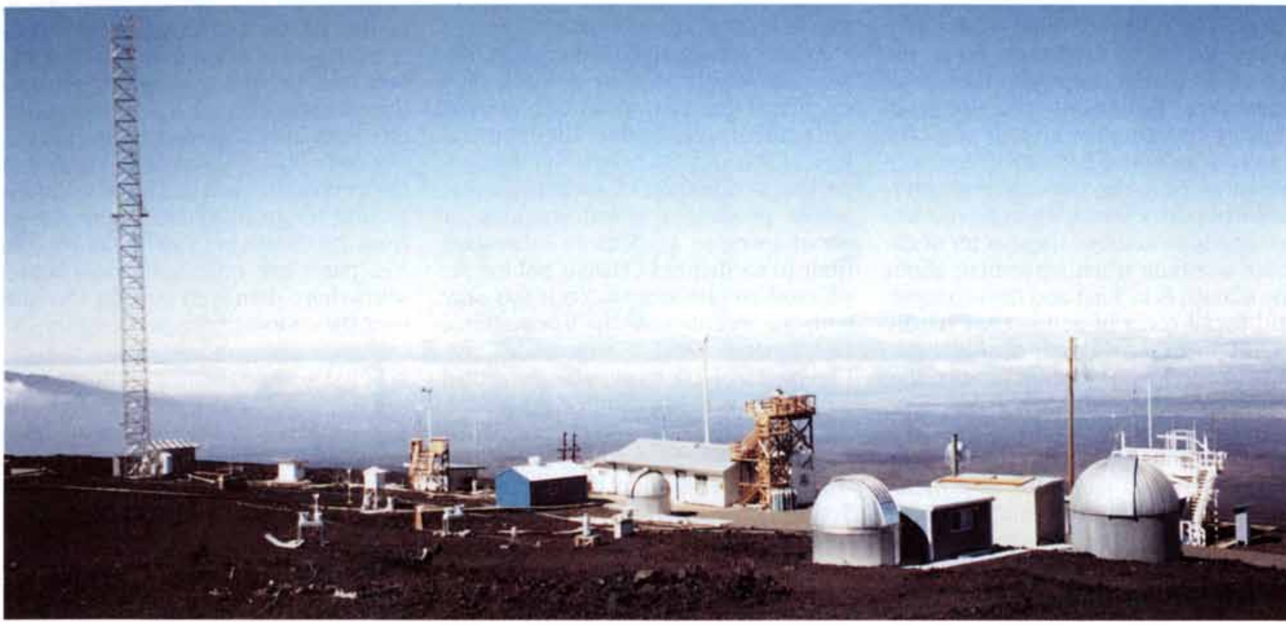
Roger Revelle, then director of the Scripps Institution of Oceanography, his colleague Hans E. Suess and C. David Keeling, his student, undertook such measurements. Revelle had long contended that humans were carrying out an unintended geophysical experiment on the atmosphere by burning fossil fuels. Determined to monitor the carbon dioxide content of the atmosphere, he persuaded Keeling to develop the instrumentation.

The measuring devices were placed in the Mauna Loa climate observatory in Hawaii at an altitude of about 11,000 feet. Beginning in 1957, the data they collected revealed a systematic increase in atmospheric carbon dioxide. Keeling's observations were verified at the South Pole and at other locations around the world. To date, the change from 290 parts per million in 1880 to 352 parts per million in 1989 represents more than a 20 percent increase over the course of the past century.

Just before the International Geophysical Year began, on the other side of the continent from Scripps, another development key to unraveling the climatic consequences of increasing carbon dioxide emissions was taking place. Under the leadership of the world-famous mathematician John von Neumann at the Institute for Advanced Study in Princeton, N.J., the first attempts were made to represent the atmosphere mathematically on digital computers.

Von Neumann's team of brilliant young scientists was headed by Jule G. Charney. Later known as the father of

BARGES STRANDED by all-time low water levels on the Mississippi River were one effect of the record-breaking drought of 1988. That long, hot summer thrust the greenhouse effect into the public eye and set off the present policy debate.



MAUNA LOA CLIMATE OBSERVATORY in Hawaii is located at about 11,000 feet. Data collected here, beginning in 1957 and continuing through the present, were the first to document a steady increase in atmospheric carbon dioxide levels.

numerical weather prediction and arguably the most important American figure in the transformation of weather prediction from art to science, Charney demonstrated the feasibility of using computers to perform the task. Von Neumann and Charney calculated the first 24-hour weather forecast in 1950 on a primitive digital computer, the ENIAC, maintained by the U.S. Army Signal Corps in New Jersey.

Looking beyond these efforts, von Neumann called climate forecasting the "infinite prediction." One of the young scientists in the Princeton group, Norman Phillips, made the first attempt at modeling the global atmosphere in 1956. It was coincidence that later, in 1963, an unusual laboratory of the National Oceanic and Atmospheric Administration (NOAA) was established on the campus of Princeton University under the leadership of Joseph Smagorinsky, a strong-willed and hard-driving young scientist who had been one of von Neumann's group. The laboratory was totally devoted to the mathematical modeling of the atmosphere using the largest and fastest digital computers available.

Called the Geophysical Fluid Dynamics Laboratory, the center harbored researchers from many nations interested in this new approach to the study of the atmosphere. Among them was a young Japanese scientist, Syukuro Manabe. Modest and retiring but completely dedicated to the work, he developed the first climate model in collaboration with his colleague Richard T. Wetherald in the 1960's. In 1975 they calculated that a doubling of the carbon dioxide content of the atmosphere would produce a global climate warming of about five degrees F (three degrees C), averaged over the surface of the earth. This calculation has been verified in many different laboratories and has not changed substantially.

Keeling's observations, together with the calculation of Manabe and Wetherald, triggered the wave of climate-change research that has marked the past two decades. Studies have since been undertaken in many parts of the world, including Europe and the Soviet Union. In the U.S. the National Research Council conducted studies in 1966, 1977, 1979, 1983 and 1987. These inquiries were chaired by such leading scientists as Gordon J. F. MacDonald, Revelle, Thomas F. Malone, Charney, William A. Nierenberg and economist William D. Nordhaus.

Yet because there were no immediate consequences for human health and no evident manifestation of climate change, the work was slow to arouse political concern. The most politically influential study was the one prepared in 1979 at the request of Frank Press, now president of the NAS, who was then White House science adviser to President Jimmy Carter. It was also in 1979 that the World Meteorological Organization in Geneva, recognizing the potential global significance of the issue, convened the first World Climate Conference.

Gradually, scientific awareness that humanity might actually be causing a planetary disruption began to register in the political world. Although there was much debate over the validity of projections from computer models, the observations of greenhouse-gas increases, however, were precise, well measured and verified in many parts of the world. These were corroborated by additional data that documented increases in other greenhouse gases such as methane, or natural gas, and CFC's.

Meanwhile mathematical-modeling groups in this country had been established not only by the NOAA but also by the National Aeronautics and Space Administration, the Department of Energy and the National Science Foundation. The leaders of these laboratories became the "gurus" of climate warming. Incisive and original in their work, Stephen H. Schneider of the NSF's National Center for Atmospheric Research in Boulder, Colo., and James E. Hansen, the leader of NASA's Goddard Institute for Space Studies, were soon to become frequent witnesses at innumerable congressional committee hearings on the subject.

Although the mathematical models of all the groups yielded similar results, the details of the geographic distribution of climate changes differed from one model to the other. All projected that an increase in carbon dioxide would bring about a gradual warming, but the timing of this warming would depend on the rate of global energy use. They all agreed that if reasonable assumptions were made about future global energy consumption, it would be around the middle of the next century that the carbon dioxide content of the atmosphere might double.

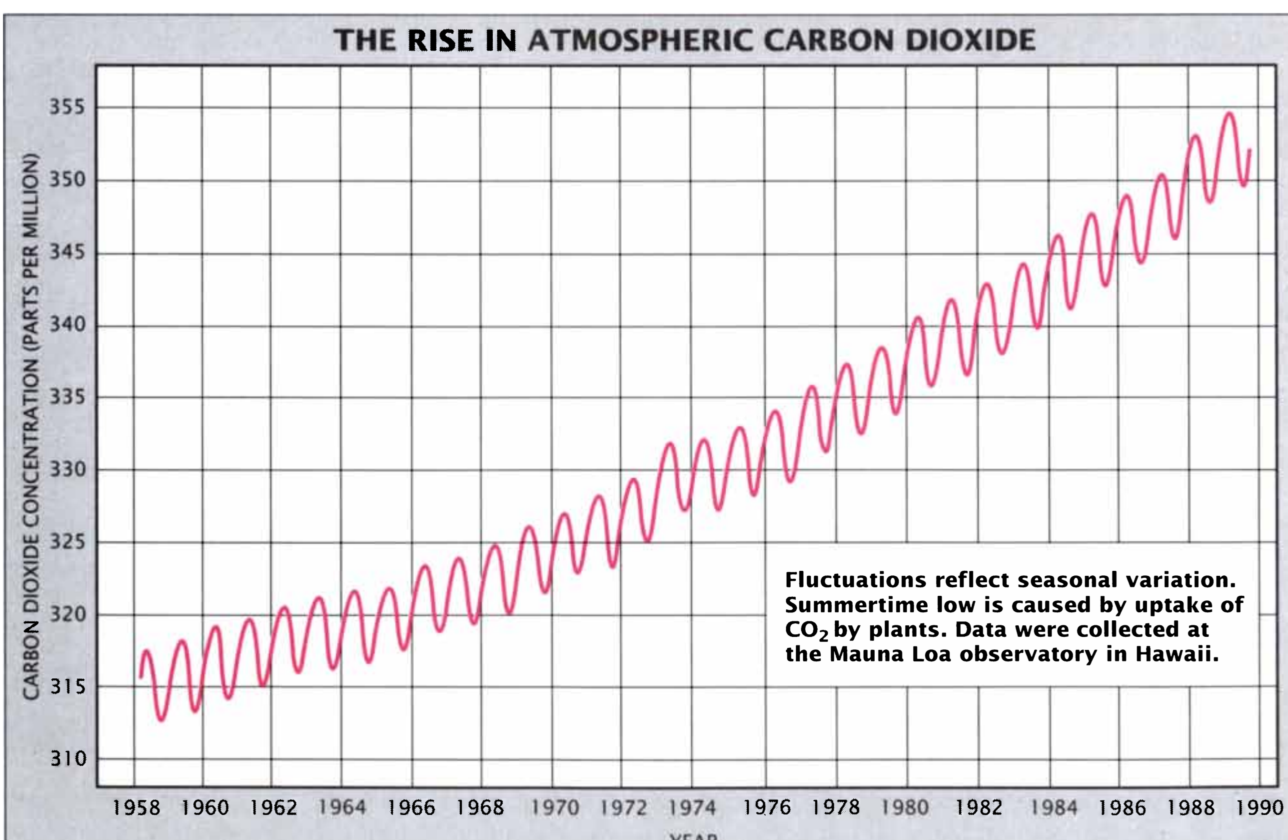
Just how much this doubling of carbon dioxide would increase temperatures, however, varied greatly from model to model. Some predicted as little as a two-degree F (one-degree C) increase, whereas others predicted increases of as much as nine degrees F (five degrees C). The differences in predictions became central elements in the debate about whether the models were sufficiently reliable to warrant policy actions. Further, it made a great difference whether the actual increase was at one or the other end of this range. At the low end, the normal resilience of society would probably be sufficient to accommodate the changed climate. Changes at the high end portended severe disruptions.

These projected temperature changes may appear innocuous because variations of this magnitude are experienced in the normal course of daily and seasonal weather. Their full implications can be appreciated by noting that it took only a two-degree F average decrease in temperatures in Europe to cause the run of several frigid centuries (from the 1400's to the 1800's) known as the Little Ice Age. Nine degrees F is believed to be the difference in temperature that separates the end of the last great ice age 12,000 years ago from the present. Further, the projections indicate that the Northern Hemisphere would experience in just a half century an unprecedented temperature change, 10 to 50 times faster than the change since the last ice age.

Those who are not familiar with mathematical models or the way computers are used to make these projections can be forgiven for being confused—or even annoyed—by the great disparities in the results. Among investigators it is understood that mathematical models are only approximations that attempt to simulate the processes that govern atmospheric behavior. The atmosphere is so complex that it is impossible to represent it in very great detail in these mathematical models. It is possible to represent only certain features and to make assumptions about how the oceans and the atmosphere interact, how the rate at which the oceans take up carbon dioxide varies and how clouds affect the exchange of energy between the earth and the atmosphere. Even the largest computers cannot represent the atmosphere, oceans and land surface in fine detail. Indeed, scientists approximate the conditions in the atmosphere by thinking of it as a set of observations spaced about 500 kilometers apart.

The political calls for action are being played out against the backdrop of that uncertainty. On one side, the view is that if there is a chance that model predictions could be correct, the consequences could be so dire that immediate action to arrest climate change would be imperative. The alternative view, equally cogent, is that commitment to action with vast economic and social consequences is unwarranted in light of both the scientific uncertainty and the absence of knowledge of the economic costs. John H. Sununu, White House chief of staff, in remarks he made at the annual meeting of the National Academy of Engineering in the fall of 1989 gave voice to this position:

> *Although I agree that* [global warming] *is a critical issue, the fact is that the models with which analysis is being done and with which policy is being moved, as good as they may be, still are based on element sizes measured in hundreds of kilometers in length and width, and tens of kilometers in thickness. I suspect that no one who has ever been involved in engineering simulation would feel comfortable making major*



decisions in which the elements were orders of magnitude greater than the details on which they were looking for information. And yet the fact is that we are moving toward binding international policy based on conclusions being drawn by policymakers who have no sense at all of the difference between the levels of confidence they should have and levels of confidence they want to have. A system is not valid just because it gives you the answers you want. And yet so much policy is being made in reaction to that principle.

The solution to the dilemma should be simple: Since the carbon dioxide content of the atmosphere has increased by more than 20 percent over the past century, we ought to be able to detect the climate warming in the global temperature record during the same period. Researchers have sought to do this, but it is a much more difficult task than it first appears. The problem is that climate is always in a state of natural fluctuation. Separating out the changes that are caused by increasing carbon dioxide from the natural changes is tricky scientific business. Moreover, the climatic temperature record is based on scattered and irregular observations not taken specifically for the purposes of determining climatic conditions.

Even so, careful analysis of these temperature records by scientists in the U.S. and in the U.K. sought to detect whether a climate warming has occurred and whether such warming is consistent with the prediction of the models. The prevailing view is that the climatic record over the past century for the entire globe reveals a net increase in temperature ranging from .5 to 1.0 degree F (from .3 to .8 degree C). But set against this conclusion is the disturbing result that similar increases in temperature cannot be detected over the past century in the U.S., where observations are numerous and accurate.

Even if the temperature rise is real, a puzzle remains that workers have been unable to unravel: Is the rise in global temperatures a natural fluctuation or a result of the increase in greenhouse gases? All that can be said is that the observed increase is consistent with the lower end of the temperature increases predicted by the computer models. Consequently, the temperature records, as well as the predictions of mathematical models, provide substance both to those who believe the evidence warrants action now and to those who believe the evidence is still too weak.

The rush to policy action was, I believe, catalyzed by the disastrous drought of the summer of 1988. During this drought, one of the worst on record, the water in the Mississippi River fell so low that navigation was impossible over long stretches, urban water supplies were threatened and crops throughout the grain belt were devastated. Both officials and the public wondered whether this was the greenhouse effect manifest. Indeed, records show that in the U.S. five of the years of the 1980's were among the hottest on record, and the average temperature for the decade as a whole was the warmest since instrumental records have been kept.

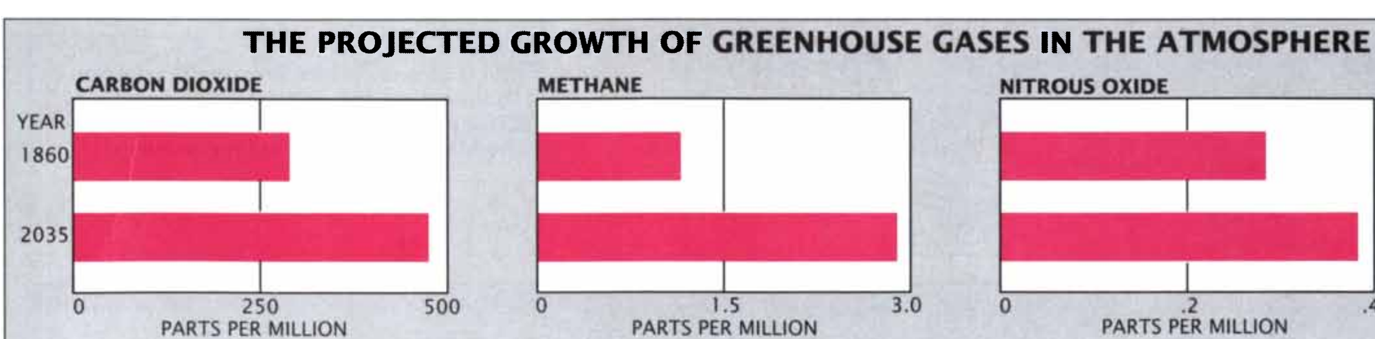
Prompted by heat and drought, congressional hearings addressed the question of whether the greenhouse effect had arrived. These hearings were unremarkable except for a statement by Hansen. When he stated that he was 99 percent certain that the greenhouse warming had begun, as evidenced by the sequence of warm years in the 1980's, the public took notice. His opinion prompted members of Congress to consider whether the prudent course was to move rapidly to legislation aimed at protecting the habitability of the planet from catastrophic consequences.

Hearings followed hearings. Both the atmospheric researchers and the more general environmental community began to choose sides on whether immediate policy action was justified. The reaction from environmentalists was quick and vociferous. Several environmental and scientific groups began to advocate international agreements restricting the emissions of greenhouse gases.

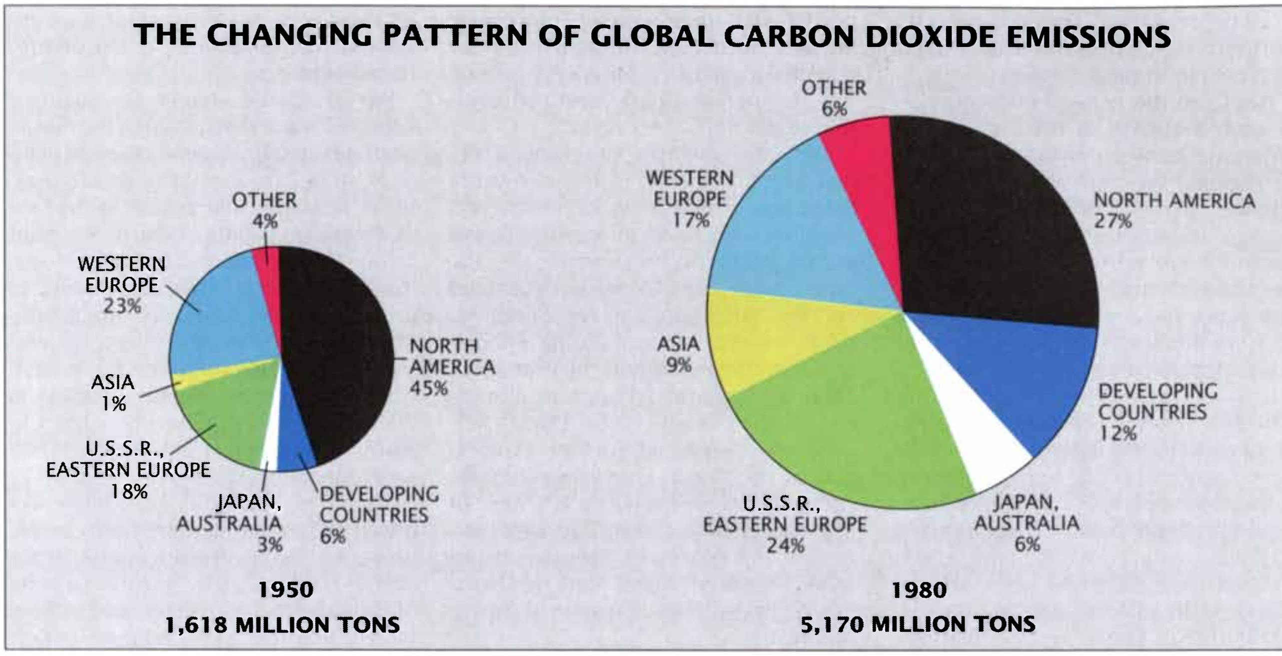
At this point, some influential atmospheric researchers, who believed that policy actions were beginning to outrun the scientific evidence, weighed in with their views. Richard S. Lindzen of the Massachusetts Institute of Technology and Jerome Namias of Scripps, the nation's most distinguished long-range weather-forecasting expert, wrote a letter to President Bush urging that no action be taken. Three other members of the NAS, including its former president Frederick Seitz, joined in a report, published under the auspices of the Marshall Institute, calling into question the scientific basis for policy actions. They recommended a major research program in mathematical modeling. They pointed out that there might be alternative explanations for the climate warming that had taken place. Thus, the great climate debate had been joined.

Meteorologists did not look with favor on the prospect of yet another public debate involving their field: they had been proved wrong many times before. As long ago as 1924, Sir Gilbert Walker, then head of the British government's Indian weather service, discerned unusually close connections between rainfall, temperature and pressures in the Pacific Ocean and the Indian subcontinent. Claims were made that the problem of forecasting the Indian monsoon was solved. Were it true, it would have been a great boon to Indian agriculture. But it was soon recognized that the correlations had little predictive power.

Later, in the 1940's and 1950's, widespread claims were made, based on the work of the late Irving Langmuir, Nobel laureate from the General Electric Company, and Vincent J. Schaefer of the State University of New York at Albany, that seeding clouds with dry ice or crystals of silver iodide could bring about an increase in rainfall. Several decades of research into the possi-



THE PROJECTED GROWTH OF GREENHOUSE GASES IN THE ATMOSPHERE



bilities of increasing rainfall, changing the intensity of hurricanes and modifying hailstorms by cloud-seeding techniques proved abortive.

Then, in the early 1980's, it was postulated that dust thrown into the atmosphere by a nuclear exchange between the Soviet Union and the U.S. would result in a "nuclear winter." This idea was deflated by Schneider and his colleague Stanley L. Thompson, who showed with the same models used in the prediction of climate that the "initial nuclear winter hypothesis can now be relegated to a vanishingly low level of probability."

Given this "cry wolf" history, it is not surprising that many meteorologists harbor deep reservations about taking costly actions on the basis of the predictions of a climate warming. But the push for policy has other constituents. Climate warming also unites those who are concerned about biodiversity and species extinction, economic development, human population growth, urban air pollution, acid precipitation and ozone depletion.

Political leaders stimulated by public concerns about environmental deterioration see these issues as important platforms and as springboards to public office. Those interested in increasing the competitiveness of American industry see greater energy efficiency as an important step toward that goal. It also serves the interest of those concerned about U.S. dependence on foreign energy sources. The issue of nuclear power is also underscored. Because fossil fuels are the main source of atmospheric carbon dioxide, strategies for stabilizing climate must envision non-fossil-fuel sources. Here at last is justification that proponents of nuclear power can forcefully advance to support expansion of nuclear power facilities throughout the world.

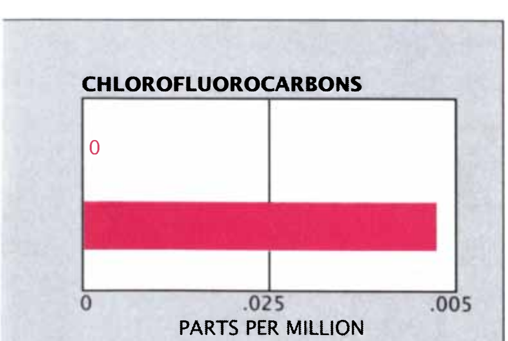
The issue of global climate warming also offers an opportunity for advancing the "new economic order" long advocated by Third World nations. International action will require technological and economic assistance to such nations if they are to participate in a global effort to reduce atmospheric pollution or arrest deforestation. In fact, the world faces the prospect that the greatest increases in emissions of carbon dioxide will occur in developing countries as their need for economic growth is followed by increased demand for energy.

In like manner, those interested in arresting population growth, especially in the Third World, point out that the climate-warming problem is probably not solvable as long as the number of human beings continues to rise. After all, it is people who consume natural resources and energy and who farm the land. Without population control, prospects for stabilizing the climate and arresting the deterioration of the habitability of the planet are abysmal.

Are the consequences of climate warming to be feared? People experience extremes of temperature in the natural course of events. The fact is that we do not know enough to predict the severity of the consequences. Because the warming would not be uniform over the surface of the earth, it would probably produce both winners and losers among regions and nations. Some parts of the earth would become warmer, some wetter and some drier. It is not possible on the basis of the evidence at hand to predict who would benefit and who would lose in such a global redistribution of so-called climatic resources.

Some aspects of global climate warming would be greatly beneficial in the view of agricultural researchers. Increased carbon dioxide will foster more active photosynthesis and enhance crop growth, to say nothing about the lowered plant requirements for water in a CO_2-enhanced atmosphere. In the words of Jerry D. Mahlman, director of the Geophysical Fluid Dynamics Laboratory, "The things we can say with confidence, the policymakers are not interested in. And the things [they] are interested in, we don't know with confidence."

Models do, however, agree that the polar regions of the world would undergo greater increases in temperature than would the tropics. Some of the projections of temperature increases in polar areas are startling in their magnitude, predicting as much as 18 degrees



F (10 degrees C) on the average in the Northern Hemisphere and only slight increases in tropical regions.

What are the general consequences of such a change in the temperature difference between equatorial and polar regions? We experience similar differences every year as the seasons change. In summer when arctic temperatures are warm, we do not suffer the great storms of winter; precipitation belts move farther north. Areas such as the southwestern part of the U.S. experience very dry conditions.

If arctic regions were to undergo significantly greater warming than equatorial regions and if precipitation belts were to move farther north, countries in the north temperate and polar zones would probably stand to benefit greatly. Their growing season would lengthen, and their precipitation would increase. With suitable soils, agriculture might thrive. These are speculations, however.

Such speculations are formulated in "scenarios" asking the question, What if? Unfortunately, an infinite number of such "what if" questions may be asked. What if the flow of rivers in the American Southwest, already fully utilized, were to be reduced by 20 percent? What if temperatures were to increase in the corn belt and precipitation were to move farther north? What, then, would be the consequences in the U.S. for agriculture, for resource availability, for energy generation, for national parks and conservation of nature?

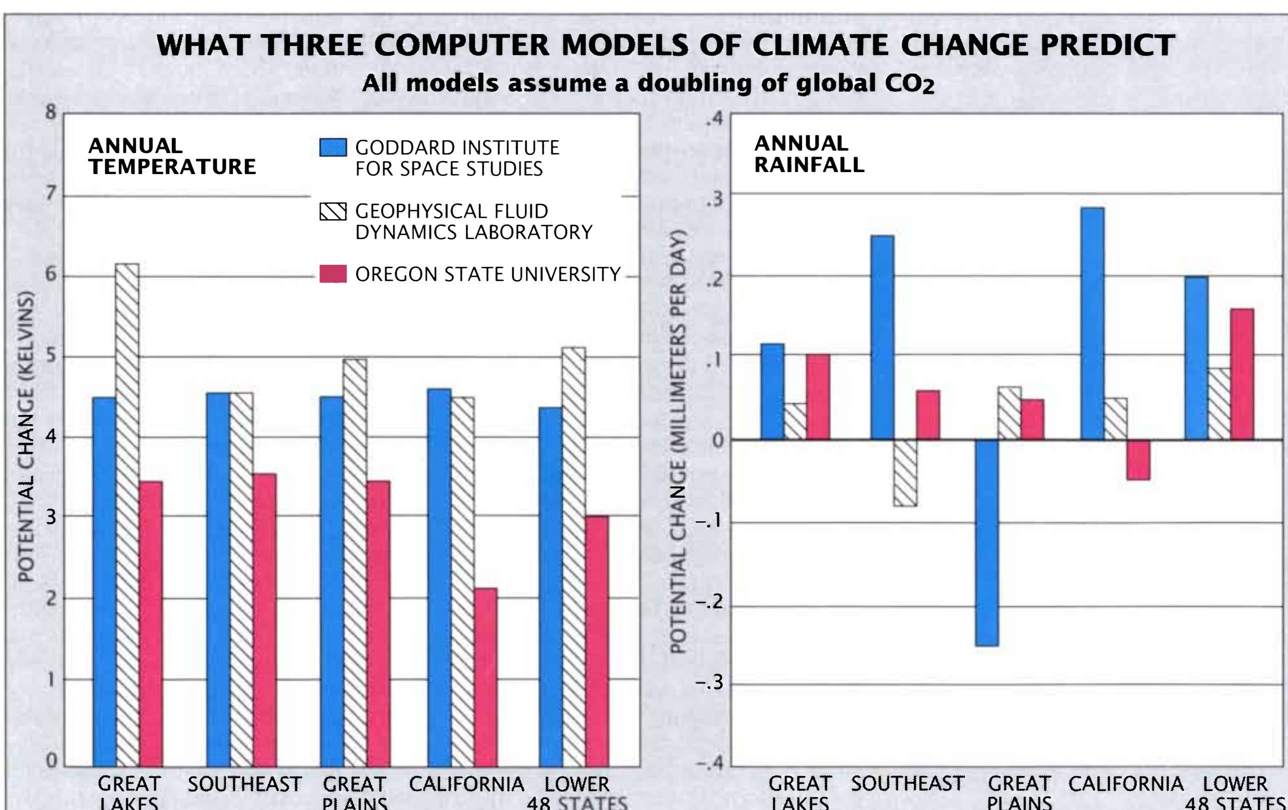
The consequences of changed climates can be seen in historical records going back thousands of years, and we have seen them in recent climatic events. We know, for example, that the Danes were able to settle Greenland and the Vikings to sail the North Atlantic to North America during a period of warm climate around the year 1000. Then a significant change in climate caused the collapse of the Danish settlements, prevented further exploration of the North American continent and ushered in the Little Ice Age. In just the past few years we have witnessed the effects of drought in the Sahel region of Africa and northeast Brazil, as well as in parts of North America.

Such scenarios can suggest apocalyptic possibilities. A recent film in the *Infinite Voyage* television series showed the U.S. Capitol under water as the result of one possible climate-warming scenario. Some foresee vast migrations of people as areas of the world become uninhabitable. Others see threats to national sovereignty and national security. President Gorbachev has stated that ecological security, not military security, will be the principal concern of all nations in the next century as environmental conditions cause disruptions worldwide.

But scenarios should be qualified with the caveat that although the events portrayed might in some cases be plausible, they are not real predictions. What, then, is a wise course in the face of great uncertainty? Clearly, it would be one that recognized uncertainty but would not permit that uncertainty to forestall action. Steps for which other economic and environmental reasons make sense would be taken first, whether or not a climate warming is taking place [see "The Changing Climate," by Stephen H. Schneider; SCIENTIFIC AMERICAN, September, 1989].

Then, as scientific knowledge reduced uncertainties, more costly measures could be taken if warranted, hence closely tying policy actions to the state of knowledge. Scientists and others have called this a "no regrets" policy. In gambling it would be known as "spreading your bets."

A recent report of the Council of Economic Advisers lends weight to this approach. It states that the cost of controlling carbon dioxide emissions and of taking other actions to address climate change would run into hundreds of billions of dollars. Because such reallocations of resources raise the specter of grave economic consequences, we need to be reasonably sure such ac-



tions are worth the cost. Any rational no-regrets policy would foster as one of its prime objectives adequate investment by governments in global monitoring and mathematical modeling to reduce the scientific uncertainties.

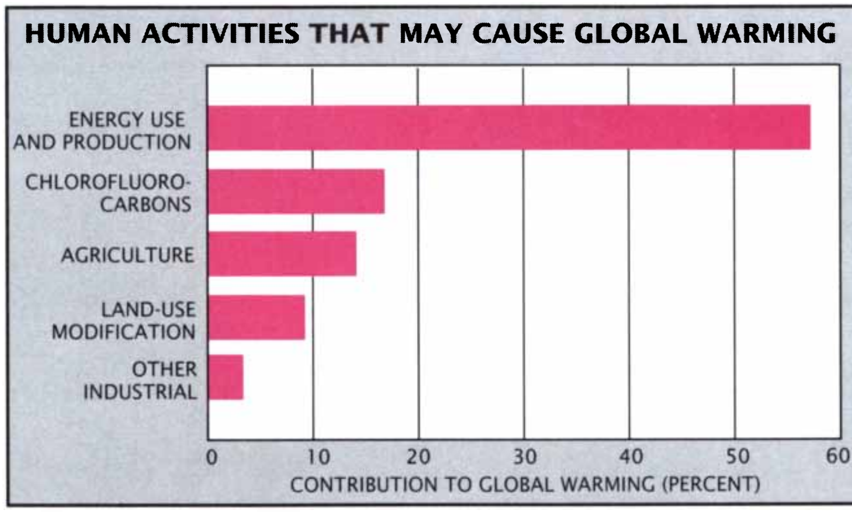
Where might we start? Energy conservation and efficiency along with the phaseout of CFC production would be the first priority for national and international action. Achieving greater energy efficiency justifies itself handily in economic terms. Increased energy efficiency would also ameliorate urban air pollution and acid precipitation. Shifts in the fossil-fuel mix from coal and oil to natural gas could significantly reduce carbon dioxide emissions per thermal unit. Technology is also available for more efficient power generation and for increased gasoline mileage.

Wisdom would also dictate major investments in nonfossil-energy sources. The circumstances favor significant new investments in passively safe, publicly acceptable nuclear power. Further development of forms of solar energy—photovoltaics or biomass, for example—makes good sense. Reforestation and forest preservation constitute a benign policy that yields many ecological and climatic benefits. Research aimed at producing stress- and disease-resistant crops would also be wise.

The public, however, must not be misled. These no-regrets initial policy steps will not solve the climate-warming problem. Their effectiveness will only modestly retard climate warming, and the future may require more drastic actions. No matter what policy actions we take, fully arresting the climate warming just does not seem to be in the cards.

The difficulties of reaching an international agreement on procedures for mitigating climate warming will be difficult and lengthy. The negotiations for the Law of the Sea Treaty provide a good lesson. They consumed 15 years and produced a treaty that the U.S. has yet to sign because of demands by developing nations for significant technology transfer and economic assistance. These issues will be even more complicated and more pervasive in any "Law of the Atmosphere Treaty."

It is likely that humanity will have to adapt to some climate changes. Modes of adaptation by society have not been well studied. Individuals, corporations and communities can adapt to climatic vicissitudes in myriad ways. Farmers can change crops, water use can be regulated and management practices can be altered.



HUMAN ACTIVITIES THAT MAY CAUSE GLOBAL WARMING

Other modes of adaptation would be needed if climate changes were severe. Sea-level rise, which is one of the predicted concomitants of a climate warming, might inundate low-lying coastal areas and cause salt water to intrude into freshwater bodies. Were this to occur, society would have to decide whether to invest in protective structures along coasts or adapt by changing land-use patterns. The North Sea dikes in the Netherlands are an outstanding example of adaptation to relative rise in sea level.

Some adaptations would take considerable time to implement. If the price of energy were to escalate, energy-efficient habitations would be necessary. Present cities with their great suburban sprawl are not energy efficient, and so we might return to more compact cities. If we chose to maintain agriculture in dry areas, society would need to decide whether to invest in the necessary irrigation systems. In fact, the economic growth of the entire western part of the U.S. has been based on major investments in water storage and transport for irrigation and industrial use.

Fortunately, time may for once be on our side. Governments generally act only when threats become real. They act in the face of military threats or when areas are endangered and destroyed by natural disasters. If the climate changes, the expectation is that it will do so gradually. We should be able to see the initial evidence of coastal inundation in an increasing frequency of high tides and in the undercutting of seacoasts. Climate warming itself should be evident in a rising frequency of heat waves or in other weather anomalies. The effects of a global climate warming are likely to take 30 to 50 years to become serious, and that is a long enough span in which actions to adapt to these changes should be possible.

What of the debate in the atmospheric, environmental and political communities? Our global environment is under attack on many fronts. Climate warming is but one, perhaps the most complex, of these issues. If the changes occurring in our atmosphere are likely to cause consequences, we must understand the problems and promote sensible policies to remedy them. What would be unwise is to lapse into apocalyptic thinking or ostrichlike denial. We like to believe ourselves far more sophisticated, more enlightened, than preceding generations. Until we can calmly and objectively approach our environmental challenges without promoting public hysteria and exciting shortsighted, self-interested reaction, we cannot claim that we are.

FURTHER READING

THERMAL EQUILIBRIUM OF THE ATMOSPHERE WITH A GIVEN DISTRIBUTION OF RELATIVE HUMIDITY. Syukuro Manabe and Richard T. Wetherald in *Journal of Atmospheric Science*, Vol. 24, pages 241-259; May, 1967.

CHANGING CLIMATE: REPORT OF THE CARBON DIOXIDE ASSESSMENT COMMITTEE. National Research Council, Carbon Dioxide Assessment Committee. National Academy Press, 1983.

THE GREENHOUSE EFFECT, CLIMATIC CHANGE, AND ECOSYSTEMS. Edited by Bert Bolin and Bo R. Doos. John Wiley & Sons, Inc., 1986.

GLOBAL TRENDS OF MEASURED SURFACE AIR TEMPERATURE. James Hansen and Sergej Lebedeff in *Journal of Geophysical Research*, Vol. 92, D11, pages 13345-13372; November 20, 1987.

THE GREENHOUSE EFFECT: SCIENCE AND POLICY. Stephen H. Schneider in *Science*, Vol. 243, No. 4892, pages 771-781; February 10, 1989.

Through the ages business and government have sought to keep private information private. Today highly specialized Contel communications networks assure that privacy.

Homeobox Genes and the Vertebrate Body Plan

This family of related genes determines the shape of the body. It subdivides the embryo along the head-to-tail axis into fields of cells that eventually become limbs and other structures

by Eddy M. De Robertis, Guillermo Oliver and Christopher V. E. Wright

Starting as a fertilized egg with a homogeneous appearance, an embryo made of skin, muscles, nerves and other tissues gradually arises through the division of cells. Long before most cells in the emerging body begin to specialize, however, a plan that designates major regions of the body—the head, the trunk, the tail and so on—is established. This plan helps seemingly identical combinations of tissues arrange themselves into distinctly different anatomical structures, such as arms and legs.

Recently embryologists have made great progress in uncovering the mechanisms that control this once mysterious process. In the past decade the powerful techniques of molecular biology have made it possible to isolate and characterize individual genes that mediate some of the developmental decisions involved in establishing the embryonic body plan. The key is a family of genes, known as homeobox genes, that subdivides the early embryo into fields of cells with the potential to become specific tissues and organs.

EDDY M. DE ROBERTIS, GUILLERMO OLIVER and CHRISTOPHER V. E. WRIGHT have collaborated for the past five years at the University of California, Los Angeles. De Robertis received his medical degree from the University of Uruguay in 1971 and his doctorate in biochemistry from the University of Buenos Aires in 1975. Between 1980 and 1985 he was professor of cell biology at the Biocenter of the University of Basel; since 1985 he has been the Norman Sprague Professor of Biological Chemistry at the U.C.L.A. School of Medicine. Oliver received his degree in biology from the University of Uruguay in 1980 and his degree in molecular biology from the University of Mexico in 1984. He is now associate professor of biochemistry at his alma mater in Uruguay. Wright received his doctorate in biochemistry from the University of Oxford in 1984. He is now assistant professor of cell biology at Vanderbilt University Medical School.

The development of *Xenopus laevis,* a South African clawed frog, stands as a good example of how a vertebrate takes shape. This amphibian is a favorite of modern embryologists, because at any time of year the female can be induced to lay about 1,500 large, easily fertilized eggs. Since all vertebrates develop similarly, most mechanisms of early frog embryogenesis also apply to fish, chickens, human beings and other animals.

One striking feature of early development is its rapid pace. A fertilized *Xenopus* egg cell divides into two after about 90 minutes. The cells then divide synchronously every 30 minutes until there are 4,000 of them. At that stage the embryo is called a midblastula and has the shape of a hollow sphere. To the unassisted eye, the cells all look identical, but some around the midblastula's equator are already destined to become a layer of cells called the mesoderm. The formation of the mesoderm is induced by protein growth factors released by the large yolky cells at the bottom pole of the embryo.

The entire mesodermal layer eventually moves into the interior of the embryo during a process called gastrulation. By the end of this process, three layers with distinct developmental potentials have been defined: the mesoderm, the endoderm and the ectoderm. The mesoderm gives rise to most of the body, including the vertebral column, the muscles, the bones and the body wall. The endoderm produces the epithelial layer of tissue lining the digestive tract as well as various other organs, such as the lungs, the liver and the pancreas. The ectoderm becomes the skin and the nervous system.

The ectoderm develops into the nervous system in response to chemical signals that diffuse out of the underlying mesoderm. The signals induce part of the ectoderm to thicken into a structure called the neural plate. (At this stage the embryo is termed a neurula.) The edges of the neural plate then fold toward one another while the middle sinks into the embryonic body. The edges finally fuse to form a neural tube, which becomes the basis for the brain and spinal cord.

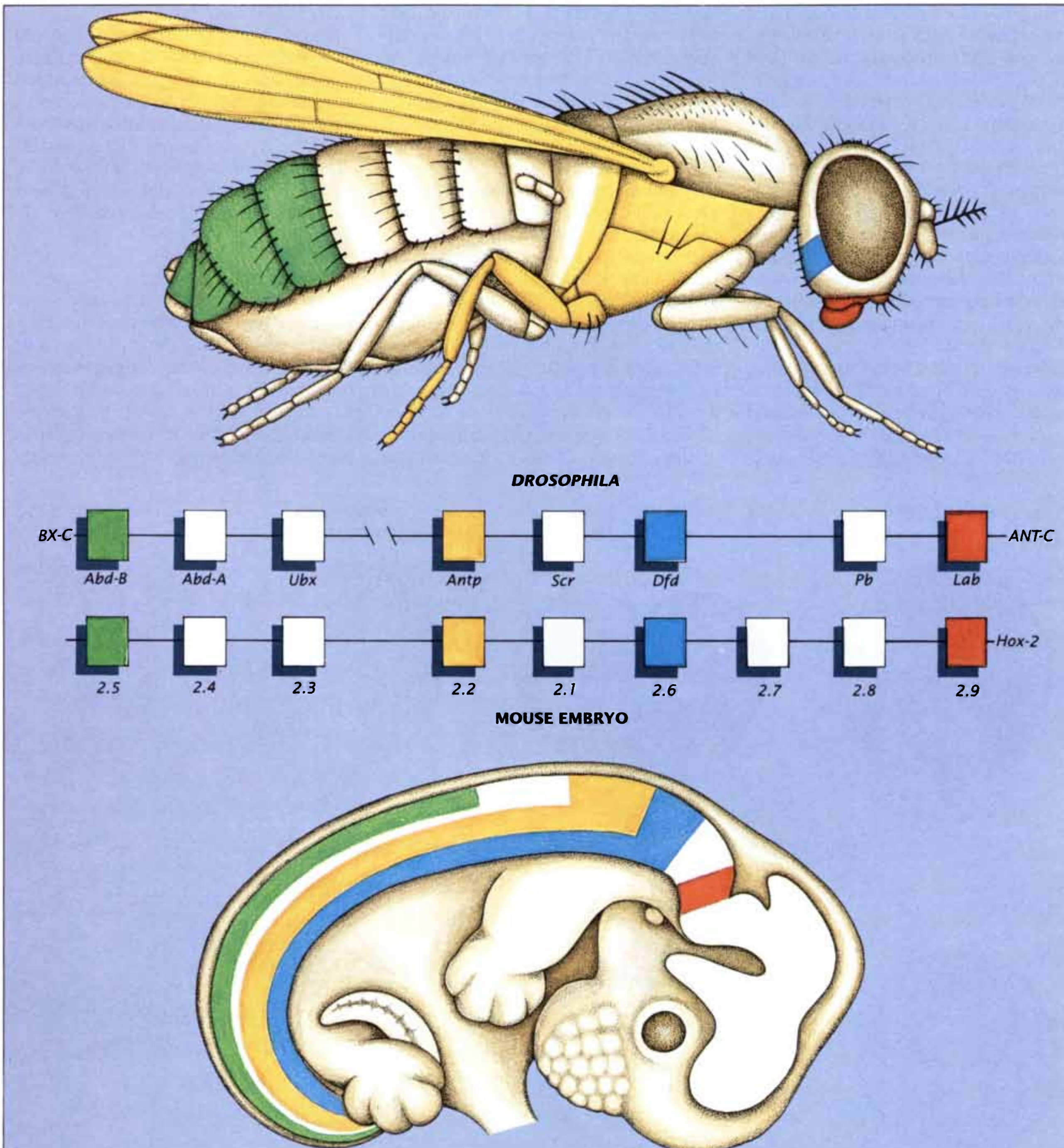
The determination of the embryo's anteroposterior (head-to-tail) axis is a milestone in development, because it provides the major line along which later structures will develop. Ross G. Harrison of Yale University was the first to show that embryonic cells commit to become limbs and other specific anatomical structures very soon after gastrulation is complete.

In 1918 Harrison took small fragments of mesoderm from the flanks of some amphibian neurulas and transplanted them into the sides of others. If the transplanted tissue came from a certain region of the donor, it always gave rise to an extra forelimb on the recipient. Harrison realized that although the mesoderm looked like a uniform layer in those embryos, the cells somehow already knew to which part of the body they belonged.

One peculiarity Harrison noted was that an embryo could still grow a forelimb even if he removed all the mesoderm that would normally give rise to it. He interpreted this finding to mean that the surrounding region of mesoderm also had a potential for inducing and directing limb growth. This broad region with the potential for expressing a structure became known as a morphogenetic field.

Following Harrison's lead, many scientists conducted transplantation experiments on amphibian embryos. Those studies established that the mesoderm is the crucial cell layer that specifies which end of the embryo is the head and which is the tail. The mesoderm of the amphibian neurula was mapped, or subdivided, into morphogenetic fields for many organs: gills, balancers (structures that help tadpoles stay upright), ears, forelimbs, hind limbs, the tail and so on.

Within each morphogenetic field, the potential for forming an organ varied gradually. It was therefore proposed that each morphogenetic field contained a "gradient-field" of infor-



HOMEOBOX GENES control development in animals as different as *Drosophila melanogaster* (a fruit fly) and a mouse. These genes divide the embryo along its head-to-tail axis into bands with different developmental potentials. The location of a homeobox gene on a chromosome corresponds to where it is expressed in the body: proceeding from left to right, the genes control body areas closer to the anterior end of the animal. All homeobox genes seem to have a common evolutionary origin. In this diagram, related homeobox genes in *Drosophila* and the mouse and the body parts they control in each animal are colored similarly. The mechanism that determines the head, trunk and tail may have arisen only once during evolution.

mation for specifying an organ. As we shall explain, the behavior of these gradient-fields corresponds closely with the patterns of expression for certain sets of genes.

Transplantation studies of the mesoderm's control over the amphibian body plan ended around the close of World War II and were replaced with genetic studies of how the body took shape. In 1948 Edward B. Lewis of the California Institute of Technology started an insightful genetic analysis (which he has continued to this day) of homeotic mutations in the fruit fly *Drosophila melanogaster.* A homeotic mutation causes a body part to be replaced with a structure normally found elsewhere on the body. For example, *bithorax* mutant flies have two pairs of wings instead of one; *Antennapedia* mutants have extra legs growing where their antennae should be.

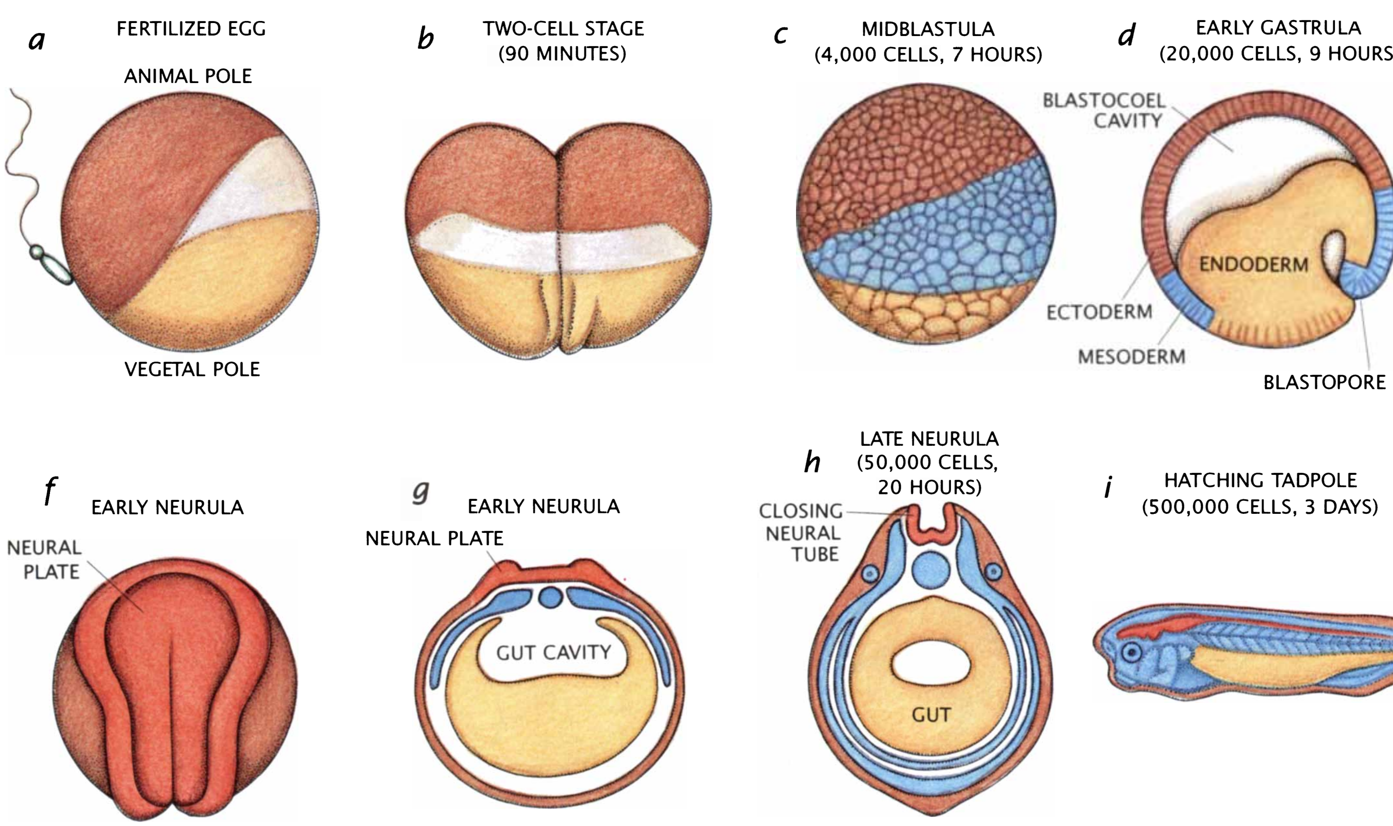
Lewis found that homeotic transformations could be caused by mutations in single genes, even though hundreds of active genes would be needed to create the abnormally placed wings and legs. It was reasonable to assume, then, that the mutations were affecting master genes that controlled the activity of many subordinate genes.

Once genetic engineering enabled scientists to isolate genes, the race to find and study the homeotic genes was on. During the early 1980's David S. Hogness and Welcome Bender of Stanford University became the first to isolate the genes *Ultrabithorax, Abdominal-A* and *Abdominal-B* in the *bithorax* complex. Walter J. Gehring and Richard L. Garber of the Biocenter at the University of Basel and Matthew P. Scott and Thomas C. Kaufman of Indiana University isolated the genes of the *Antennapedia* complex, including ones called *Labial, Proboscapedia, Deformed* and *Antennapedia*.

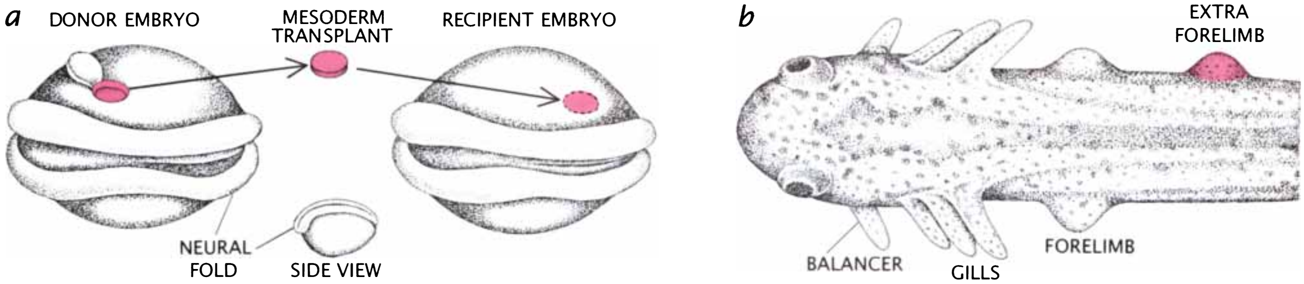
A crucial discovery came in 1983, when Gehring and his colleague William J. McGinnis found the *Antennapedia* gene contained a DNA sequence that was also found in another development-controlling gene. (Similar DNA sequences in different genes are said to be conserved.) Because conserved DNA sequences can hybridize, or bind, to one another, one could label the conserved DNA sequence from *Antennapedia* radioactively and use it as a probe to locate other genes containing the same region. In this way, Gehring and McGinnis isolated *Ultrabithorax, Deformed* and other homeotic genes. The conserved DNA region was identified independently by Scott, who was then at the University of Colorado at Boulder.

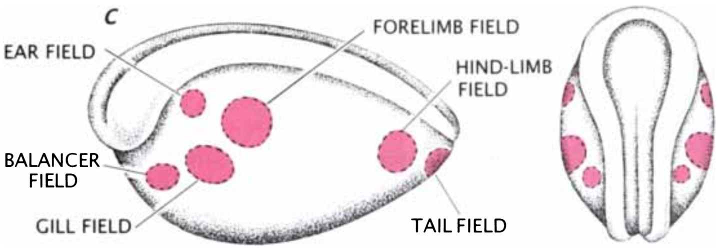
Significantly, McGinnis also showed that other invertebrates—such as centipedes and earthworms, from which insects are thought to have evolved—also contained the same conserved region of DNA. Clearly, the molecular structures of many genes known to control embryonic cell development were related. The conserved DNA region in each homeotic gene was dubbed the homeobox.

The homeobox encodes a sequence of 60 amino acids that is very similar in the protein products of most homeotic genes. That sequence in a protein is known as the homeodomain. Its function is to recognize and bind to specific DNA sequences in those genes regulated by the homeotic genes [see "The Molecular Basis of Develop-



VERTEBRATE BODY PLAN is generated through the chemically induced formation and movement of cell layers, as seen in the development of *Xenopus laevis,* a South African frog. Through rapid cell divisions (*a–c*), a fertilized egg becomes a hollow ball of cells. The large yolky cells at the bottom pole of the embryo release protein growth factors that induce the overlying cells to become the mesoderm layer (*blue*). The mesoderm is the critical layer that determines the embryo's anteroposterior polarity. Two other layers of cells—the ectoderm (*brown*) and the endoderm (*yellow*)—are established during gastrulation, the process by which the mesoderm migrates into the interior of the embryo (*d–e*). During the neurula stage (*f–g*), the meso-

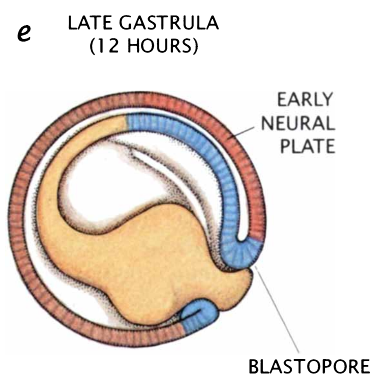




TRANSPLANTS in which mesoderm from one neurula-stage embryo is grafted onto another will induce the formation of additional limbs or organs. If mesoderm from the forelimb area is transplanted, for example (*a*), the recipient embryo will have an extra forelimb (*b*). Through such experiments on amphibians, embryologists have identified morphogenetic fields that specify the development of various structures (*c*).

ment," by Walter J. Gehring; SCIENTIFIC AMERICAN, October, 1985].

The polypeptide chain in the homeodomain consists of four helixes, one of which is responsible for recognizing a specific DNA sequence. Because this helix is nearly the same in all homeodomain proteins, the proteins all bind to fairly similar DNA sequences. When they bind to genes in a cell, homeodomain proteins activate or repress the expression of those subordinate genes.



derm induces part of the ectoderm to become the neural plate (*red*). As seen in cross section (*h–i*), the neural plate closes on itself and becomes a neural tube, which is the forerunner of the brain and the spinal cord in the mature animal (*j*).

We began research on the homeobox in 1983, when one of us (De Robertis) had his laboratory at the biocenter on the same floor as Gehring. We had been interested for some time in the development of *Xenopus laevis* [see "Gene Transplantation and the Analysis of Development," by Eddy M. De Robertis and J. B. Gurdon; SCIENTIFIC AMERICAN, December, 1979]. As we followed the great advances being made in studies of *Drosophila*, it became evident that we would have to identify master genes in vertebrates if we were ever to gain a comparable understanding of their development.

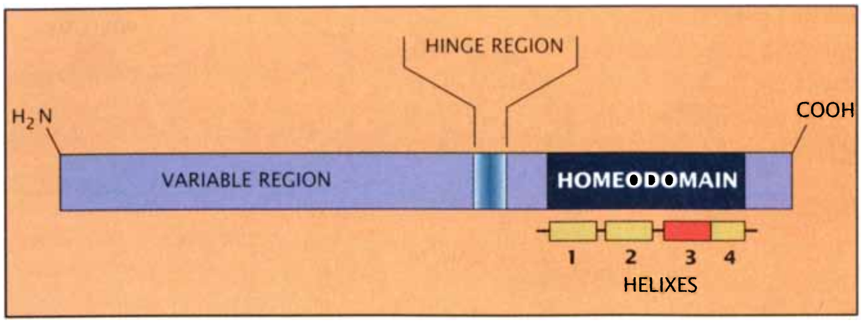
Yet the lack of knowledge about frog genetics seemed to be an insurmountable barrier to further progress. Even though the genetics of mice had been reasonably well studied, there were no real candidates for master genes controlling embryogenesis.

We decided to try what seemed, at the time, a crazy experiment: to isolate a gene similar to *Antennapedia* from frog DNA with McGinnis and Gehring's fruit fly homeobox probes. There was little reason to believe that the frog DNA contained such a gene or that the genes of such unrelated species would be significantly similar. Still, we felt it was worth the attempt. Some of our colleagues were skeptical that such an experiment could ever work, and two of our students declined to help on those grounds.

We were soon celebrating with a bottle of champagne. On our very first attempt, while working with Andrés E. Carrasco, a postdoctoral student in our laboratory, we succeeded. We analyzed the DNA sequence in the frog gene that our experiment had isolated, which is now called *XlHbox 1*, and confirmed that it contained the homeobox region. That finding strongly suggested that a gene directly controlling vertebrate development might at last be at hand.

Little did we imagine after our first experiment that it would take six more years and the efforts of laboratories throughout the world before it was certain that vertebrate homeobox genes were directly involved in the control of development. Even so, initial progress was swift in studies of mammals. Working with mice, Frank H. Ruddle of Yale University (who was then on sabbatical at the biocenter) and Peter Gruss of the Max Planck Institute for Experimental Medicine in Göttingen, West Germany, isolated many genes containing homeoboxes. Dado Boncinelli of the University of Naples had similar success with human genes. The proteins encoded by all these homeobox genes differ greatly from one another except at the highly conserved homeodomain.

The roster of proteins known to contain homeodomains grew in 1988, when researchers purified transcription factors for the first time. These factors are proteins that increase the expression of particular target genes. When the transcription factors were sequenced, some were found to contain homeodomains, which indicated that they were products of genes with homeobox regions. These biochemical



HOMEODOMAIN PROTEINS bind to DNA and regulate gene expression. They are composed of a variable region, which determines a protein's specific activity, a small connective hinge region and a homeodomain, a 60 amino acid sequence that is similar in all proteins of this type. This sequence is encoded by the homeobox regions of genes. The homeodomain consists of four alpha helixes (*1–4*), one of which (*red*) recognizes and binds to a specific DNA sequence in the target genes.

studies independently confirmed that homeobox genes regulate the activity of other genes.

But how do homeobox genes orchestrate cellular differentiation during development? An inkling of the answer comes from observing the regions in which the proteins made by homeobox genes are located in the embryonic body at various developmental stages. The *Xenopus* XlHbox 1 protein, for example, is found in a narrow band of cells just behind the frog embryo's head. This band consists of both the mesoderm and the anterior spinal cord and neural crest.

The anterior and posterior boundaries of *XlHbox 1* expression in these tissue layers are neatly aligned. Because mesoderm is known to induce the anteroposterior characteristics of neural tissue, it seems possible that the mesoderm expressing *XlHbox 1* also induces cells in the overlying neural plate to express the gene as well.

Other homeobox genes are active in different regions. On the basis of homeobox gene expression patterns, therefore, one can view the vertebrate embryo as subdivided into anteroposterior fields of cells with different developmental capacities. This subdivision of the embryonic body precedes the formation of specific organs or structures.

Even though the homeodomains encoded by different homeobox genes are very similar to one another, characteristic differences in their amino acid sequences can be used to identify them. Some homeodomains resemble one another much more closely than others. Interestingly enough, on the basis of these similarities and differences, some mammalian homeodomains strongly resemble those produced by particular fruit fly genes.

When patterns of expression for many homeobox genes in mouse embryos were analyzed, a remarkable observation was made independently by Robb Krumlauf of the Medical Research Council in London and Denis Duboule of the European Molecular Biology Laboratory in Heidelberg. Investigators had previously shown that in both vertebrates and invertebrates, homeobox genes cluster in complexes, or groups, on a chromosome. In other words, the homeobox genes are arranged in a precise order, left to right, on the linear DNA molecule that makes up a chromosome.

Krumlauf and Duboule made the unexpected discovery that in mice the order of the homeobox genes in a cluster corresponds directly to where the genes are expressed. Homeobox genes located near the left end of a complex are expressed in posterior parts of the body and genes to the right are expressed closer to the head. Lewis had noticed the same pattern in *Drosophila* many years earlier.

All vertebrates have four homeobox complexes, each located on separate chromosomes. These complexes probably arose during evolution through duplications of the single cluster of homeobox genes in invertebrates. Consequently, every human being has four genes that resemble the fruit fly gene *Abdominal-B,* for example, and four others that resemble *Deformed.*

One unifying principle applies to all homeobox complexes: genes expressed posteriorly are located at the left, and those expressed anteriorly are at the right. Homeobox genes are therefore arranged in the chromosomal DNA in the same order in which they are expressed along the anteroposterior body axis. This extraordinary arrangement may have come about because homeobox genes must be activated in a particular order.

Evidence for how this sequential deployment of homeobox genes occurs is accumulating. In vertebrate embryos, retinoic acid (a compound related to vitamin A that can sometimes cause severe birth defects) and peptide growth factors are good candidates for providing such positional clues. They could convey such information

Consensus
```
RKRGRTTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN
```
Labial
```
NNS---NF-NK-LT--------------A------NT-Q-N-T-V----------Q--RV
PGGL--NF-TR-LT---------K--S-A--V---AT-G-N-T-V----------Q--RE
```
Hox-2.9

Deformed
```
P--Q--A---H-I--------Y--------------T-V-S-----------------D-
P--S--A---Q-V--------Y---------V--------S-----------------DH
```
Hox-2.6

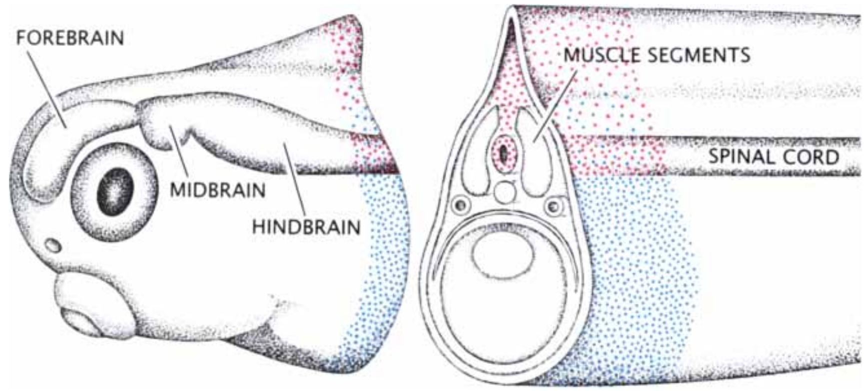
Antp
```
-----Q------------------------------------------------------
-----Q---------------Y--------------T-----------------------
```
Hox-2.3

Abd-B
```
VRKK-KP-SKF---------L--A-VSKQK-W-L-RN-Q-----V----------N--NS
SRKK-CP--K----------L--M----D--H-V-RL-N-S---V----------M--L-
```
Hox-2.5

ALL HOMEODOMAINS are fundamentally similar, but homeodomains made by some insect and mammalian genes are particularly alike. Amino acid sequences in several homeodomains are shown here. The genes *Labial, Deformed, Antp* and *Abd-B* are from *Drosophila* fruit flies, and the four analogous *Hox* genes are from mice. In each sequence the letters stand for amino acids. A hyphen indicates that the amino acid is the same as in the consensus string, an average of all homeodomain sequences.

by activating homeobox genes selectively in the mesoderm, the key element in determining the body plan.

By adding retinoic acid to cultured embryonic cells, Boncinelli's research group has shown that the compound can activate many homeobox genes. In frog embryos, Douglas A. Melton of Harvard University has proved that fibroblast growth factor (which induces the formation of the mesoderm in early embryos) can activate posterior homeobox genes selectively. In our laboratory at the University of California at Los Angeles, Ken W. Y. Cho has shown that a protein resembling transforming growth factor β activates only anterior genes.



HOMEOBOX GENES are expressed in discrete bands along the anteroposterior axis of an embryo. In a *Xenopus laevis* tadpole, for example, the *XlHbox 1* gene is expressed in a region in the anterior trunk of the body. The protein produced by the gene is found in the cell nuclei of both mesodermal (*blue*) and ectodermal (*red*) tissues. The forelimb grows entirely from mesoderm cells expressing the XlHbox 1 protein.

Once activated, do homeobox genes directly specify the identities and fates of embryonic cells, thereby shaping the body and guiding the formation of organs? Or are their effects indirect? The results of two experiments argue for their having a direct role.

In the first experiment, we injected antibodies directed against the XlHbox 1 protein into single-cell *Xenopus* embryos. The antibodies bound to the protein and inactivated it during the crucial period in which the body plan is established. When we examined the tadpoles that developed, we discovered the tissues that normally expressed *XlHbox 1* and that should have become a section of the anterior spinal cord had instead become hindbrain structures. In effect, the "loss-of-function" of *XlHbox 1* changed part of the spinal cord into a more anterior structure.

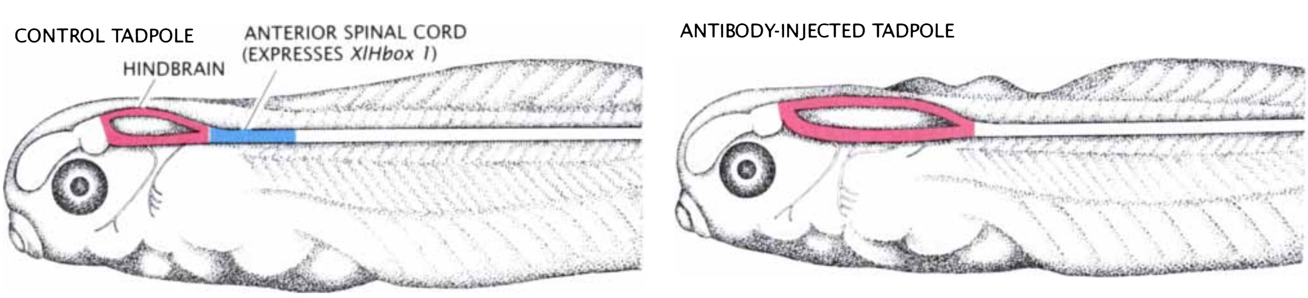
In the second experiment, Gruss and Michael Kessel of the Max Planck Institute injected DNA containing a mouse homeobox gene into mouse embryos. The piece of DNA was designed so that the homeobox gene would be expressed throughout the body, even in regions where it normally would not, such as the head and neck. The resulting mice frequently had severe head defects, such as cleft palates. Even more interesting, they also had an extra vertebra and intervertebral disk at the base of the cranium, and some had an extra pair of ribs in the neck region. In this way, the "gain-of-function" of a homeobox gene induced homeotic transformations precisely like those observed in fruit fly mutations.

Other work also suggests a role for homeobox genes in specifying cell identity. As previously described, homeobox genes are strongly expressed in bands along the anteroposterior axis early in development. Later, when organs are forming in these regions, the same homeobox genes are once again expressed intensely. At these later stages, homeobox genes seem to provide molecular tags that remind cells of where in the embryo they originated.

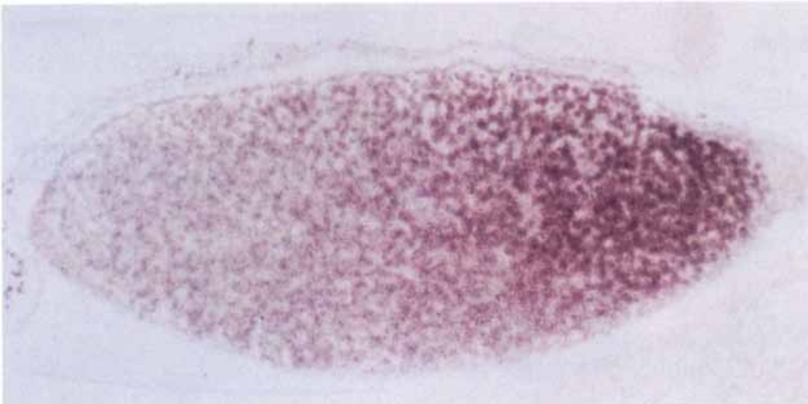
The development of the forelimb is a particularly informative case. The entire forelimb field is derived from the band of mesoderm that expresses *XlHbox 1*. Cells proliferate in the band and form a small forelimb bud that appears on *Xenopus* by the third week after fertilization.

At this stage the forelimb bud mesoderm appears uniform, but it contains a gradient of XlHbox 1 protein. That is, the protein is most abundant in cell nuclei along the anterior side of the limb bud—the side that gives rise to the thumb—and least abundant in nuclei on the posterior side, which gives rise to the smallest digit. As the limb extends and takes shape, the concentration of XlHbox 1 protein stays highest near the shoulder, at the proximal end of the arm. In contrast, the protein from another gene, *Hox 5.2*, establishes a gradient that is highest along the posterior side and the distal end of the limb—a pattern precisely the reverse of that for *XlHbox 1*.

Gradients of XlHbox 1 and Hox 5.2 proteins can be detected in frog, chicken and mouse embryos. Other homeobox genes are also involved in forelimb development: Duboule has identified three other homeobox genes adjacent to *Hox 5.2* that turn on sequentially as the limb tip extends. The order in which the genes are ac-



INHIBITION of a homeodomain protein can alter the developmental fate of embryonic tissues. In normal tadpoles (*left*), XlHbox 1 protein is expressed in a defined region of the anterior spinal cord (*blue*). If antibodies against XlHbox 1 protein are injected into a one-celled embryo, then in the resulting tadpole (*right*), that region is transformed into hindbrain (*red*).



GRADIENTS of homeodomain proteins can be seen in these stained limb buds. In the chick wing bud (*left*), the concentration of Hox 5.2 protein is highest in the cell nuclei near the posterior region (*right edge*). In the pectoral fin bud of a zebrafish (*right*), the concentration of XlHbox 1 protein is highest in the anterior region (*left edge*). The fish pectoral fin is the evolutionary precursor of the tetrapod forelimb. Gradients are efficient mechanisms for conveying positional information.

tivated corresponds to their order in the DNA.

Gradients of proteins or other molecules are good signposts for specifying the positions of cells and efficiently directing their fates. For example, cells in a limb bud can form separate digits by responding differently to varying amounts of a single protein. It would be less economical if a distinct protein had to specify each digit.

In conclusion, analysis of homeodomain protein gradients during limb development reveals that the same set of homeobox genes that establishes the head-to-tail axis is used again later to specify the positions of cells during limb development. Homeodomain proteins are found in cell nuclei, as would be expected of DNA-binding proteins that turn genes on and off. How gradients of nuclear proteins are established in limb buds is not yet known. Intercellular communication signals, similar to those involved in axis formation and perhaps mediated by growth factors or retinoic acid, are probably involved.

In addition to what the study of homeobox genes has explained about embryonic development, it has provided insights into evolution. Because the order of homeobox genes is similar in vertebrates and invertebrates, the first homeobox complexes must have evolved eons ago in flatworms or other primitive organisms that were the common ancestors of both human beings and insects. It would be interesting to know whether the most primitive multicellular organisms with anteroposterior polarity, such as rotifers, also have homeobox gene complexes. The amazing conservation of the complexes throughout evolution suggests that once an efficient way of specifying the anteroposterior axis was found, it was easier to produce new body shapes by modifying that system than to develop entirely new strategies.

Homeobox gene activity also offers clues about how specific anatomical structures might have evolved. Scientists have long wondered, for example, about the origin of the arm, or forelimb, in tetrapods. Because the primitive fish called coelacanths have pectoral fins with bony joints, investigators assumed for many years that the arm evolved from pectoral fins.

Support for this theory has come out of work by Anders Molven and Charles B. Kimmel of the University of Oregon. In zebrafish embryos, *XlHbox 1* is first expressed in a circular region of the lateral mesoderm corresponding to the pectoral fin field. At this stage the expression of the gene corresponds exactly with a morphogenetic field defined by Harrison in 1918.

As the cells proliferate, XlHbox 1 protein forms a steep gradient in the pectoral fin bud, similar to the gradients in frog, chicken and mouse forelimb buds. This pattern suggests that *XlHbox 1* is an ancient gene, whose function in the limb gradient-field antedates the appearance of tetrapod structures such as digits. Much can probably be learned by reexamining the comparative embryology of vertebrates at the level of gene expression.

Although it will take a long time to understand exactly how genes cooperate to organize cells from an apparently homogeneous egg into a swimming tadpole, molecular analysis of vertebrate development has already made a great leap forward. The expression of homeobox genes may provide a molecular explanation for the gradient-fields recognized by experimental embryologists many decades ago. The genes that control the anteroposterior axis are conserved in the zoological spectrum to a degree beyond investigators' wildest expectations. Molecules that may be involved in transmitting positional information, such as retinoic acid and growth factors, have been identified. Possibilities are now open for analyzing how body shape changes during the course of evolution. Students starting work in laboratories today may one day be able to answer simple questions, such as what makes an arm different from a leg. What a good time to begin!

FURTHER READING

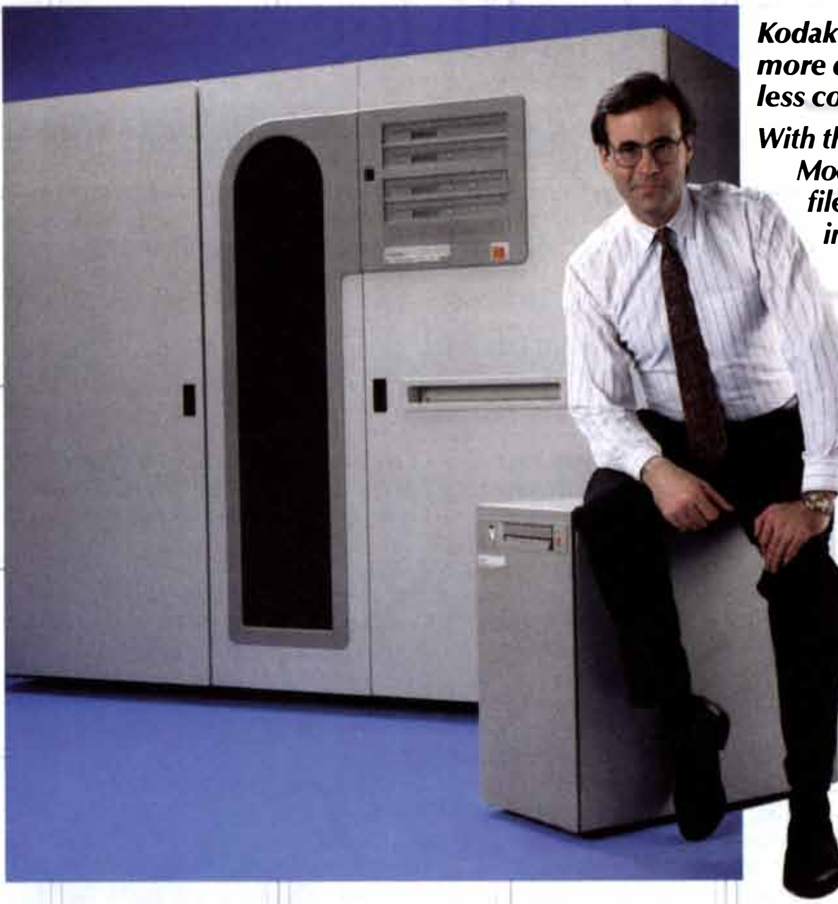
CLONING OF A X. LAEVIS GENE EXPRESSED DURING EARLY EMBRYOGENESIS CODING FOR A PEPTIDE REGION HOMOLOGOUS TO DROSOPHILA HOMEOTIC GENES. Andrés E. Carrasco et al. in *Cell*, Vol. 37, No. 2, pages 409-414; June, 1984.

A GRADIENT OF HOMEODOMAIN PROTEIN IN DEVELOPING FORELIMBS OF XENOPUS AND MOUSE EMBRYOS. Guillermo Oliver et al. in *Cell*, Vol. 55, No. 6, pages 1017-1024; December 23, 1988.

THE MURINE AND DROSOPHILA HOMEOBOX GENE COMPLEXES HAVE COMMON FEATURES OF ORGANIZATION AND EXPRESSION. Anthony Graham et al. in *Cell*, Vol. 57, No. 3, pages 367-378; May 5, 1989.

INTERFERENCE WITH FUNCTION OF A HOMEOBOX GENE IN XENOPUS EMBRYOS PRODUCES MALFORMATIONS OF THE ANTERIOR SPINAL CORD. Christopher V. E. Wright et al. in *Cell*, Vol. 59, No. 1, pages 81-93; October 6, 1989.

VARIATIONS OF CERVICAL VERTEBRAE AFTER EXPRESSION OF A *HOX 1.1* TRANSGENE IN MICE. Michael Kesset et al. in *Cell*, Vol. 61, No. 2, pages 301-308; April 20, 1990.



The LEP Collider

The Large Electron-Positron Collider will churn out a million Z^0 particles this year. The abundance of data will enable physicists to probe their present understanding of physical law as never before

by Stephen Myers and Emilio Picasso

Although it was midafternoon on July 14, 1989, the bicentennial of the French Revolution, engineers and scientists crowded into a control room at the European laboratory for particle physics (CERN) in Prevessin, France, just across the border from Geneva, Switzerland. Word had got out that the laboratory's Large Electron-Positron (LEP) Collider was ready to take its first beam, and people had flocked there to witness the historic event. CERN's director-general, Carlo Rubbia, and his predecessor, Herwig F. Schopper, who had actively supported the LEP through its early phase, were among the anxious onlookers.

At five minutes to four, the first cluster of positrons completed its maiden circuit around the giant, ring-shaped accelerator. Every one of the hundreds of thousands of components along the 26.7-kilometer circumference of the machine was working flawlessly.

One month later in the same, but less crowded, control room, workers toiled late into the evening, preparing to bring two counterrotating beams, one of electrons and one of positrons, into collision for the first time. At five minutes to midnight, the electrostatic plates that steer the beams apart were shut off, and the beams were allowed to collide at the centers of the four gigantic LEP detectors: Aleph, Delphi, L3 and Opal.

Ten suspenseful minutes later, the control-room telephone rang. It was the Opal detector group calling to announce the clear sighting of the first Z^0 particle (pronounced "Z zero" or "Z naught") produced by the LEP. During the time it took to carry a photocopy of the event to the control room, the telephone rang to announce five more possible events. Amid the ensuing euphoria, Rubbia, who shared a 1984 Nobel prize for his role in discovering the Z^0, remarked that he had gone to Stockholm on the strength of only five events.

In the four months that followed, the LEP detectors observed more than 100,000 Z^0 events. Already these observations have yielded a major result: evidence that there are probably only 12 elementary particles, organized into three "generations" [see "Science and the Citizen," SCIENTIFIC AMERICAN, December, 1989]. All in all, some 30 articles have been published based on the Z^0 events observed last year.

This year the LEP is expected to generate about a million more Z^0's. This abundance of data will enable physicists to investigate rare decay modes, determine the precise mass and lifetime of the Z^0, study the processes underlying the production of hadrons (particles made up of quarks) and probe subtle aspects of the prevailing theory. Equally important, experimenters at the LEP will try to discover new states of matter—in particular, the Higgs boson and perhaps also the top quark. Both of these massive particles are predicted by theory but have yet to be observed. Experimenters will also be searching for phenomena that are not predicted by existing theories.

The LEP Collider is destined to be the preeminent tool for particle physics in the next decade. Its spectacular performance to date is a tribute to the ingenuity and determined effort of hundreds of technicians, engineers and scientists from more than 25 nations who collaborated on the LEP and its detectors.

Seen from the air, the LEP is a large, nearly circular octagon that straddles the Franco-Swiss border. The ring consists of eight straight sections, each 500 meters long, joined by eight arcs 2.8 kilometers long. The accelerator components are in a subterranean tunnel about four meters wide at a depth, on average, of 100 meters.

With an average diameter of 8,486 meters and a beam-collision energy of 100 billion electron volts, or gigaelectron volts (GeV), the LEP is the largest and most powerful electron-positron storage ring ever built. It marks the culmination of a history that began 30 years ago, in March, 1960, when Bruno Touschek gave a seminar at the Italian National Laboratory in Frascati, in which he argued the importance of studying collisions between electrons and their antimatter counterparts, positrons. He proposed an accelerator in which beams of electrons and positrons of the same energy would be directed in opposite directions around a circular track and collided head-on. The detritus from the collisions would be studied to gain insight into the fundamental laws of nature.

The first electron-positron storage ring was built at Frascati and christened ADA (*anello di accumulazione*). The ADA measured 1.6 meters in diameter and produced a maximum beam

STEPHEN MYERS and EMILIO PICASSO are at the European laboratory for particle physics (CERN) near Geneva. Myers received a Ph.D. in 1972 from Queens University in Belfast and joined CERN in the same year. Since 1979 he has played a key role in major aspects of the LEP's design, building and commissioning. He is now deputy division leader of the Super Proton Synchrotron–LEP division. Picasso obtained his doctorate in physics from the University of Genoa in 1956 and joined CERN in 1964; at that time he worked on precision tests of quantum electrodynamics and special relativity. From 1981 to 1989 he served as director of the LEP project.

LARGE ELECTRON-POSITRON storage ring 26.7 kilometers in circumference lies in a tunnel 100 meters underground (drawing not to scale). Counterrotating bunches of electrons and positrons collide at energies near 100 billion electron volts. Shown at the bottom, from left to right, are the beam-focusing magnets, a tram system inside the tunnel and a row of radio-frequency cavities, which power the beams.

energy of 250 million electron volts (MeV). This pioneering machine was followed by many larger and more energetic electron-positron storage rings in the U.S.S.R., France, the U.S., Italy, West Germany and Japan.

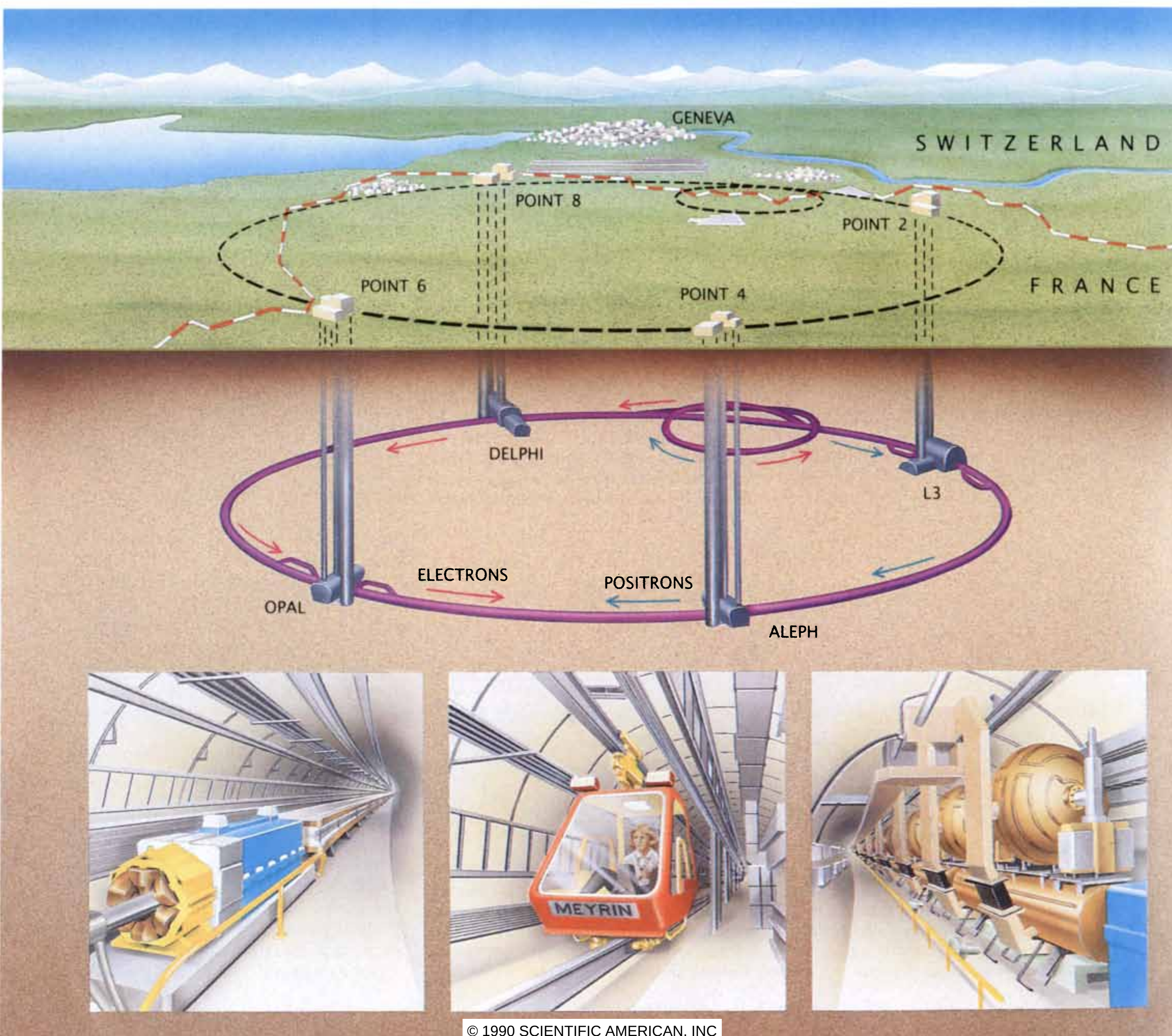
In all these machines the goal is to create new particles from the energy liberated by the annihilation of the electron by its antiparticle. In accordance with Einstein's fundamental formula, $E = mc^2$, the greater the energy released, the more massive the particles created can be. Because the electrons and positrons have the same energy and are colliding head-on, the total energy available in the center of mass of the system is simply twice the beam energy. The LEP, with present collision energies of up to 110 GeV, generates Z^0 particles. It will eventually be upgraded to 200 GeV and generate particles known as the W^+ and W^-. Each of these three particles weighs about 90 GeV, or about 100 times the mass of the proton.

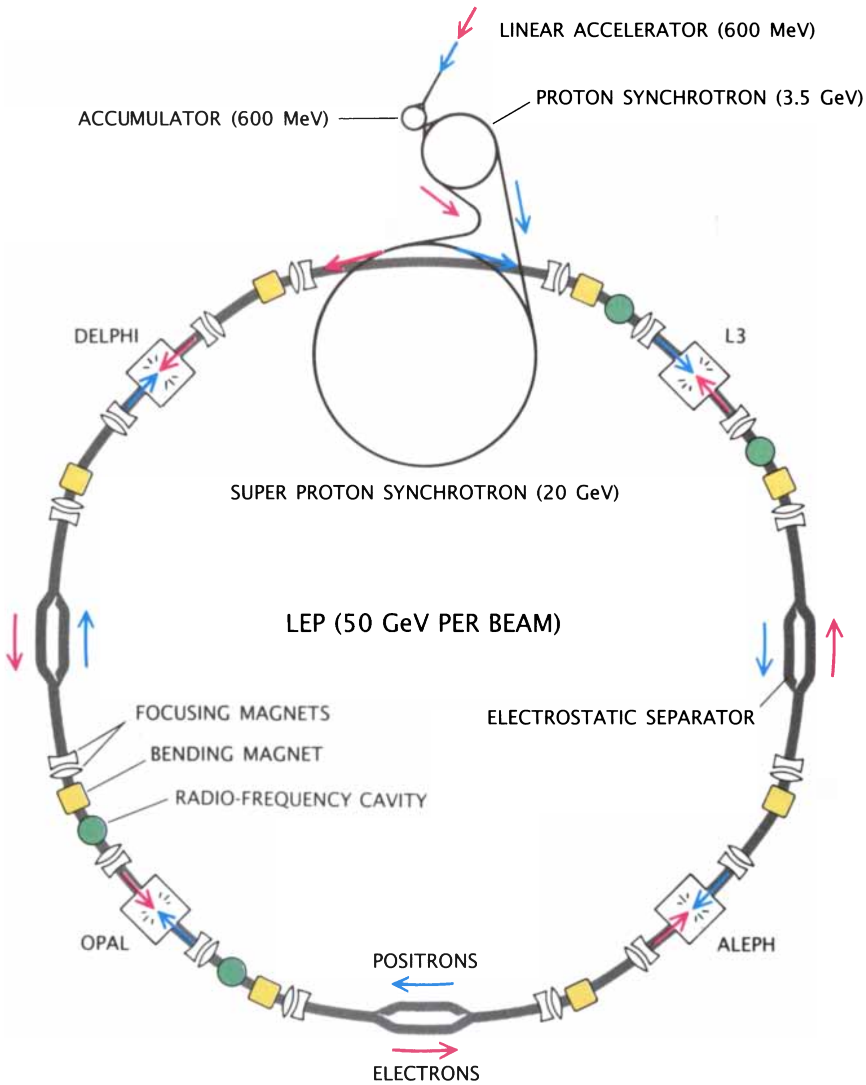
To understand the significance of these particles, we must refer to the prevailing theory of matter and forces, called the Standard Model. According to the theory, there are two basic types of particles: fermions and bosons. Fermions are the basic building blocks of matter. Bosons are "carriers of force": they are exchanged between fermions, generating attraction and repulsion among different types of fermions.

There are two classes of fermions: leptons and quarks. Leptons consist of charged particles, such as electrons and muons as well as uncharged, virtually massless particles called neutrinos. Quarks are constituents of hadrons, such as the familiar proton and neutron.

Fermions are arranged in generations, each consisting of two quarks and two leptons. The first generation consists of the up and down quarks (which make up protons and neutrons), the electron and the electron neutrino. The second generation is occupied by the charm and strange quarks, the muon and the muon neutrino. The third generation contains the bottom quark and top quark (not yet observed), the tau and the tau neutrino.

According to the Standard Model, the four fundamental forces of nature are said to be "carried" by bosons. Electromagnetism is carried by photons (particles of light). The strong force, which binds protons and neutrons into atomic nuclei, is carried by gluons. Gravity





BIRD'S-EYE VIEW of LEP storage ring shows how electron and positron bunches are accelerated by a series of smaller machines. Four bunches each of electrons and positrons circulate in opposite directions and collide inside the four gigantic detectors: Aleph, Delphi, L3 and Opal. The bunches also cross paths at four intermediate sites, where they are prevented from colliding by electrostatic separator plates.

	Z^0 MASS (GeV)	Z^0 WIDTH (GeV)	NEUTRINO FAMILIES	TOP QUARK	HIGGS BOSON	SELECTRON	WINO
				LOWER MASS LIMIT (GeV)*			
ALEPH	91.182 ± .056	2.541 ± .056	3.01 ± .20	45.8	24.0	44.2	46.0
DELPHI	91.171 ± .060	2.511 ± .065	2.97 ± .26	44.0	14.0	42.0	45.0
L3	91.160 ± .054	2.539 ± .054	3.29 ± .17	SEE CAPTION		41.0	44.0
OPAL	91.154 ± .051	2.536 ± .045	3.09 + .25 − .31	44.5	25.0	43.4	45.0

*95 PERCENT CONFIDENCE LIMIT

RECENT RESULTS from the LEP detectors are shown here. The mass and "width" of the Z^0 imply there are only three light neutrinos. Lower mass limits on the top quark and Higgs boson from L3 are still pending. (Lower limits on the top-quark mass are below those obtained from proton-antiproton colliders, but they also depend less on theoretical models for top-quark decay.) LEP data also rule out the selectron and wino (supersymmetric partners of the electron and *W*) below 44 GeV.

is carried by gravitons. One of the achievements of the Standard Model was to unify the electromagnetic and weak force, which is responsible for radioactive decay. In doing so, the theory predicted that the weak force would be carried by the intermediate vector bosons: Z^0, W^+ and W^-.

The discovery of these three particles in 1983 at CERN triumphantly validated the Standard Model. The theory also makes precise predictions about how the Z^0, W^+ and W^- may be generated and detected. For all its success, however, the Standard Model is not entirely satisfactory. For example, we do not know from first principles why elementary particles have the masses they do. Further experiments at the LEP will explore the range of validity of the theory with unprecedented precision.

The LEP's capacity to generate numerous Z^0's has already contributed to a landmark achievement: the determination that there are only three types of light neutrinos. According to the Standard Model, there is a neutrino associated with each generation of particles. The theory, however, does not specify the number of generations, and so the only way to determine this number is experimentally, by counting neutrino species.

Neutrinos are electrically neutral, almost massless particles that scarcely interact with matter; streams of neutrinos continuously pierce through the earth unimpeded. This makes them difficult to observe directly. The LEP Collider provided an indirect count of neutrino types without actually detecting the elusive particles themselves. It did so by measuring the "width" of the Z^0 "resonance" accurately.

What is this Z^0 resonance, and what does its width have to do with the number of neutrinos? The resonance refers to the fact that as the LEP collision energy is gradually raised near 91 GeV, one sees increasing numbers of Z^0 events. The number reaches a peak, or "resonates," at around 91.2 GeV and then declines as the energy is raised further. A histogram of the number of Z^0's plotted against the beam energy looks like a bell-shaped curve.

The width of this curve is the crucial evidence limiting the number of possible elementary particles. Why would that be? The width is actually a measure of the uncertainty in the energy, which, according to the uncertainty principle, is inversely related to time: the wider the curve, the shorter the lifetime of the particle. The lifetime in turn indicates the number of decay modes available to the Z^0: the more particles the Z^0 can decay into, the briefer its lifetime—and the wider the curve—will be. Hence, the number of neutrino types will affect the width.

Measurements of the Z^0 width from the four LEP detectors indicate only three types of light neutrinos. A fourth light neutrino is definitely excluded. Future theories will have to account for this finding. (The Z^0 resonance does not reveal whether there are any heavy neutrinos weighing more than half the Z^0 mass. Such neutrinos would be billions of times heavier than the known neutrinos.) This is a conclusion with consequences for cosmology and astrophysics. For example, one must now rule out the possibility that a fourth neutrino weighing up to about 10 GeV could provide the "missing mass" needed to halt the expansion of the universe.

LEP researchers are making concerted efforts to put other pieces of the Standard Model puzzle into place. The search is on for two particles predicted by theory, the top quark and the Higgs boson (which comes in neutral and perhaps also charged versions). Particle physics data from around the world, combined with the precise Z^0 mass, have led theorists to predict a top-quark mass of about 139 GeV, give or take 30 GeV. So far LEP experiments have ruled out a top quark weighing less than about 45 GeV and a neutral Higgs boson weighing less than 25 GeV. If the particles are not found at the energies that the LEP is capable of producing, it may simply mean that the particles weigh more—or it may mean the Standard Model is fatally flawed.

In designing the LEP to fulfill its scientific goals, researchers had to consider numerous factors. The most crucial factor determining the size of the LEP storage ring is the phenomenon of synchrotron radiation. Any charged particle bent into a circular orbit will radiate photons (particles of light) and lose a fraction of its energy. If this energy is not replenished, the particles will rapidly decelerate and spiral into the accelerator wall. At a beam energy of 55 GeV, the LEP radiates some 200 million volts per turn and requires 16 megawatts of power at 350 megahertz. This power is supplied by a series of radio-frequency (RF) cavities. At their future energy of 100 GeV, the LEP beams will radiate more than 2,800 megavolts per turn.

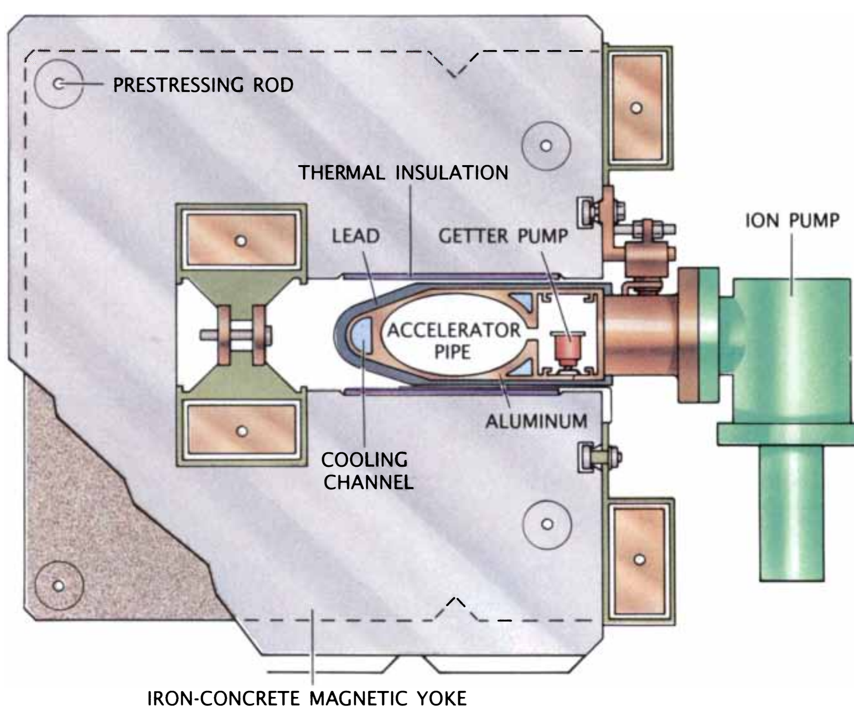
The cost of operating an electron-positron storage ring, then, is largely determined by the cost of replenishing the energy lost to synchrotron radiation. For a particle that is being bent into a circular path, this loss is proportional to the fourth power of the particle's energy divided by the radius of the circle. Clearly, for a given beam energy, one can reduce the loss by increasing the radius of the circle. Just as clearly, a larger ring is more expensive to build. In economic terms, there is a trade-off between the cost of a larger ring and the cost of building and operating the RF system.

Synchrotron radiation, although detrimental to the cost of maintaining beam energy, has beneficial effects as well. The radiation "damps" the oscillations of individual particles in the beam. As new particles are injected into the beam, their energies oscillate about the central energy of the beam. The larger the oscillations, the more quickly the oscillations will be reduced by radiation damping. Hence, radiation damping allows many particles to be accumulated by injecting new particles repeatedly into the beam.

The damping force, which on its own would reduce the particle oscillations to zero, is offset by a diffusive force, which causes the particles to increase their oscillation amplitudes. Hence, the beam reaches an equilibrium state in which the two opposing forces are balanced. The situation is like that of a balloon. The air pressure in the balloon is analogous to the diffusive force, and the elasticity of the rubber is analogous to the radiation damping. If the pressure increases, the balloon becomes larger, whereas if the balloon is made of a less elastic material, the size will be smaller at a given pressure.

The beauty of this equilibrium state is that it results in particle beams that are almost impervious to small perturbations, such as fast fluctuations in the guiding magnetic field, just as a balloon that is deformed returns quickly to its previous shape. At LEP energies, this property exists only for electrons and positrons, not for protons. For an accelerator physicist, this is the principal difference between electron and proton storage rings. Protons have a practically infinite "memory" for any perturbation, whereas electrons "forget" perturbations, typically in hundredths of a second.

The LEP designers had to produce a machine that could achieve the required energy and luminosity (beam intensity) yet balance these requirements against considerations of construction cost, operating expense



ACCELERATOR PIPE, bending magnet and vacuum-pump system are shown in a cross-sectional view. Particle beams travel down the center of the pipe. The aluminum wall is clad with lead to prevent synchrotron radiation from penetrating to the outside and generating corrosive gases. Surrounding the pipe is a large bending magnet; since it has to generate only a weak field, the magnet was made less costly by interleaving plates of iron with concrete. To achieve a high vacuum of 10^{-11} torr, "getter" pumps were employed to react with and bind stray molecules.

and future flexibility. Early in 1976 CERN designers proposed a machine 50 kilometers in circumference. This was the optimum size to attain the top beam energy of 100 GeV using a conventional accelerating system. This proposal was followed by one for a smaller, 22-kilometer-circumference ring with a beam energy of 70 GeV, which could be boosted to 100 GeV once a suitable superconducting acceleration system became available.

By the middle of 1979, however, the design study concluded that a somewhat larger circumference of 26.7 kilometers was preferable. The production of W^+/W^- pairs could then be attained by room-temperature accelerating systems and would not have to depend on the development of superconducting technology.

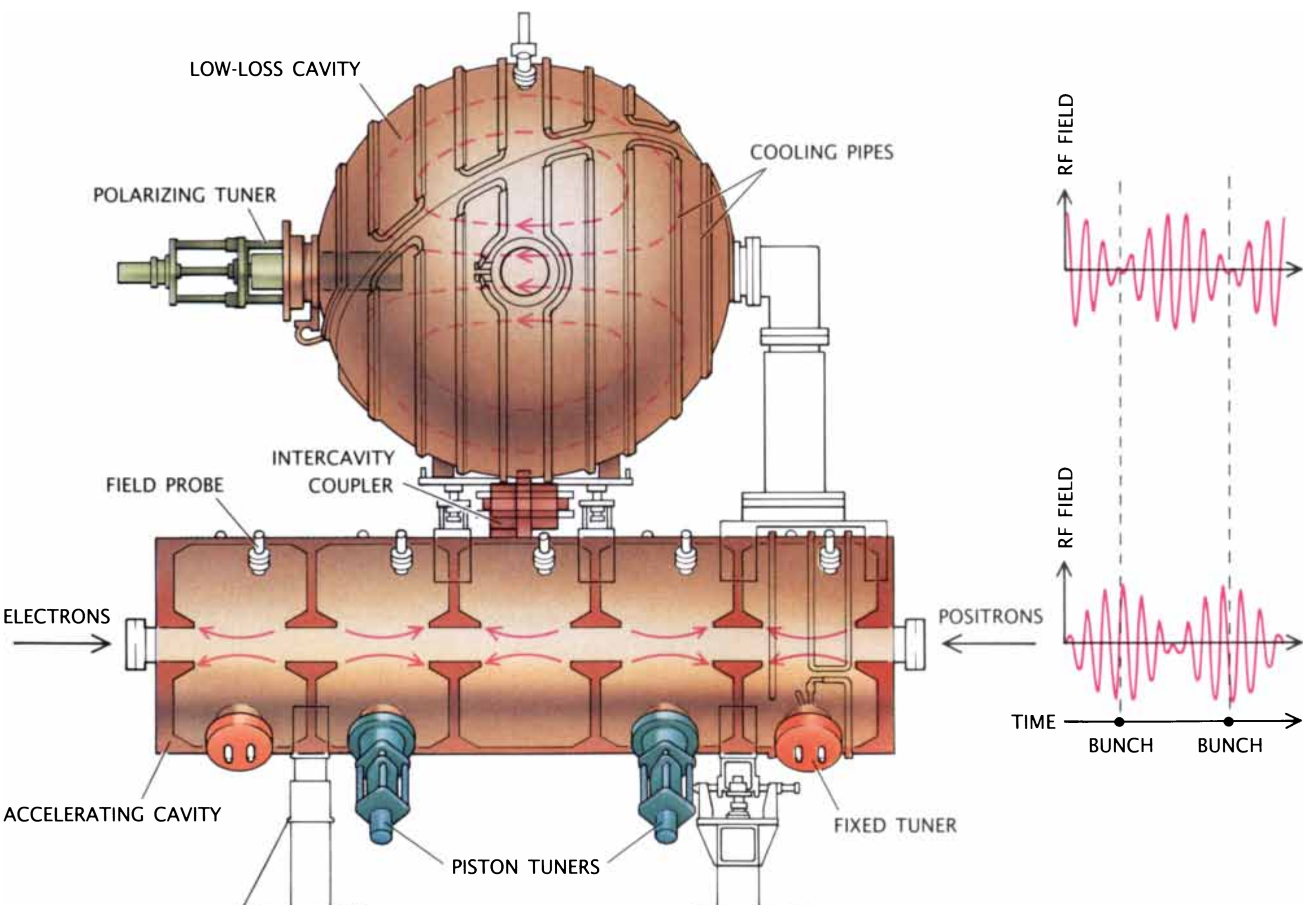
In December, 1981, the CERN council approved this design. The collider would initially operate at a maximum center-of-mass energy (total head-on collision energy) of 110 GeV (phase one) and would later be developed to attain energies of up to 200 GeV (phase two), with the possibility of still further increases in energy to about 240 GeV. The second phase will permit workers to study the strength of the interaction among the Z^0, W^+ and W^-, which is crucial to establishing the validity of the electroweak theory.

Construction work began soon after the LEP design was selected. From 1983 to 1988 the LEP was the largest civil engineering project in Europe. The engineering and infrastructure construction alone consumed more than half of the total construction budget. The main-ring tunnel formed the most impressive part of the work, even though it represented less than half of the 1.4 million cubic meters of rock and soil that had to be excavated. The remainder of the underground work included the four caverns to house the detectors, 18 pits and some 60 chambers and alcoves.

Three tunneling machines drilled through the rock at a rate of about 25 meters per day. They were guided on their desired trajectory to a precision of one centimeter. Collider components within the tunnel were aligned to within .1 millimeter. This remarkable precision was achieved using a laser-interferometry system set up in the hills surrounding the site.

To save on time and cost, the planners decided to use existing smaller accelerators for the particle-injector system, which boosts particles up to a high energy before injecting them into the LEP. Two linear accelerators drive the particles, first to 200 MeV and then to 600 MeV. A 600-MeV storage ring accumulates bunches of particles. The CERN proton synchrotron is enlisted to accelerate the bunches to 3.5 GeV. These bunches are injected into the Super Proton Synchrotron, where they are boosted to 20 GeV before finally being injected into the LEP.

Once particles are inside the LEP ring, they must be guided around the trajectory and focused at the collision sites. A system of electromagnets and electrostatic plates performs this task. Each of the eight arcs, which account for more than three fourths of the LEP circumference, consists of 31 standardized modules. Each module measures 79.11 meters in length and contains



RADIO-FREQUENCY (RF) CAVITY generates an oscillating electric field (*red arrows*) to accelerate particle bunches. To reduce heat loss in the copper cavity wall, the RF field switches to an adjoining low-loss spherical cavity whenever there are no bunches to accelerate. The spherical cavity dissipates less heat because the field gradient is strongest at the hollow center of the sphere rather than at its metal surface. The graphs show the field strengths in relation to bunches.

magnets in the following order: a vertically focusing quadrupole (a four-pole magnet that squeezes the beam in its vertical dimension), a vertical orbit corrector, a group of six bending dipoles, a horizontally focusing sextupole, a horizontally focusing quadrupole, a horizontal orbit corrector, a second group of six bending dipoles and finally a vertically focusing sextupole.

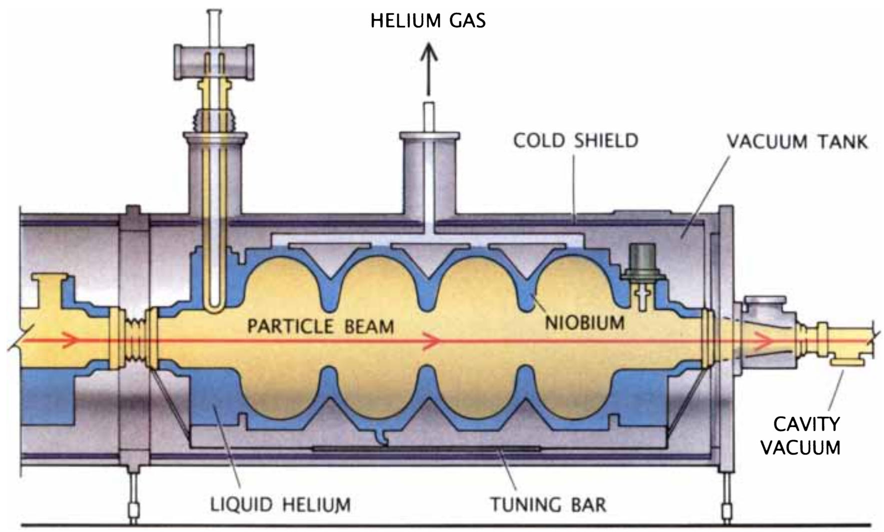
The bending dipole magnets guide the electrons and positrons around the arcs. The bending radius is large in order to reduce synchrotron radiation, and so the magnetic field of these dipoles is unusually low, about .1 tesla. The low field allows the dipole magnet core to have a novel design, in which the gaps between the magnetic steel plates are filled with mortar, like the custard between the layers of a Napoleon pastry. This proved to be 40 percent less expensive than conventional steel cores.

At the center of each collision site, the particle beams must be squeezed to tiny dimensions in order to increase the beam luminosity. This squeezing is accomplished by a set of superconducting quadrupole magnets. These focus the beam diameter to about 10 microns in the vertical plane and 250 microns in the horizontal.

A huge superconducting solenoid magnet encloses the collision site and generates a uniform magnetic field through the detectors. This field bends the paths of charged particles emerging from the collisions and so enables experimenters to deduce their mass and electric charge.

All magnets in the LEP are accurately adjusted by controlling the current flowing in their coils. This is accomplished by more than 750 direct-current power supplies, which range from less than one kilowatt to seven megawatts. The power supplies are accurate down to two parts in 100,000. They are precisely synchronized during the critical stage in which the beams are "ramped up" to peak energy.

At the LEP, four equally spaced bunches of electrons and four bunches of positrons circulate in opposite directions at the same time. The opposing beams cross paths at the heart of each experimental detector and at four intermediate points as well. During injection, accumulation and energy ramping, however, the counterrotating beams cannot be allowed to come near each other, because the electromagnetic fields associated with each bunch would deflect particles in the opposing bunch, causing the bunches to explode against the accelerator pipe wall. This "beam-beam effect" is overcome by subjecting the beams to an electrostatic field that acts like an invisible highway divider, forcing the counterrotating beams apart at all eight possible collision points. Just before data are ready to be recorded, the electrostatic plates at the experimental collision points are turned off and the beams allowed to collide. (The separators can also be used to steer the beams more accurately into collision.)



SUPERCONDUCTING RF CAVITY is designed to raise the LEP's energy to 100 GeV per beam. The liquid-helium-cooled niobium walls will reduce heat losses to zero. The flanges are rounded to eliminate sharp field gradients. Otherwise, electrons would accumulate and heat the niobium above the superconducting temperature.

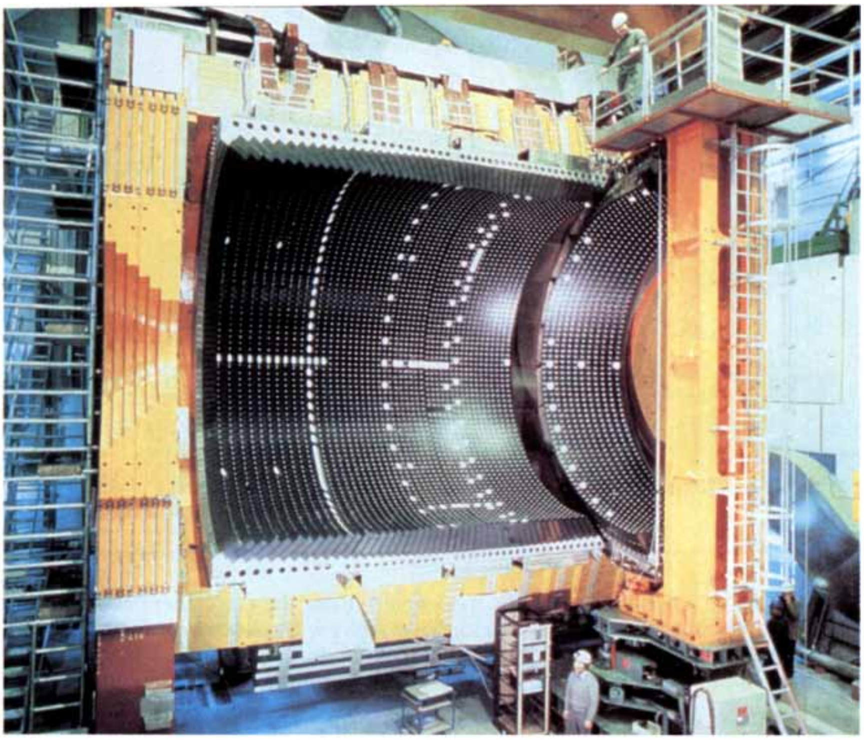
As mentioned earlier, the electrons and positrons circulating in the storage ring lose energy continuously in the form of synchrotron radiation. It is the job of the radio-frequency resonant cavity (RF system) to supply fresh energy to the beams.

The present RF acceleration system consists of 128 five-cell copper cavities powered by 16 one-megawatt klystrons. The klystrons generate radio-frequency power, which is directed into the cavities, where it produces a large electric field. The field oscillates at just the right frequency so that as a particle bunch enters a cavity, it gets accelerated forward by the field, like a surfer who catches a perfect wave.

A major problem for the RF system is that the accelerating electric fields heat up the copper cavity walls and thereby dissipate a large amount of power. To reduce the problem, LEP engineers mounted a spherical cavity atop each accelerating cavity in such a way that the RF power oscillates between the two cavities. The spherical cavity is designed to dissipate much less power than the accelerating cavity. The oscillations are timed so that the power is at a peak in the accelerating cavities at the instant that a particle bunch passes through. For half of the time, the power is in the spherical cavity. Hence, the bunches receive the maximum acceleration, yet the heating loss is greatly reduced because the RF power spends half of the time in the low-loss cavities.

The RF acceleration system also performs the very important role of concentrating particles into tight bunches. The RF system operates at a frequency of 352.21 megahertz, which means it generates 31,320 oscillations over one circuit around the LEP. Each oscillation contains a region within which each particle can perform stable oscillations with respect to the particle at the center of the bunch. This stable region is called the RF bucket. The RF bucket has properties that allow it to focus particles into discrete bunches of about one to two centimeters in length in which all the particles have the same energy, to within .1 percent.

Because there are 31,320 oscillations in one circuit around the LEP, there are 31,320 RF buckets. One might picture the storage ring as a pair of gigantic ferris wheels, each with 31,320 gondolas. The wheels rotate in opposite directions. If two friends want to ride in separate ferris wheels and pass each other at a precise point, they must sit in just the right gondolas. Similarly, in order for electron and positron bunches to collide at the centers of the detec-



CURVED WALL of lead glass and iron from the Opal detector looms some 10 meters high and 10 meters long. The lead glass absorbs electrons and gamma rays, and thick, yellow-painted iron slabs measure the energy deposited by hadrons. The entire section joins together with an identical structure to enclose a large solenoid magnet and gas-filled chambers that disclose the tracks of charged particles.

tors, particles must be injected and accumulated in only the correct buckets. This is achieved by very precise synchronization between the RF system and the injector system.

After accumulating enough particles and ramping them up to the required energy, the LEP beams are typically circulated for around six hours during which the beams are collided repeatedly in the detectors. During this time each of the 10^{12} particles in the beams will traverse the LEP circumference more than 240 million times, traveling altogether about 6,500 million kilometers—farther than the distance to Neptune. The entire accelerator pipe must be evacuated to very low pressures, because over the course of such a long trajectory, even a few gas molecules will collide with and erode the beam.

The pressure inside the LEP accelerator pipe is less than 10^{-11} torr. About 20 kilometers of the 27-kilometer length of the LEP vacuum chamber, however, is subject to synchrotron radiation, which provokes the desorption of gas from the vacuum chamber wall. Hence, in the presence of the beam the pressure rises to 10^{-9} torr.

Only about half of the radiated power is absorbed by the aluminum walls of the vacuum chamber. The remainder would normally pierce through and escape into the tunnel, where the high radiation would rapidly destroy organic materials such as rubber gaskets, cables and electronic insulation. The radiation could also react with air, forming ozone and nitric oxides, which, in humid conditions, produce highly corrosive nitric acid. To prevent the radiation from escaping, the aluminum chamber is covered with lead cladding.

Previous electron storage rings attained ultrahigh vacuums by linear sputter-ion pumps, which need to operate in the field of the bending magnets. But at the LEP the bending fields are too weak. Instead the LEP incorporates a novel system, adopted for the first time in an accelerator, which relies on "nonevaporable getter" strips. These strips are three centimeters wide and together extend over 20 kilometers. They consist of constantan (a copper-nickel alloy) coated with a zirconium-aluminum alloy. The material acts like molecular flypaper, forming stable chemical compounds with most active gases so that residual gas molecules simply "stick" to the ribbon.

Finally, to achieve pressures of less than 10^{-11} torr, the interior surfaces of the LEP vacuum chamber had to be immaculate. Every component was carefully cleaned with chemicals after manufacture and then stored in chemically inert conditions. After being installed in the tunnel, all the chambers were "baked out" at 150 degrees Celsius for 24 hours by pumping superheated water at a pressure of eight bars into the cooling channels of the aluminum-lead chambers. Components such as valves, gauges, stainless-steel chambers and electrostatic separators were baked out by electric heating.

To produce enough Z^0's to test the Standard Model precisely, the beams must be highly luminous. One way to increase the luminosity is to maximize the number of electrons or positrons in a bunch. This is done by injecting fresh particles into each circulating bunch. Luminosity is increased further by focusing the bunches down to a very small spot just at the point of collision by a series of magnetic lenses. In its first phase the LEP is designed to provide a peak luminosity of around two times 10^{31} particles per square centimeter-second.

Achieving this high luminosity has proved a considerable challenge. A variety of phenomena arising from the complex physics of accelerated beams conspire to diminish the luminosity. The first of these is the short-term "wake field." Wake fields are induced in the accelerator components by the electromagnetic fields of particles at the front of a bunch. The wake fields excite the particles at the tail end of the bunch, thereby increasing the amplitude of their transverse oscillations. The inherent longitudinal motion within an RF bucket sends these particles from the rear of the bunch to the front. There, because of their greater transverse oscillations, the particles generate yet stronger wake fields. This unstable growth in oscillations eventually drives the particles into the accelerator pipe wall.

The wake-field effect can be reduced by ensuring that the accelerator pipe is perfectly symmetric cross-sectionally and uniform longitudinally. In practice, there are many nonuniformities, which occur at regions such as the RF cavities, the electrostatic separators and the bellows that link the vacuum chambers. Substantial effort went into designing these components so as to minimize the effect of the wake field. The vacuum bellows in particular are of a completely new design. During the first months of operation, we took measurements of the beam and found that the wake-field effect at the LEP was about 30 percent less than had been anticipated.

The low wake-field effect may make it possible to increase the beam current in the future. Another problem must first be addressed, however. The LEP has so far attained a maximum current of 2.2 milliamperes, somewhat below the design goal of three milliamperes. We tentatively attribute this reduced current to the magnetization of the nickel that was employed to bind the lead cladding to the aluminum vacuum chamber wall. A mere 10 microns thick, the nickel creates feeble magnetic fields, yet these are strong enough to disperse the beam significantly. The effect can be suppressed either by introducing compensating magnetic fields or by demagnetizing the nickel.

Another important factor affecting the luminosity is the number of particle bunches circulating within the storage ring. Clearly, the more bunches there are, the more frequently they will collide. Four factors influence the number of bunches. The first is beam power. Increasing the number of bunches increases the beam current and hence the beam power. During phase one the beam power at the LEP is only 1.2 megawatts. During phase two the beam power will be 12 megawatts for the same beam current. But if the number of bunches is increased by, say, tenfold, then the beam power for phase one (55 GeV) would be 12 megawatts, and the same number of bunches at 100 GeV would consume 120 megawatts.

The second factor is the extraneous collision points that must be eliminated. For a given number of circulating bunches of each particle type, there are twice as many collision points. Because of the beam-beam effect mentioned previously, the beams must be kept apart by a complicated system of electrostatic plates at all collision points except in the detectors. If we increase the number of circulating bunches, then we must also install more separators.

A third factor is the long-term wake field. The short-term wake field disrupts the tail end of a particle bunch. The long-term wake field disrupts subsequent particle bunches, just as turbulence from a jet aircraft jostles any aircraft in its wake. The strength of this effect is directly related to the spacing between bunches. Already a feedback system is needed to correct for the long-term wake field generated by the four bunches each of electrons and positrons that now circulate in the LEP. The cost and complexity of this feedback system will increase as more bunches are introduced into the LEP.

Finally, we must consider the ability of the detectors to cope with an increased rate of collisions. A finite time is required for the electronics of these detectors to record and measure an event. Greater numbers of bunches will increase the rate of events, and so the detectors will require more expensive and more complex electronics.

For all these reasons, the first phase of the LEP was designed to handle only four bunches per beam. As more power becomes available in the second phase, it may be possible to operate the machine at phase-one energies and introduce more bunches. This option is now being studied. It will require the installation of a new system of separators and beam-feedback systems. In addition, the detector electronics will have to be modified substantially. It now appears that a scheme with 36 bunches per beam will be feasible.

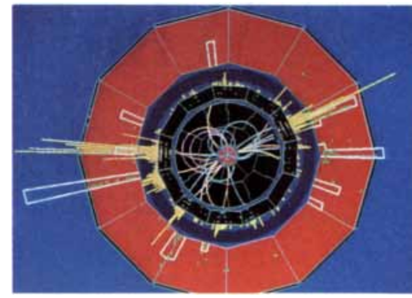
The first phase of the LEP has been successfully completed as far as the machine itself is concerned. We are now striving to increase the luminosity up to, and perhaps beyond, the design value. The second phase will take the beams up to energies of 100 GeV each. To reach such energies, however, the LEP will need substantially more power to replenish synchrotron radiation losses. What is more, with conventional RF systems, the power lost through heating would become enormous. Hence, it is essential to reduce heat losses in the accelerating cavity walls to almost zero. This reduction can be achieved only with superconducting RF cavities.

We have been improving the design and manufacture of superconducting cavities and hope to raise the voltage in these cavities from the present five megavolts per meter to seven megavolts per meter. In phase two of the LEP some 200 such cavities will be installed. Many other systems will also have to be upgraded. In particular, some of the superconducting quadrupoles must be replaced by more powerful ones, and new klystrons are required to provide the additional power. If all goes according to plan, the LEP physicists will be able to study the physics of W^+ and W^- in 1994.

We are also examining a scheme to create polarized beams, which consist of particles with longitudinally aligned spins. They will enable us to study possible deviations from the Standard Model with greater sensitivity.

Finally, a group is studying a proposal to build a large proton collider within the LEP tunnel. The Large Hadron Collider (LHC) may be able to collide protons at a center-of-mass energy of 16,000 GeV. This capability would place it between the energies of the Tevatron at the Fermilab National Accelerator Laboratory in Batavia, Ill., and the Superconducting Supercollider. Another exciting option is the possibility of colliding the 8,000-GeV proton beams of the LHC with the 100-GeV electron beams of the LEP.

The decision to build the LEP, or any new accelerator for that matter, is difficult to justify a priori. The correctness of the choice can be judged only in retrospect after about a decade of operation. The first few months at the LEP have already vindicated many of the scientific arguments for building it. Perhaps the most exciting possibility is that the LEP will yield unforeseen discoveries and so open new vistas in the understanding of nature's laws.



COMPUTER DISPLAY depicts a spray of particles forged by an electron-positron annihilation in the Aleph detector.

FURTHER READING

LEP, CERN'S LARGE ELECTRON POSITRON COLLIDER: THE EXPERIMENTAL PROGRAMME. R. Turlay in *Europhysics News*, Vol. 20, No. 6, pages 75–79; June, 1989.

THE LEP MACHINE DESIGN. G. Plass and E. Picasso in *Europhysics News*, Vol. 20, No. 6, pages 80–91; June, 1989.

ALEPH COLLABORATION: A PRECISE DETERMINATION OF THE NUMBER OF FAMILIES WITH LIGHT NEUTRINOS AND OF THE Z BOSON PARTIAL WIDTH. CERN-EP/89-169, December 19, 1989. Submitted to *Physics Letters B*.

L3 COLLABORATION: MEASUREMENTS OF Z^0 DECAYS TO HADRONS AND A PRECISE DETERMINATION OF THE NUMBER OF NEUTRINO SPECIES. L3 Preprint No. 004, December 24, 1989.

OPAL COLLABORATION: A COMBINED ANALYSIS OF THE HADRONIC AND LEPTONIC DECAYS OF THE Z^0. CERN-EP/90-27, February 23, 1990. Submitted to *Physics Letters B*.

DELPHI COLLABORATION: A PRECISE MEASUREMENT OF THE Z RESONANCE PARAMETERS THROUGH ITS HADRONIC DECAYS. CERN-EP/90-32, March 9, 1990. Submitted to *Physics Letters B*.

What Causes Diabetes?

For insulin-dependent diabetes, the answer is an autoimmune ambush of the body's insulin-producing cells. Why the attack begins and persists is now becoming clear

by Mark A. Atkinson and Noel K. Maclaren

People stricken with what is today called insulin-dependent, or type I, diabetes once faced certain death within about a year of diagnosis. They died because the pancreas lost its ability to make insulin, which is required for normal metabolism. Isolation of the hormone from animals in 1921 made treatment possible and has since meant survival for millions of diabetics.

Yet neither animal insulin nor the more modern, human, form offers a cure. Injections must be taken once or more a day for life. Moreover, many diabetics eventually suffer from devastating complications. As the disease persists, blood vessels can be damaged, leading to heart disease, stroke, blindness or kidney failure. Nerve damage is also common.

Improved techniques for delivering insulin might slow the development of complications, but the ideal solution is to prevent diabetes itself. To do that, however, investigators must uncover the root causes. Recently, research into that problem has advanced with remarkable speed. It is now clear that insulin-dependent diabetes stems from an autoimmune, or self-directed, attack on the pancreas. Dozens of laboratories in the U.S. and abroad—including our own at the University of Florida—are currently attempting to clarify the details of that process. We are trying to determine which components of the immune system are the major agents of attack, what triggers the autoimmune reaction and what allows it to persist.

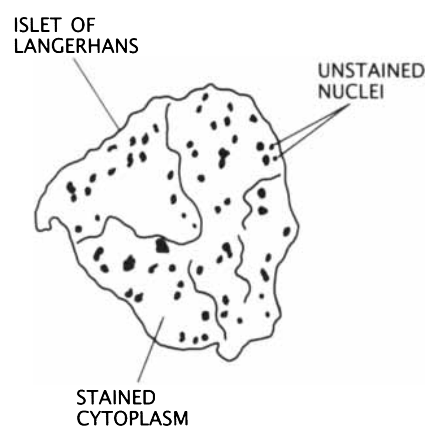
On the basis of such work, physicians should soon be able to identify the one person in 300 who will acquire insulin-dependent diabetes, so that treatment can be initiated at the very first sign of disease. By the turn of the next century, it should also be possible to offer such individuals safe preventive therapy.

The autoimmune process that causes insulin-dependent diabetes is highly selective and frequently begins before adulthood (which is why the disease was formerly called juvenile-onset diabetes). The attack does not affect the majority of pancreatic cells, which secrete digestive enzymes. Instead it restricts itself to the hormone-producing cells. These are clustered in spherical groupings—the islets of Langerhans—scattered throughout the pancreas. Even in the islets, three out of the four cell types are spared; only the insulin-producing beta cells, which make up the large core of an islet, are targeted.

Insulin helps most cells of the body take up biological fuels, including the sugar glucose. As the beta cells are killed and the pancreas stops producing this crucial hormone, glucose accumulates in the blood, giving rise to the abnormally high glucose levels that are a hallmark of diabetes. Then the body becomes dehydrated as the kidneys overwork to filter the excess glucose into the urine. Meanwhile body cells starve in a sea of plenty, and so they uncontrollably break down their stores of fat and protein to provide more fuel. If the breakdown of fat continues unchecked, acidic by-products called ketones build up. These, combined with dehydration, can induce coma and, finally, death.

Insulin injections can halt this lethal sequence and prevent it from recurring, but they cannot mimic the normal pattern of insulin release by the pancreas. Nor can they normalize metabolic functioning well enough to prevent the long-term complications of diabetes. These are generally believed to be caused or exacerbated by chronically elevated blood glucose levels.

At one time, the beta cells were thought to be killed suddenly, perhaps by a virus or some ingested toxic chemical. That view was based on the observation that symp-



ISLETS OF LANGERHANS house the hormone-producing cells of the pancreas and include the beta cells, which make insulin. This islet was made visible in a slice of a normal pancreas by fluorescently labeled antibodies that bind to a component of the islet-cell cytoplasm (*green*). The antibodies came from the blood of an insulin-dependent diabetic, and their recognition of a healthy islet indicates that an autoimmune attack of normal tissue contributes to diabetes. The disease emerges after most of the pancreatic beta cells are destroyed.

MARK A. ATKINSON and NOEL K. MACLAREN are, respectively, assistant professor and professor in the department of pathology and laboratory medicine at the University of Florida College of Medicine in Gainesville. Atkinson received a B.S. from the University of Michigan and a Ph.D. in pathology from the University of Florida. He is currently focusing his attention on the pancreatic proteins involved in the development of insulin-dependent diabetes. Maclaren, who chairs his department, received much of his training in New Zealand, earning an M.D. at the University of Otago Medical School in 1963 and then specializing in internal medicine at Wellington Hospital. He also received training in pediatric endocrinology at Johns Hopkins University before moving to Florida in 1978 to pursue research into the genetics and pathogenesis of autoimmune endocrine diseases.

toms often arise abruptly. First the cardinal signs of the disease appear: ravenous hunger (resulting from energy-starved tissues), frequent urination (a consequence of extra work by the kidneys) and unquenchable thirst (caused by the body's need to replace water lost in the urine).

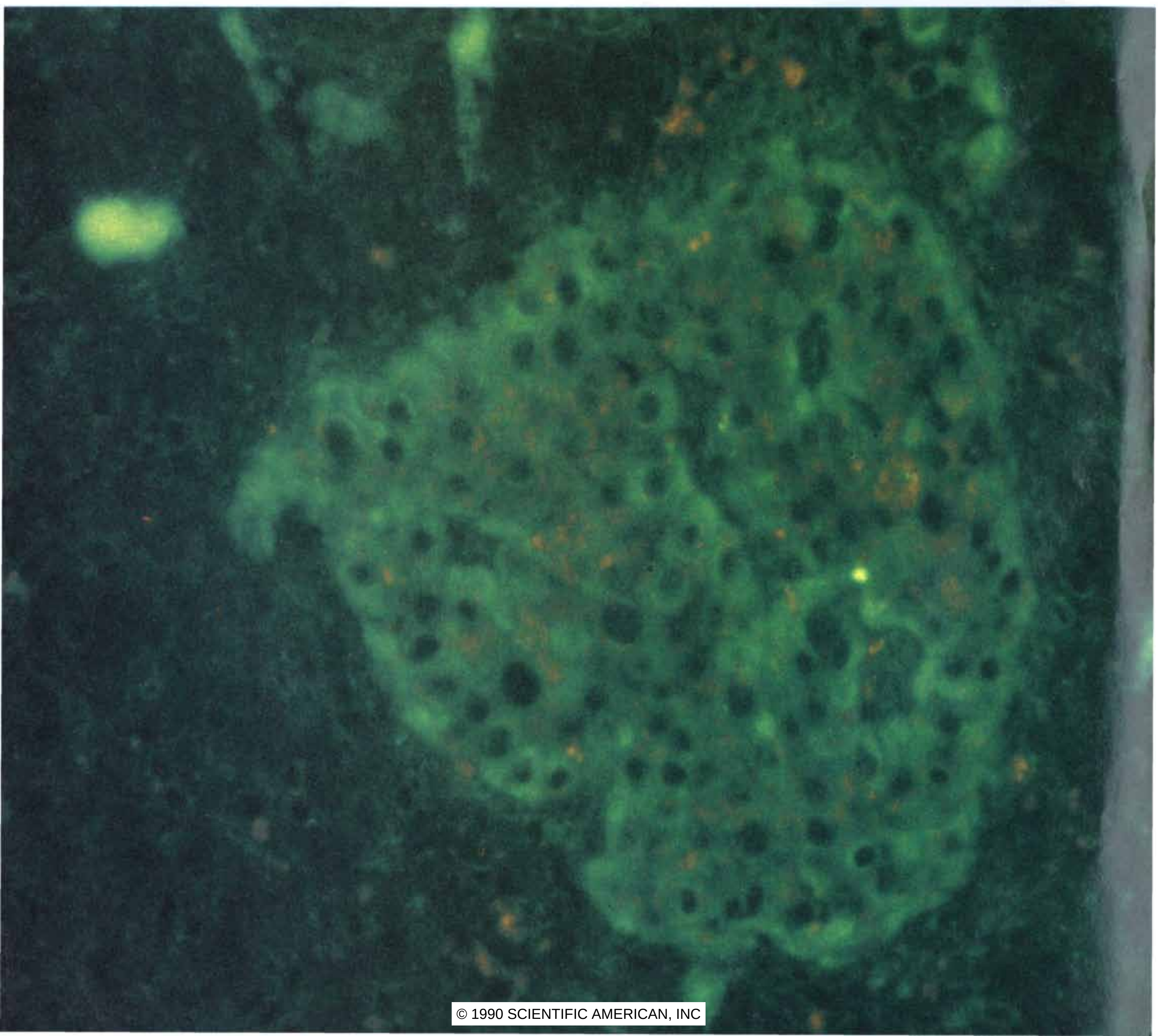
Patients may also lose weight with astonishing rapidity; it is not unusual to shed as many as 15 pounds within two weeks. Some people progress to coma quickly as well and, if they are deprived of prompt care, may die soon afterward, even though they may have seemed completely well just days or weeks before.

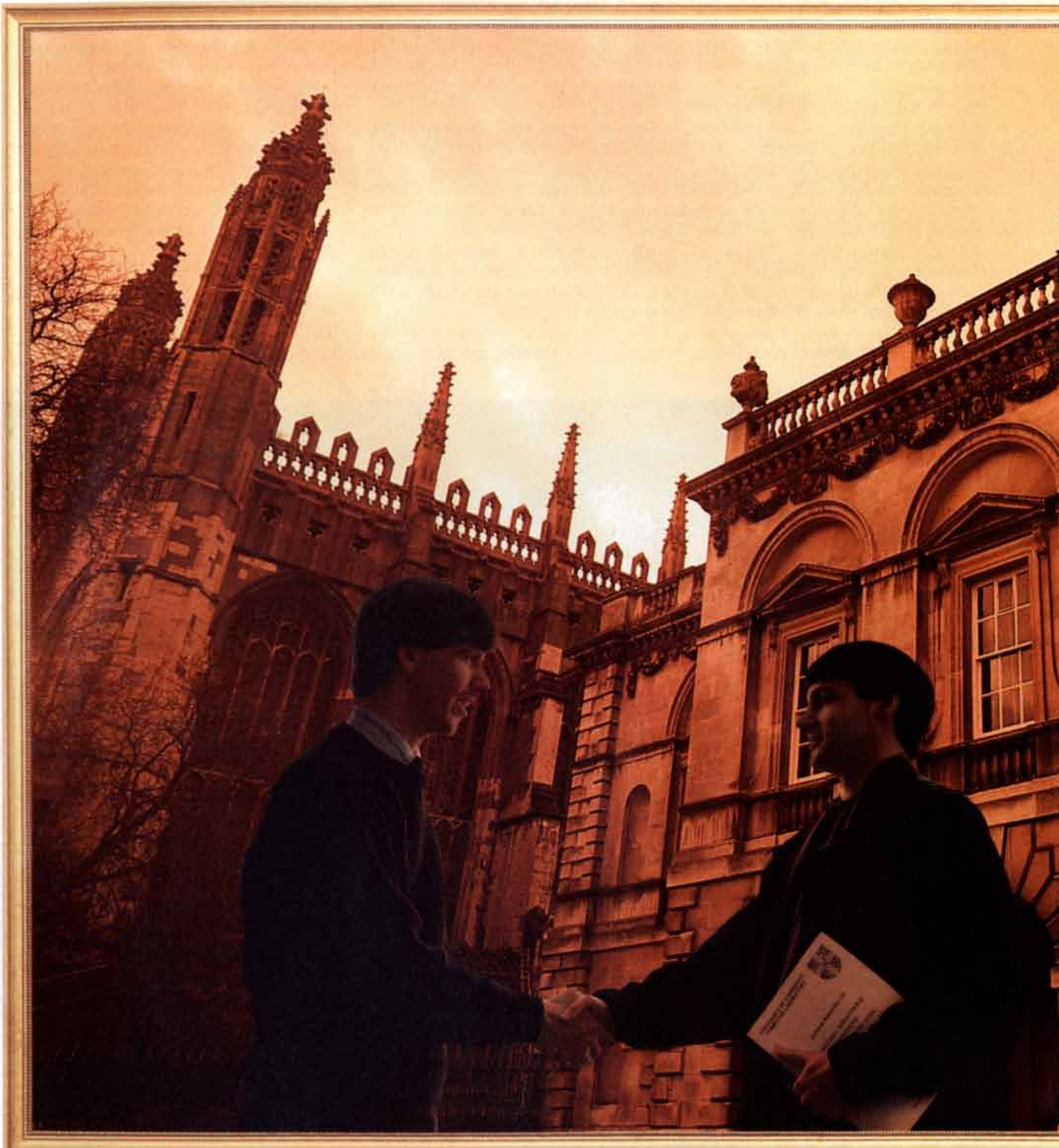
In contrast, the symptoms of non-insulin-dependent (type II) diabetes—a disease caused by entirely different mechanisms (still under study)—are more subtle. Non-insulin-dependent diabetics, who are usually older than 40 and overweight at diagnosis, produce various quantities of insulin, but they typically use it inefficiently. Some of them are treated with insulin to control blood glucose levels, but they do not require the drug for day-to-day survival.

Despite appearances, insulin-dependent diabetes actually does not develop suddenly at all. It typically "brews" silently for several years, as the immune system slowly eliminates the beta cells. The classic symptoms appear only when at least 80 percent of those cells are gone; the remainder are eliminated over the next two or three years.

The first major indication of immune-system involvement came from studies of pancreatic tissue of patients who died close to diagnosis. The examinations revealed that the islets were inflamed: they were filled with white blood cells (lymphocytes and monocytes), which normally attack microbe-infected cells and otherwise help the body to fight infection. The islet inflammation, called insulitis, could in theory be explained as an appropriate defense against a foreign organism. Such a response might occur if an organism had invaded the beta cells and then displayed antigens—substances that mark an invader as "nonself"—on the cell surface.

Further evidence, however, eventually pointed to autoimmunity. For instance, many insulin-dependent diabetics have other autoimmune disorders. Investigators also detected antibod-





digital

"Cambridge University teamed up with Digital for world-class computing research."

"Like most universities, Cambridge has relationships with all sorts of companies. With Digital, we at the Computer Lab have something special going. It's been a long-running, extremely close association that can best be described as a corporate friendship.

"There are intellectual and practical benefits to both us and Digital. Some of our graduate students spend summers in Palo Alto, California, at Digital's Systems Research Center. Several of the world's best computer scientists work in Digital's laboratories. Digital sponsors projects in our lab and sends top engineers to do Ph.D. work with us. The result is some of the best and brightest minds get to work together on research that will have significant impact on the world of computing.

"We're both interested in the same things and we're both good at them. I can't think of a better basis for real progress."

The rewards of working together.

Digital believes that working together with the world's leading universities is an important way to create innovative computing solutions—solutions that make information more of a shared resource.

Today, Digital gives you an elegantly simple way for your people to work together more productively, more creatively, more efficiently, more competitively.

For a white paper on Digital's university-based research, please call 1-800-343-4040 ext. 161.

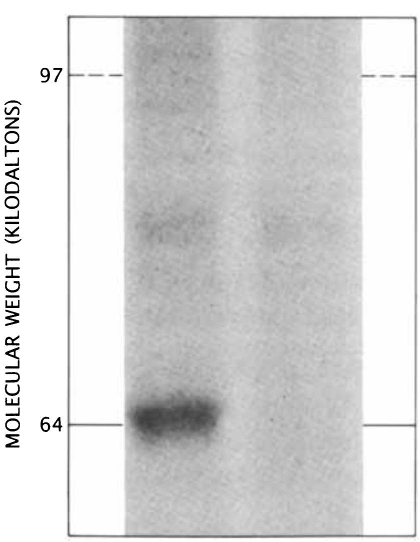
A way to work together like never before.

Digital has it now.

Roger Needham
Head of the
Computer Laboratory
Cambridge University

ies—or what may be called "autoantibodies"—against natural components of islet cells ("autoantigens") in newly diagnosed individuals and in those who were destined to become diabetic years later. Because antibodies mediate a part of the immune response (the humoral response), discovery of these anti-islet autoantibodies meant the immune system was in fact aroused against the self.

The fact that islet-cell autoantibodies are produced does not necessarily mean they are the main agents of beta-cell destruction, but the possibility has to be considered. Antibodies, which are produced by activated *B* (or bone marrow–derived) lymphocytes, bind to a single antigen. When they recognize that antigen on a target cell, they may interfere with the function of the cell directly or enlist the destructive powers of other components of the immune system, such as macrophages (scavenger cells), natural killer (NK) cells and lethal proteins known collectively as complement. These proteins attach to the free end of an antibody and then punch holes in the cell.



PROTEIN weighing 64 kilodaltons (K) and present only on beta cells elicits the production of a specific antibody against itself in diabetics and prediabetics but not in healthy people. This "autoantibody" appears before others, which suggests that the 64-K protein helps to trigger the autoimmunity that leads to diabetes. The autoantibody is detected indirectly, by mixing a subject's blood with radioactively labeled islet proteins. Autoantibodies in the blood precipitate any proteins they recognize. Then their identity is inferred by separating the proteins (according to weight) by gel electrophoresis. The gel samples above show that the blood of a diabetic (*left*) reacted with the 64-K protein (*dark band*) but that the blood of a nondiabetic did not (*right*).

Three autoantibodies, all of which are found more often in the blood of diabetics and their relatives than in the general population, have been studied most extensively. One of them, the cytoplasmic islet-cell autoantibody (ICA), was discovered in the 1970's by Deborah Doniach, Gian Franco Bottazzo and their colleagues at the Middlesex Hospital in London. It reacts to the cytoplasm of all islet cells and may be directed against a complex lipid known as a ganglioside.

About a decade later a second autoantibody was discovered by Steinunn Baekkeskov and Åke Lernmark of the Hagedorn Research Laboratory in Denmark. It is called the 64-K autoantibody because it reacts to a native islet protein about which little is known with certainty except that it weighs 64 kilodaltons (K). This protein is found only in the plasma membrane of beta cells.

The third autoantibody, discovered by Jerry P. Palmer and his colleagues at the University of Washington, reacts to insulin. Virtually all insulin-treated patients produce standard antibodies to injected insulin; however, this autoantibody can be identified in many insulin-dependent diabetics before they ever receive their first insulin injection.

If any of these autoantibodies leads the attack on the beta cells, it is probably the 64-K variety. It is beta cell–specific (unlike the cytoplasmic islet-cell autoantibody), and we have shown that it appears before the other two autoantibodies and well before diagnosis. The 64-K type is also the only one of the three found in both of the two rodent models of insulin-dependent diabetes: the BB (BioBreeding) rat and the NOD (nonobese diabetic) mouse; the latter makes the insulin autoantibody as well. In these species the 64-K autoantibody is produced early in life, coincident with the onset of insulitis.

There is, however, little evidence that even this autoantibody is the primary agent of beta-cell killing. If autoantibodies were major players, their injection into healthy animals might be expected to cause diabetes. But when they are administered to healthy rodents, they do not produce disease. Similarly, no infant born to an insulin-dependent diabetic mother has yet been reported to have acquired the disease in utero, even though maternal antibodies readily pass through the placenta to the fetus.

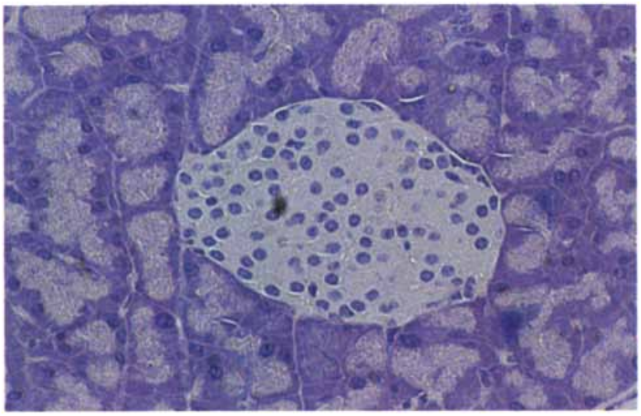
The issue is not yet closed. Nevertheless, we and many other scientists suspect that cell-mediated, rather than antibody-mediated, immunity is primarily responsible for the development of insulin-dependent diabetes. Under normal conditions, the cellular response, effected primarily by lymphocytes, is crucial to the eradication of viruses as well as other pathogens that infect cells.

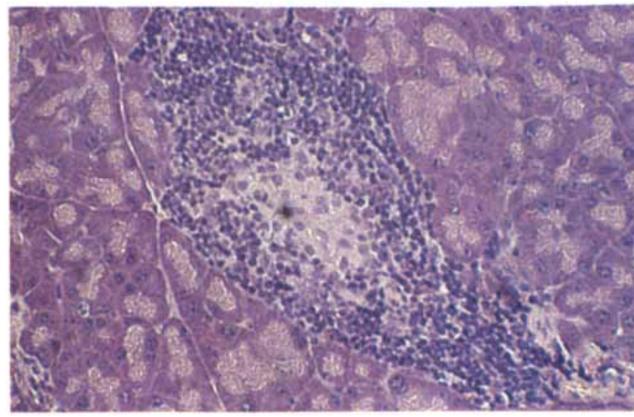
The helper *T* lymphocyte, which is produced by the bone marrow but matures in the thymus, is "command central." Such cells bear receptors that recognize a specific antigen. The *T* cells become activated only when the receptor binds an antigen that is itself associated with what is called a class II MHC molecule on macrophages and other antigen-presenting cells. The MHC, or major histocompatibility complex, is the genetic region on chromosome 6 that encodes tissue-typing proteins, or markers of self that distinguish the tissues of one individual from those of another.

Once the helper *T* cells are activated, they secrete cytokines (peptide mediators) that enhance the immune response. One cytokine, interleukin-2, promotes the helpers' own proliferation and that of cytotoxic, or killer, *T* cells. The cytotoxic cells recognize antigen only if it is bound by what is known as a class I MHC molecule. Other interleukins promote the secretion of antibodies by *B* cells and thereby greatly strengthen the humoral immune response.

Strong evidence implicates lymphocytes as the major agents of beta-cell killing. In both BB rats and NOD mice, the transfer of cells from the spleen (which is rich in NK cells and *T* cells) of newly diabetic animals into healthy animals can induce diabetes in the recipients. Similarly, preliminary data gathered by Kevin J. Lafferty of the Barbara Davis Center for Childhood Diabetes in Denver suggest that lymphocytes taken from the peripheral blood of newly diagnosed insulin-dependent diabetic humans can induce insulitis in the so-called nude strain of mice, which lack normal *T* cells.

Whether it is NK cells or *T* cells that effect most of the beta-cell destruction in humans is not clear, but we favor the *T* cells. Aldo A. Rossini and Arthur A. Like of the University of Massachusetts Medical School and Alexander Rabinovitch of the University of Alberta Medical School in Edmonton have found evidence that NK cells are most active during the development of diabetes in BB rats. Yet other work has shown that *T* cells are more active in NOD mice, whose diabetes more closely mimics that of humans. Moreover, *T*





ISLET of a normal mouse (*left*) contrasts with that of a NOD (nonobese diabetic) mouse, a model of insulin-dependent diabetes (*right*); lymphocytes (*dark spheres*) have infiltrated the latter islet. Such "insulitis" is evidence of autoimmunity.

cells are the most prevalent cells in human insulitis.

What exactly arouses the immune system against the body's own beta cells? At one time, it seemed possible that the trigger could be an abnormal substance on the beta cell, such as a protein encoded by a mutated gene, but recent findings argue against that idea. If an abnormal antigen participated in the diabetic process, newly diagnosed or prediabetic individuals would presumably generate an autoantibody against it. Yet every islet-cell autoantibody identified thus far reacts to some normal constituent in the islet.

What is more, our colleagues Ammon B. Peck and Drake M. LaFace have shown that when the immune system of still healthy (not yet diabetic) NOD mice is destroyed and reconstituted with bone-marrow cells from normal mice, the NOD mice develop no insulitis and no diabetes. If the beta cells of the recipients displayed an abnormal antigen, a normal immune system would certainly have noticed and become aroused for attack.

Conversely, when the immune system of normal mice was replaced with bone-marrow cells from NOD mice, the islets of the recipients became inflamed, and many of the animals developed overt diabetes even though their beta cells were normal at the start of the experiment. These studies also cast doubt on the once popular notion that a viral infection of beta cells (which would induce the display of foreign antigens) is the usual trigger of insulin-dependent diabetes.

The autoimmunity of diabetes may well be evoked by a process known as molecular mimicry [see "The Self, the World and Autoimmunity," by Irun R. Cohen; SCIENTIFIC AMERICAN, April, 1988]. In such a process a foreign antigen—such as a component of a virus or another microorganism—would provoke a normal immune response somewhere in the body. If in chemistry or conformation this antigen were essentially a twin to a component of the beta cell, the antigen would also stimulate an immune attack against the beta cell. Molecular mimicry is clearly involved in other autoimmune disorders, such as rheumatic heart disease, which can occur following a streptococcal infection of the throat.

We consider it likely that the 64-K protein is a mimic of some foreign antigen and as such provokes the autoimmune attack on the beta cells. We think so because the consistently early appearance of autoantibodies against this protein suggests that the 64-K protein is important in the initial stages of beta-cell attack.

Irun R. Cohen and his colleagues at the Weizmann Institute of Science in Israel recently found that certain bacterial heat-shock proteins, which are produced during stress, may be among the offending foreign antigens. When the team immunized normal mice against the proteins and then introduced *T* cells from those animals into healthy mice, the transferred cells induced both insulitis and high blood glucose levels in the recipient animals.

In broad outline, then, one can say that once some trigger (molecular mimickry?) induces a (cell-mediated?) immune response against the beta cells, the autoimmune ambush proceeds as it would if the insulin producers were infected by a virus. Helper *T* cells directed against beta cells multiply and call forth various other soldiers in the body's defensive army.

Yet this simple portrait must be amended in two ways. First, beta cells seem to hasten their own destruction by being unusually susceptible to immune-mediated damage. Second, the immune attack itself is more vigorous and prolonged than a normal response to a virus would be.

Several characteristics of beta cells account for their great vulnerability. When the cells are damaged, they display excessive numbers of class I MHC molecules. Such display promotes attack by cytotoxic *T* cells. Beta cells may also be particularly susceptible to direct damage by cytokines. For instance, Jørn Nerup of the Steno Memorial Hospital in Denmark and Rabinovitch have shown that the cytokine interleukin-1, which is secreted by activated macrophages, is directly toxic to beta cells but less toxic to other cells. Finally, the beta cells seem unusually prone to damage by free radicals: highly reactive oxygen molecules released by damaged cells and macrophages.

Many of us suspect that the exaggerated quality of the immune campaign against the beta cells similarly has more than one cause. It seems to stem in part from an overreaction of the helper *T* cells to beta-cell autoantigens. Meanwhile the natural suppressor system, which should hold autoimmune activity in check, fails to act or does so only weakly. Little is known about the normal processes that suppress autoimmune activity, and so no one fully understands why those processes do not halt the autoimmune destruction of beta cells. Nevertheless, studies of NOD mice suggest that the failure may have some genetic basis.

In these animals, genetic abnormalities in the class II MHC region contribute to the propensity of the animals to develop diabetes. Mouse genes encode two types of class II MHC proteins. One, called IE, seems to be important in normal suppressor activity.

The other, IA, is important in the recognition of antigens by *T* cells.

It turns out that NOD mice do not express their IE gene (they make no IE protein). Yet Hirofumi Nishimoto of Osaka University in Japan showed that when NOD mice are induced by gene transfer to express the gene, they develop little or no insulitis. Hence, the absence of the IE gene must contribute to diabetes. The lack of the gene is not the sole cause of diabetes, however, or the absence should lead to diabetes in other strains of mice. Yet many otherwise normal strains do not express IE proteins and develop neither insulitis nor diabetes. Whether autoimmunity in human diabetes is related to some flaw in the expression or structure of a gene analogous to the IE gene is unclear.

Workers have been more successful at unearthing clues to why the autoimmune response to beta cells is exaggerated. In mice, inheritance of the genetic codes for a specific variety of IA protein—the one in all NOD mice—increases susceptibility to diabetes.

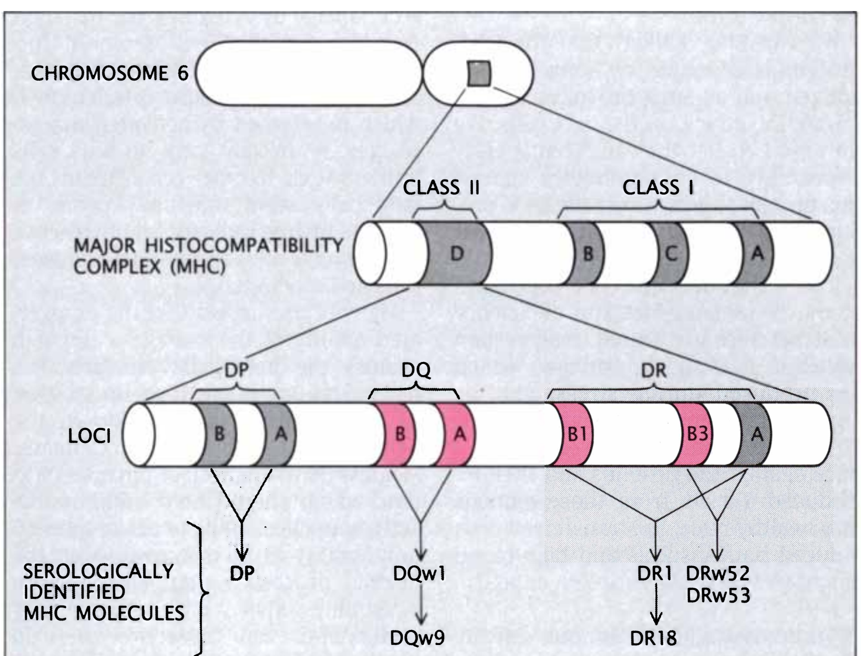
Edward H. Leiter of the Jackson Laboratory in Bar Harbor has shown that at least two other genes (whose functions are not yet clear) contribute to diabetes in NOD mice. Once the genes are better characterized, it may be possible to use them as probes for identifying similar genes in human beings.

The class II MHC region has been associated with susceptibility to insulin-dependent diabetes in humans as well. That region, known as the D region, includes three loci—DP, DQ and DR—each of which gives rise to a molecule containing two amino acid chains: A and B. So far only the DQ and DR loci have been connected to diabetes. The genes in both loci have a number of variants, or alleles; hence, many different DQ and DR proteins (which are numbered) occur in the population. Because the DQ and DR genes inherited from both parents are expressed, each individual can produce two DQ and two DR types.

It was discovered long ago that the genes for the proteins DR1, 3 and 4 are particularly common in insulin-dependent diabetics, whereas those for DR2 and 5 are uncommon. The finding suggested that the DR1, 3 and 4 genes increase susceptibility to diabetes or are inherited systematically with some true susceptibility gene and that the DR2 and 5 genes confer some protection or are linked to a protective gene.

Now researchers are homing in on the specific susceptibility genes. For instance, the genetic instructions for the DR4 protein are usually passed from one generation to the next together with the codes for a particular DQ protein: DQw3.2. (The "w" represents a provisional designation; the decimal distinguishes this protein from other DQw3 proteins that seem alike on the basis of serological tests but actually have small differences in their amino acid sequences.) Individuals who inherit the codes for the DQw3.2 protein are quite likely to acquire diabetes, whereas the relatively few DR4-positive individuals who produce the DQw3.1 variant are less likely to acquire the disease.

About two years ago John A. Todd and Hugh O. McDevitt of Stanford University identified one region of DQ molecules that increases susceptibility to diabetes. In people who inherit the genes for DR proteins 1 through 6, they found that susceptibility is usually determined by the 57th amino acid on the DQB chain. When this position is occupied by aspartic acid, which is negatively charged, the likelihood of diabetes developing is low. But a noncharged amino acid, such as valine or serine, at the same position (as occurs in the DQw3.2 protein) raises the risk.

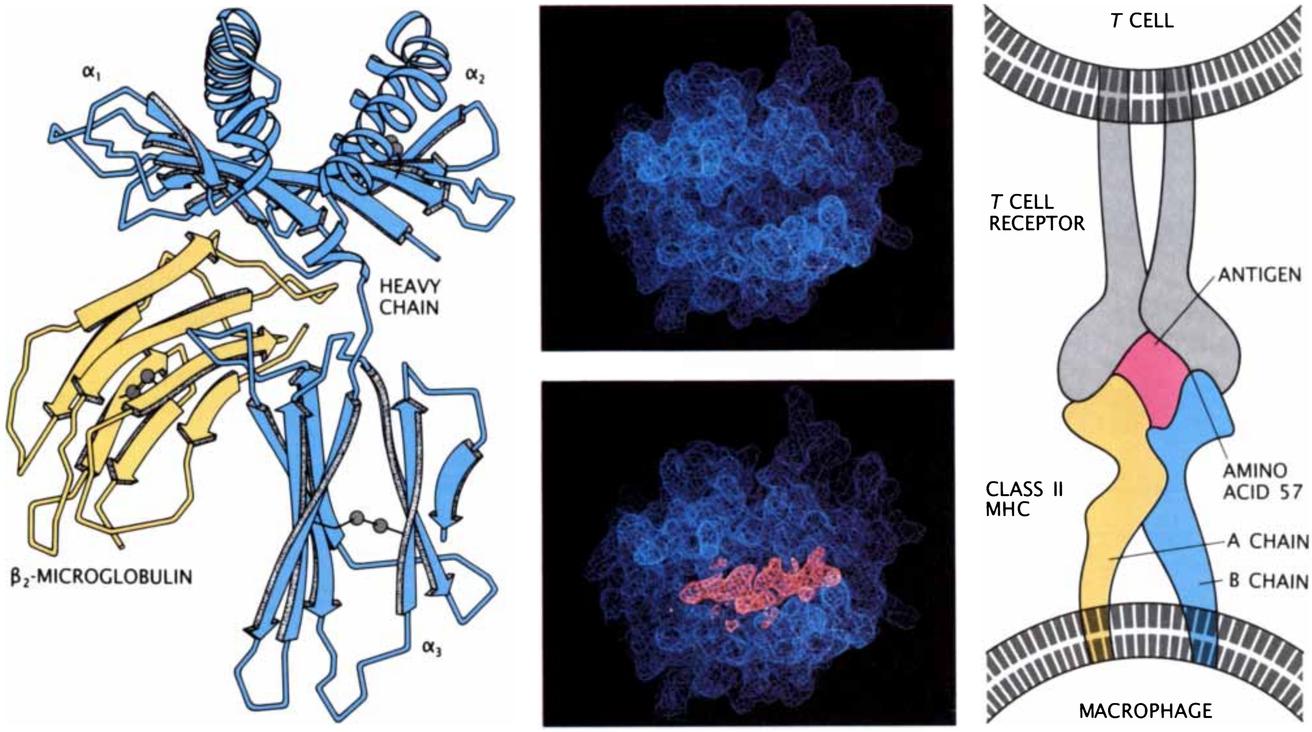


GENES in the major histocompatibility complex (MHC) on chromosome 6 have been associated with diabetes. The suspect genes lie in the D region, which is divided into three loci: DP, DQ and DR. All three encode so-called class II MHC molecules, each consisting of two amino acid chains (A and B). Because several of the genes for the chains are polymorphic, or variable (*color*), an assortment of molecular types occur (these are numbered). The genes encoding DR1, 3 and 4 are common in insulin-dependent diabetics. They probably do not confer susceptibility directly, but the DQ genes usually inherited in conjunction with them probably do. Genes that yield aspartic acid at the 57th position of the DQB chain tend to be protective; genes giving rise to other amino acids in that slot increase susceptibility.

Since then, exceptions have been discovered to the general rule that aspartic acid at that position is protective. Nevertheless, the lack of aspartic acid apparently contributes to diabetes in many people, and analyses of the three-dimensional structure of MHC molecules are beginning to explain why that is the case.

The basic structure of class I MHC molecules has been studied more thoroughly than that of class II molecules, but guided by that work, X-ray crystallographers have developed a fairly good idea of the conformation of the class II molecule. They have determined that the A and B chains combine to form roughly an X, with the A chain making up one side and the B chain the other. Together the upper segments of the chains form something of a "hot-dog bun," the inner cleft of which binds antigen (the "hot dog").

This model places the 57th amino acid of the B chain on the surface of the cleft, at a spot accessible to both an antigen and a *T* cell receptor. The properties of the amino acid in that position could strongly affect the conformation of the cleft (such as the angle between the two cut surfaces of the "bun") and, thus, the tightness with which different



ANALYSES of the structure of MHC molecules are helping to explain why a lack of aspartic acid at position 57 of the class II DQB chain can contribute to diabetes. Don C. Wiley of Harvard University and his colleagues found that a class I molecule (*left*) consists of a heavy alpha chain (*blue*) and a smaller, reversibly bound polypeptide, β_2-microglobulin (*yellow*). The first two domains of the alpha chain form a cleft for binding antigen. Representations by Wiley and his co-workers (*center*) also show the empty cleft from above (*top*) and occupied by antigen (*pink*). Class II molecules, which on macrophages present antigens to helper *T* cells (*right*), are probably similar to class I molecules, except that their A chain resembles a combination of the first alpha domain and the β_2-microglobulin, and the B chain resembles the rest of the alpha chain. The 57th amino acid on the B chain is thought to lie on the surface of the class II binding cleft, where it could affect the tightness of antigen binding or the recognition of antigen by *T* cells. Aspartic acid might result in weak binding of a diabetes-related antigen and no autoimmune response. Other amino acids might yield tight binding and a strong response.

antigens are bound. Hence, a cleft with an amino acid of neutral charge at that position might bind the autoantigen responsible for diabetes particularly tightly. This tight binding would increase the likelihood of recognition by *T* cells. It is also possible that direct contact between the 57th amino acid and the *T* cell receptor could influence *T* cell activity.

Further structural studies may help to explain why having aspartic acid at position 57 is not protective for everyone. Because there are several points of contact between any antigen and the binding cleft, the charge or shape of a particular amino acid at any of those points could influence the strength of antigen presentation by DQ molecules.

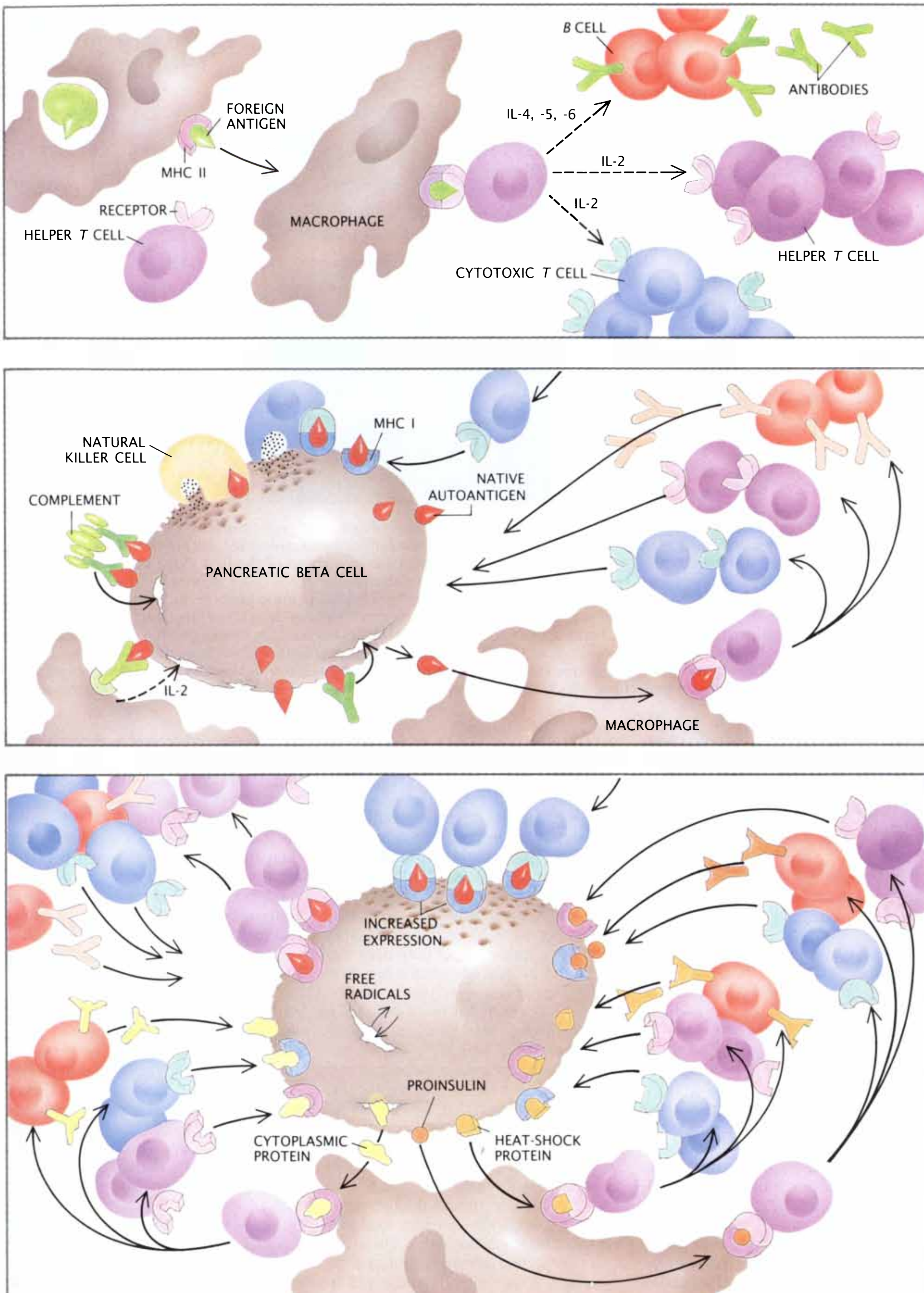
Associations between human diabetes and genes outside the major histocompatibility complex will probably be uncovered as well. As is true of NOD mice, many genes unquestionably contribute to susceptibility.

The genetic findings and other data collected thus far can be assembled into several scenarios of how insulin-dependent diabetes develops. In the view we favor, an individual who is genetically predisposed to diabetes is exposed to a foreign antigen that closely resembles a component of the beta cell (perhaps the 64-K protein). A macrophage or other antigen-presenting cell soon displays the foreign antigen in association with a class II MHC molecule (such as DQw3.2) that binds that antigen more tightly than is normal. Recognition by a helper *T* cell then unleashes an unusually vigorous and prolonged immune response to the foreign antigen, which may be further exacerbated if the *T* cell receptors themselves bind tightly to the antigen–class II MHC unit.

Eventually, the aroused immune-system cells, as well as antigen-specific antibodies, are carried by the bloodstream to the pancreas, where at least some beta cells are displaying a mimic of the foreign antigen. Because most cells in the body make class I MHC proteins, some molecules of the native autoantigen naturally become associated with the MHC molecules. These autoantigens then provoke an attack by cytotoxic *T* cells already sensitized to the foreign "twin" antigen. At the same time, antibodies raised against the foreign pathogen bind to the beta cell and begin to attract macrophages, complement and possibly NK cells.

So far the injured beta cells may be able to withstand the assault, but then the macrophages begin presenting the native autoantigen to helper *T* cells (perhaps at the same time damaging cells directly by releasing interleukin-1). At this point, the helper *T* cells multiply and amplify the immune response against the autoantigen.

Soon the damaged cells, perhaps in an attempt to identify themselves as self, increase their production of MHC molecules—including not only class I molecules but also, according to recent evidence, class II molecules. The class I molecules elicit increased bombardment by cytotoxic *T* cells. Normally the display of class II MHC molecules by nonimmune cells is thought to arouse the suppressive arm of the immune system. In this case, however, the effort by the beta cells backfires: because of a genetic defect in the immune system, the suppressor response does not occur. Instead the display of the class II molecules attracts even more helper


B CELL
ANTIBODIES
FOREIGN ANTIGEN
MHC II
IL-4, -5, -6
IL-2
IL-2
RECEPTOR
HELPER T CELL
MACROPHAGE
HELPER T CELL
CYTOTOXIC T CELL
MHC I
NATURAL KILLER CELL
NATIVE AUTOANTIGEN
COMPLEMENT
PANCREATIC BETA CELL
IL-2
MACROPHAGE
INCREASED EXPRESSION
FREE RADICALS
PROINSULIN
CYTOPLASMIC PROTEIN
HEAT-SHOCK PROTEIN

BETA CELLS UNDER ATTACK: A SCENARIO

1

INITIATING EVENT is a vigorous immune response against a foreign antigen (*green*) that closely resembles a normal component of beta cells. After macrophages ingest an invader somewhere in the body and present the mimic antigen (in tight association with a class II MHC molecule) to helper *T* cells, the helper cells secrete peptides—interleukins (IL's)—that promote activity by other helper cells as well as antibody-producing *B* cells and cytotoxic *T* cells.

2

IN THE PANCREAS, the sensitized cytotoxic *T* cells easily recognize the natural "twin" (*red*) of the foreign antigen—an "autoantigen"—on beta cells wherever the twin is bound by class I MHC molecules, which are ubiquitous. Antibodies bind the beta cells as well and thus impair them directly or by eliciting help from other components of the immune system, such as the toxic proteins that are collectively known as complement. Soon macrophages become involved and stimulate helper *T* cells to amplify the cell-mediated destruction.

3

FURTHER ATTACK follows as the beta cells become more and more damaged. The cells overproduce class I MHC molecules (thus promoting more activity by cytotoxic *T* cells). They also display class II molecules. Such display by cells not a part of the immune system should induce the system to suppress autoimmunity, but in prediabetics the opposite occurs: the display stimulates more helper *T* cell activity. As the beta cells become ever more damaged, substances normally sequestered within them, such as proinsulin, begin to appear at the surface of the cell. They arouse new waves of attack that in the end overcome the beta cells.

T cells and amplifies the immune response still further.

By now the hobbled beta cells and the activated macrophages surrounding them produce free radicals in quantity, and the beta cells are unable to inhibit the toxic effects. The beta cells may also display or leak proteins that are normally kept sequestered within the cell (such as heat-shock proteins, cytoplasmic gangliosides and proinsulin). Because such proteins are unfamiliar to the immune system, they are perceived as foreign. Macrophages engulf them and present them to helper *T* cells, thereby triggering another round of attack. The beta cells cannot withstand such a virulent assault, and they soon die.

As more and more insulin producers succumb, the remaining healthy cells must overwork in order to supply the needed insulin. This hyperactivity stresses the cells and perhaps leads to increased display of autoantigens. Hence, the surviving beta cells are killed with increasing speed. Eventually, too few cells remain to supply insulin to the body, and the symptoms of diabetes "suddenly" emerge.

In order to prevent the immune destruction of beta cells in susceptible individuals, physicians must first be able to identify those at risk. As we and several other groups have demonstrated, tests for detecting autoantibodies to the cytoplasmic autoantigen and to insulin can predict whether diabetes will develop in close relatives of insulin-dependent diabetics (who are at highest risk for the disease) as well as in groups of people chosen from the general population.

The 64-K autoantibody is also a predictor. Indeed, in the end, it may be the best choice for mass screening; it appears early, and few if any individuals who acquire insulin-dependent diabetes fail to produce it before diabetic symptoms appear. For now, the laboratory tests required to detect the autoantibody are too costly and cumbersome to be feasible for widespread use. Yet we and other workers are already deciphering the amino acid sequence of this molecule. Once that is accomplished, a simple screening tool can be devised.

There are no approved preventive therapies for people identified by such tests, but clinical trials are examining treatments that hold promise. For example, in some newly diagnosed diabetics, surviving beta cells have been protected by immunosuppressive agents such as cyclosporine, steroids and azathioprine, all of which reduce the overall population of *T* cells or inhibit activation of those cells. Spurred by this success, several groups are contemplating delivering such drugs experimentally as a preventive measure. The approach will undoubtedly succeed, but the benefits of the therapy will have to be weighed quite carefully against its considerable drawbacks. The drugs probably have to be taken indefinitely, and because they suppress the immune system, they increase the risk of infection and cancer.

In the future, once the trigger autoantigen for diabetes and its structure are identified with certainty, investigators should be able to construct selective agents: ones that will specifically eradicate the relatively small number of *T* cells bearing receptors for the autoantigen but that will also leave most *T* cells intact. For instance, one might create hybrid molecules consisting of the autoantigen and a bound toxin. When autoreactive *T* cells became bound to the antigen, the toxin would kill them. Such a treatment would be safe because erasing a small subset of *T* cells would not compromise immunocompetence significantly.

Just 10 years ago few investigators believed insulin-dependent diabetes would soon become preventable. Today there is growing confidence that before another 10 years pass, research into the causes of the disease will lead to safe preventive therapies. It has to. If the ravages of insulin-dependent diabetes cannot be prevented by insulin injections, then the disease itself must be eliminated.

FURTHER READING

EPIDEMIOLOGIC APPROACH TO THE ETIOLOGY OF TYPE I DIABETES MELLITUS AND ITS COMPLICATIONS. Andrzej S. Krolewski et al. in *New England Journal of Medicine*, Vol. 317, No. 22, pages 1390–1398; November 26, 1987.

A HYPOTHETICAL MODEL OF THE FOREIGN ANTIGEN BINDING SITE OF CLASS II HISTOCOMPATIBILITY MOLECULES. Jerry H. Brown et al. in *Nature*, Vol. 332, No. 6167, pages 845–850; April 28, 1988.

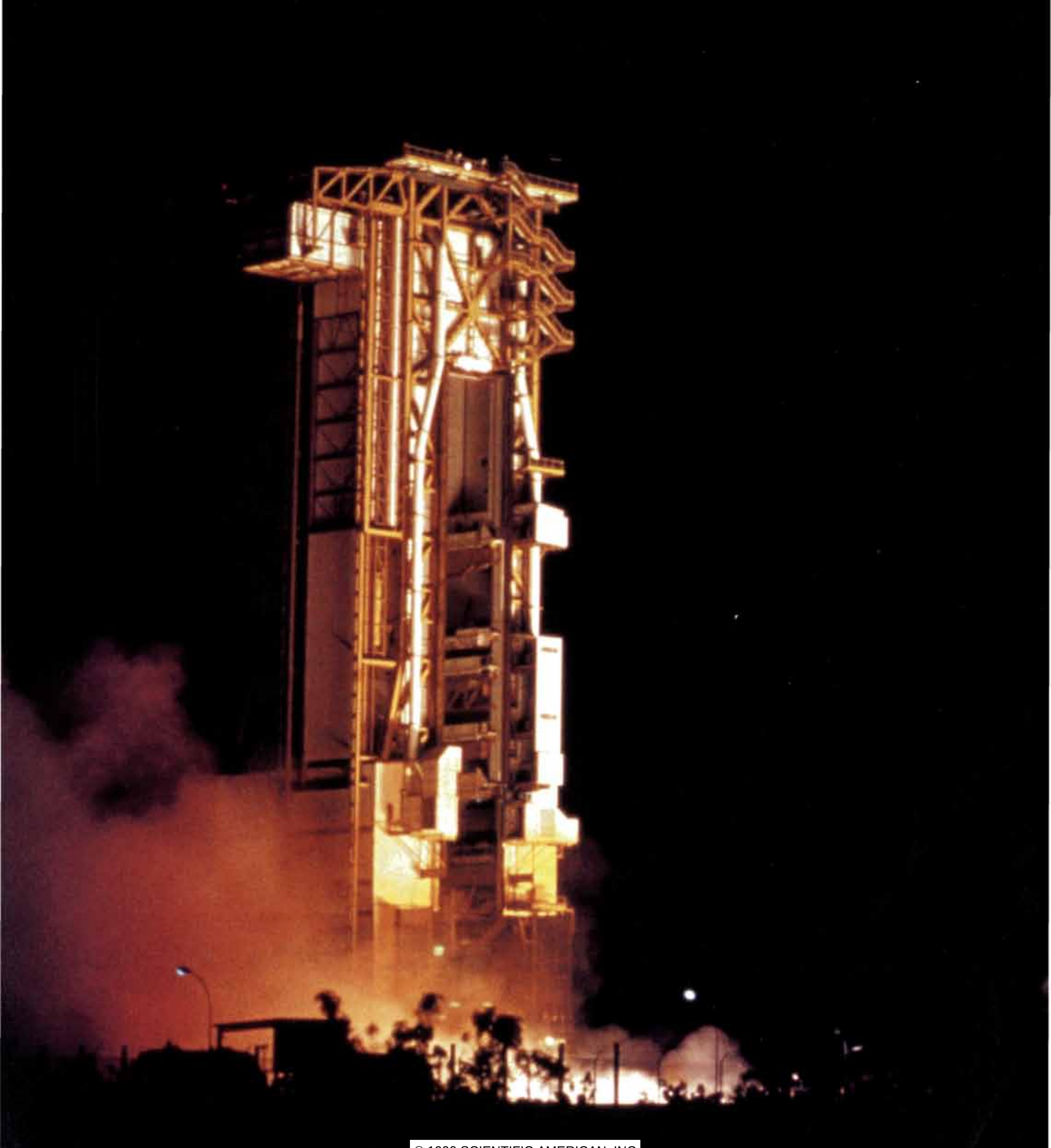
HOW, WHEN, AND WHY TO PREDICT IDDM. Noel K. Maclaren in *Diabetes*, Vol. 37, pages 1591–1594; December, 1988.

IMMUNOLOGIC TOLERANCE: COLLABORATING BETWEEN ANTIGEN AND LYMPHOKINES. G.J.V. Nossal in *Science*, Vol. 245, No. 4914, pages 147–153; July 14, 1989.

GENETIC CONTROL OF AUTOIMMUNITY IN TYPE I DIABETES. John A. Todd in *Immunology Today*, Vol. 11, No. 4, pages 122–129; April, 1990.

THE NEW SPACE RACE

by Elizabeth Corcoran and Tim Beardsley



October, 1989, launch of Ariane 4 from K

Getting off the ground has been the easy part. Now space explorers are seeking more commercial reasons for going into orbit.

On the evening of April 7, peasants clustered on the darkening hillsides surrounding China's Xichang launch site, waiting. Towering more than 43 meters above the launch pad sat China's Long March 3 rocket, ready for lift-off. Its mission: to boost *Asiasat I,* a communications satellite made by Hughes Space and Communications Group, into orbit.

The spectators did not wait long. Indeed, the launch may have been the easiest part of the exercise. Chinese officials had already weathered a political tempest in Washington, D.C., and secured special permission from President Bush to launch the American-made satellite. At 9:30 P.M. Long March soared eastward, jettisoning its payload 21 minutes, 27 seconds later. Within weeks, the satellite began transmitting data, television broadcasts and telephone conversations from countries across East Asia. Equally important: the launch put China among the runners in the space marathon of the 1990's.

The first space race was a contest to prove technological prowess and military might. Now the missiles of the past have blossomed into commercial launch vehicles. Even though national prestige remains important in the ongoing space race, this time the runners are struggling for positions that will command future commercial opportunities. Some half a dozen nations are competing to build businesses—or at least trying to earn some money—by launching commercial payloads. But they face a big problem: "commercial space" is still an oxymoron.

There are too few customers to fill all the open slots on launch schedules. Communications satellites are now the only space venture that the private sector can afford, but the world's appetite for these will soon be sated by about 15 launches a year, according to the Paris-based research company Euroconsult. In contrast, the number of available rides to space this year is closer to 40. Making matters worse, satellites are lasting longer because their components are becoming more durable. Fiber-optic cables are an attractive earth-based alternative for long-distance transmissions (except in remote areas such as Indonesia).

Even so, governments are determined to stay in the space race. Communications and satellite intelligence have become critical not just to national security but to countries' economic health. And then there is national pride. "We have a choice now," says D. Allan Bromley, President Bush's science adviser and a member of the National Space Council. "We can either be a major player on this new frontier, or we can pull back."

So nations are turning to the commercial arena to subsidize their launch capability. New—albeit uncertain—markets hold promise. Among the possibilities are direct-broadcast satellites for television and satellites for consumer navigation systems. The military is exploring "lightsats," small satellites weighing a few hundred kilograms that could replace systems damaged during a crisis. Networks of these could also be used in civilian communications.

And then there is the elusive goal of manufacturing in space. The best hope rides with experiments that investigators and small companies are now putting into orbit. Given a decade or so and enough experiments, some might prove that working in a microgravity environment is lucrative.

Outside the U.S., governments are working with industry to finance space programs. European countries, particularly West Germany and Italy, are vigorously pursuing microgravity experiments. So is Japan. The Soviet Union—with its fleet of reliable rockets, its *Mir* space station and its reentry capsules for bringing experiments back from space—seems to have the most robust space infrastructure.

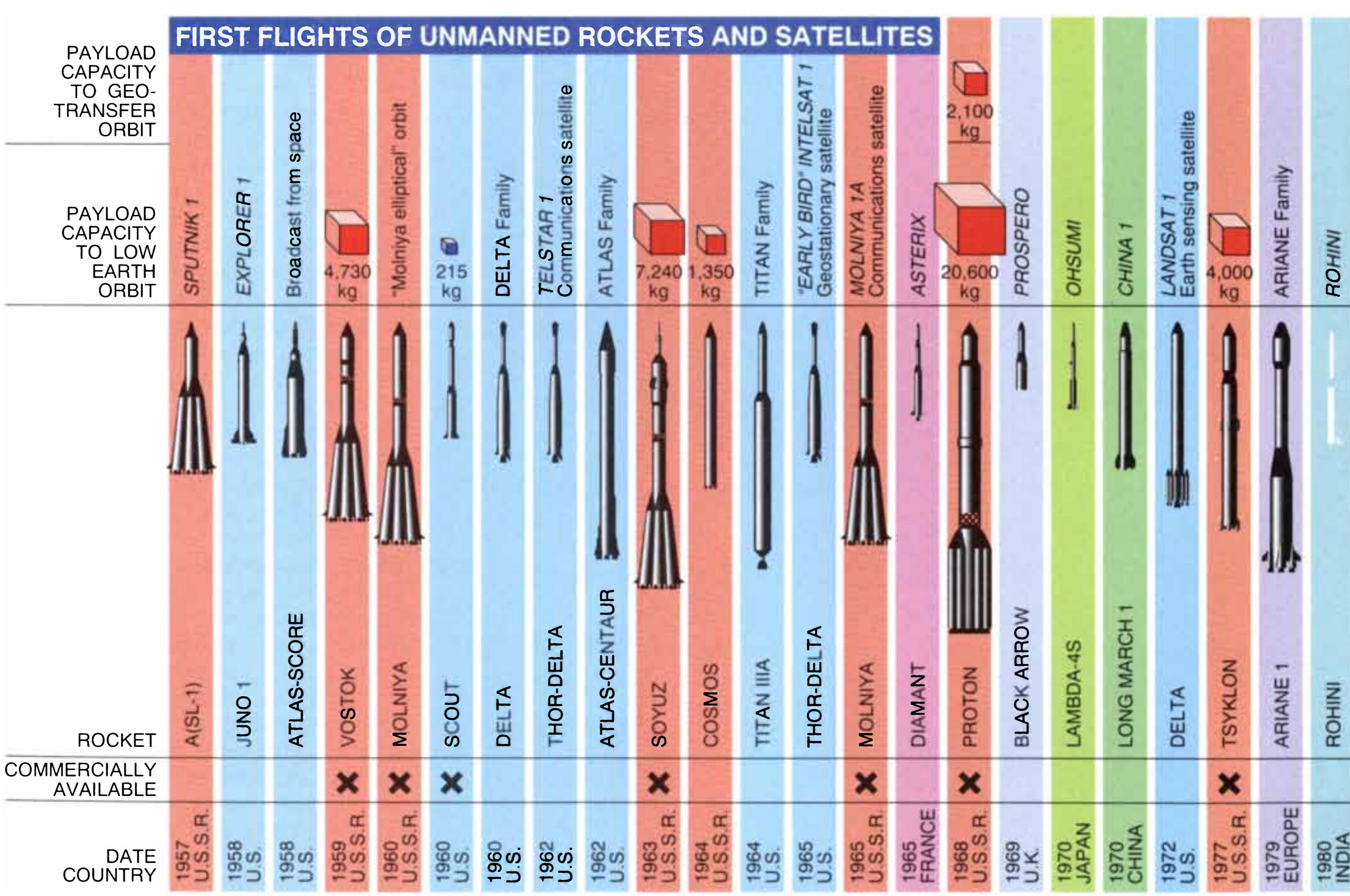
In the U.S. the world's largest rocket makers await signals from the government. Large corporations are shying away from experimenting in space because their calculations indicate the risks are too great. As a result, start-up companies seeking niches in space look promising. And they are learning a lesson the large companies have yet to take to heart: building a launching business means developing a service, not just hardware. "You can't just build metal, add 10 percent to its cost and call it a business," says Charles Chafer, who in 1980 helped found Space Services in Houston, Tex.

NASA Fumbles

Unfortunately, slapping a price on a piece of metal has been the way the space business was run for decades. Through the 1970's any company in the world hoping to launch a commercial payload had one choice: the U.S. National Aeronautics and Space Administration. The U.S. offered variations of a quartet of vehicles: heavy-lift Titans (built by Martin Marietta), mid-size Deltas (from McDonnell Douglas) and Atlases (from General Dynamics), and tiny Scouts (made by LTV Aerospace and Defense Company). The contractors built the vehicles; NASA acted as a broker and ran the launches.

Then NASA tried to replace these expendable vehicles with the space shuttle. It had argued the shuttle would be a cheaper way to get into orbit. (This proved true only when the launches were subsidized; one estimate of the cost of a dedicated shuttle flight today is far more than $350 million.) As a result, commercial and almost all military satellites were booked onto the shuttle. But NASA had trouble running a commercial launching business. Flights were often delayed, upsetting the plans of satellite owners.

So many satellite builders began looking to the fledgling European rocket Ariane 1 as an alternative. Incorporated in 1980, Arianespace's founders took a businesslike approach from the beginning. They deliberately kept the bureaucrats of the European Space Agency (ESA) out of the day-to-day business of building and launching rockets. Ari-



anespace helped customers arrange financing for flights. By the end of 1985 Arianespace had won about half the commercial contracts for launching satellites. "NASA treated its clients like a lord accepting peasants on his land," asserts Charles Bigot, managing director of Arianespace. "Arianespace treated them like a retailer."

Although the Reagan administration permitted U.S. rocket makers to compete for commercial launching contracts, not one was willing to take on the shuttle. The explosion of the *Challenger* shuttle in 1986 changed that; the U.S. government subsequently agreed to ban NASA from launching commercial payloads. But U.S. contractors still faced stiff competition. Arianespace had won a loyal following and was steadily improving its vehicles. ESA covered the heavy technology-development costs. Arianespace earned its operating expenses with satellite flights (and now needs to win about six or seven annual launch contracts).

Time is not helping U.S. launching companies' competitiveness. The government continues to have trouble figuring out what it means by a "commercial" launching industry. Contradictory priorities from different federal agencies buffet U.S. space policy. The military, anxious to maintain a mixed fleet of rockets, has awarded almost enough contracts to sustain the three large-booster companies for the next five years. But the contracts may have dampened the companies' interest in ferreting out commercial contracts.

Mixed Signals

"Most of our customers are repeat customers" who flew payloads on board Deltas when NASA ran the launches, says Samuel K. Mihara, marketing director for McDonnell Douglas's Delta. One commercial launch a year would be average business, he says. More than one would be good. Martin Marietta has secured 42 firm contracts for Titan launches, scheduled to take place through the late 1990's. Forty-one are devoted to Air Force cargo; one will carry NASA's *Mars Observer* probe. Industry sources suggest that Marietta may drop out of the commercial market altogether—a rumor the company denies.

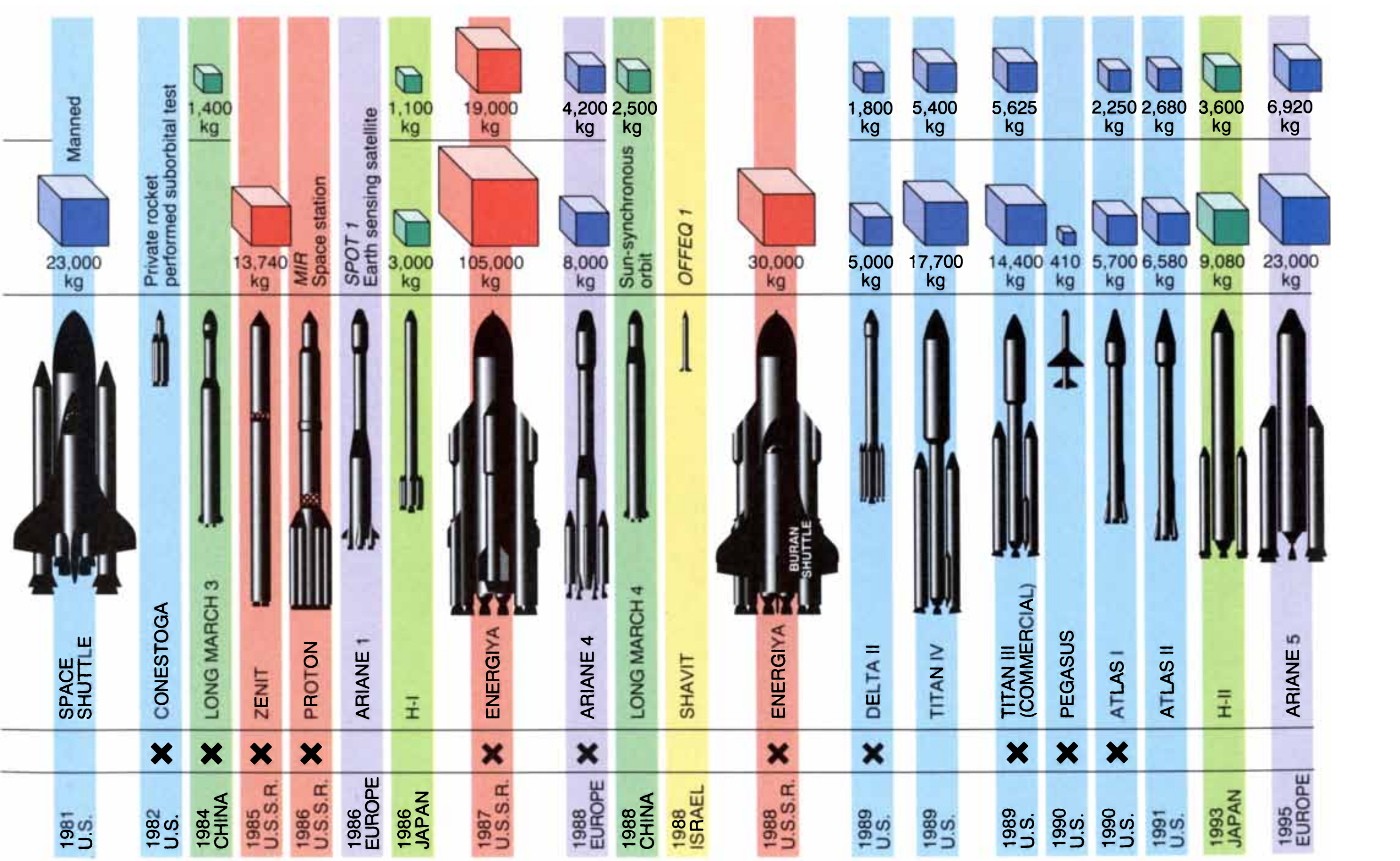
The mixed signals have not helped the traditional aerospace contractors adjust to their new roles. "To this day, I still haven't figured out what the hell commercial space is," says Jack Whitelaw, a manager for LTV.

Nor have the large U.S. companies been willing to lend much support to start-up companies keen to venture into space. "Large U.S. aerospace firms have not played a strong strategic investment role in new commercial space ventures," notes a recent report from the Department of Commerce.

So even small U.S. companies planning space ventures are looking overseas for funds. SPACEHAB, a company in Washington, D.C., that is building a module to fly on the shuttle, found willing investors in Japan. Geostar, also in Washington, D.C., which is developing a satellite-based tracking system for the trucking and railroad industry, similarly relies on support from Japanese and French investors.

Meanwhile other governments have begun hoping that commercial launches may supplement their space budgets or add to their hard-currency earnings. Both the U.S.S.R. and China are now wooing commercial satellite customers with the hallmark tactics of capitalism—low prices and polite service. These countries worry more about the political barriers to winning contracts than breaking even on launches.

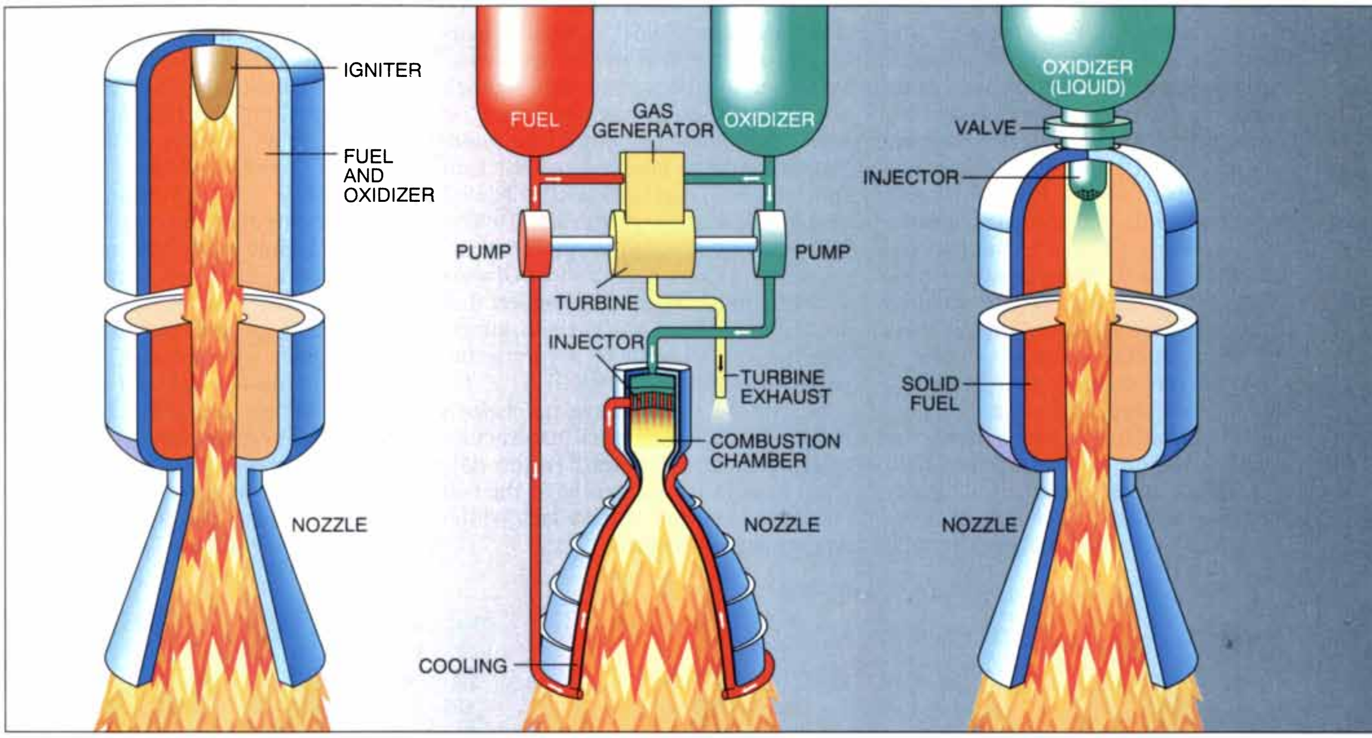
Initially the Soviets tried to attract commercial satellites by offering bargain-basement prices. They were rou-



tinely rejected. Now Glavkosmos, the Soviet marketing arm for space services, has honed its pitch. The Soviets are betting that the reliability of their rockets and diversity of space infrastructure will draw customers. Glavkosmos has had some success. INTOSPACE in Hanover, West Germany, has flown experiments with the Soviets, as has Payload Systems in Cambridge, Mass.

The Chinese, on the other hand, are still apparently offering cut-rate prices—which would violate their pledge to the U.S. to sell services at prices comparable to those charged by the West. A recent Chinese bid to fly an Arabsat communications satellite "was about half of what we or McDonnell Douglas would have to charge," asserts Douglas Heydon, president of Arianespace's U.S. subsidiary. The office of commercial space in the Department of Transportation was investigating the concerns in early May.

And yet another competitor, namely, Japan, is coming up. Its National Space Development Agency (NASDA) has managed a small but rapidly growing launch program since 1969. In exchange for U.S. rocket technology in



Engines to beat gravity...

Like jet airplane engines, chemical rockets work by burning a fuel to produce hot gases. But unlike jets, which consume oxygen from the air, rockets carry an oxidizer on board so they can operate outside the earth's atmosphere. A key objective of rocket design is to create engines that produce the largest sustained force, or thrust, from the smallest weight of fuel. Engineers call the ratio between thrust and fuel consumption specific impulse.

The least complicated rocket engines—and generally those with the lowest specific impulse—are solid-fuel boosters. They consist of a hollow cylinder of a rubbery propellant that is a mixture of fuel and an oxidizer, such as ammonium perchlorate. Although solid rockets are highly reliable, when ignited they burn like giant firecrackers—and are equally hard to extinguish.

Liquid-fuel rockets have a higher specific impulse but are more complicated and expensive. Both fuel and oxidant, which are carried in separate tanks, must be pumped or sprayed rapidly into a combustion chamber. Controlling the rate of pumping allows the thrust produced by the engines to be regulated, or throttled. One common pairing of liquid fuel and oxidizer is unsymmetric dimethylhydrazine and nitrogen tetroxide, which ignite spontaneously when mixed. Using a liquid hydrogen-oxygen combination provides even greater specific impulse, but the materials are dangerous and difficult to handle.

Within the family of liquid rockets are several variations. One type powers the fuel and oxidizer pumps by circulating fuel through pipes embedded in the nozzle wall. The nozzle heats the fuel, building up enough pressure to turn the turbine that drives the pumps. The liquid hydrogen-oxygen engine in the first stage of the Ariane 5 will use a gas generator—a closed chamber in which a small amount of fuel and oxidizer is burned to produce high-pressure exhaust. The exhaust then drives the pumps.

Japan's developmental LE-7 engine is like the space shuttle's main engine—a staged-combustion liquid engine. Some of the hydrogen is ignited in a preburner by a small stream of oxygen. The resulting hot exhaust drives the pump turbines. Then the exhaust, with its large component of unburned hydrogen, is channeled into the main combustion chamber, where it burns with additional liquid oxygen.

Hybrid rockets use a rubbery fuel resembling solid-rocket propellant. But the oxidant, typically liquid oxygen, is sprayed from a nozzle down the central bore of the fuel cylinder. The engine burns like a solid rocket from the inside outward.

On paper, hybrid rockets offer many advantages. They can be throttled and have more specific impulse than do solid rockets. But apart from powering some Air Force drone missiles and a prototype launch vehicle designed by a start-up company, hybrid engines have yet to find much use.

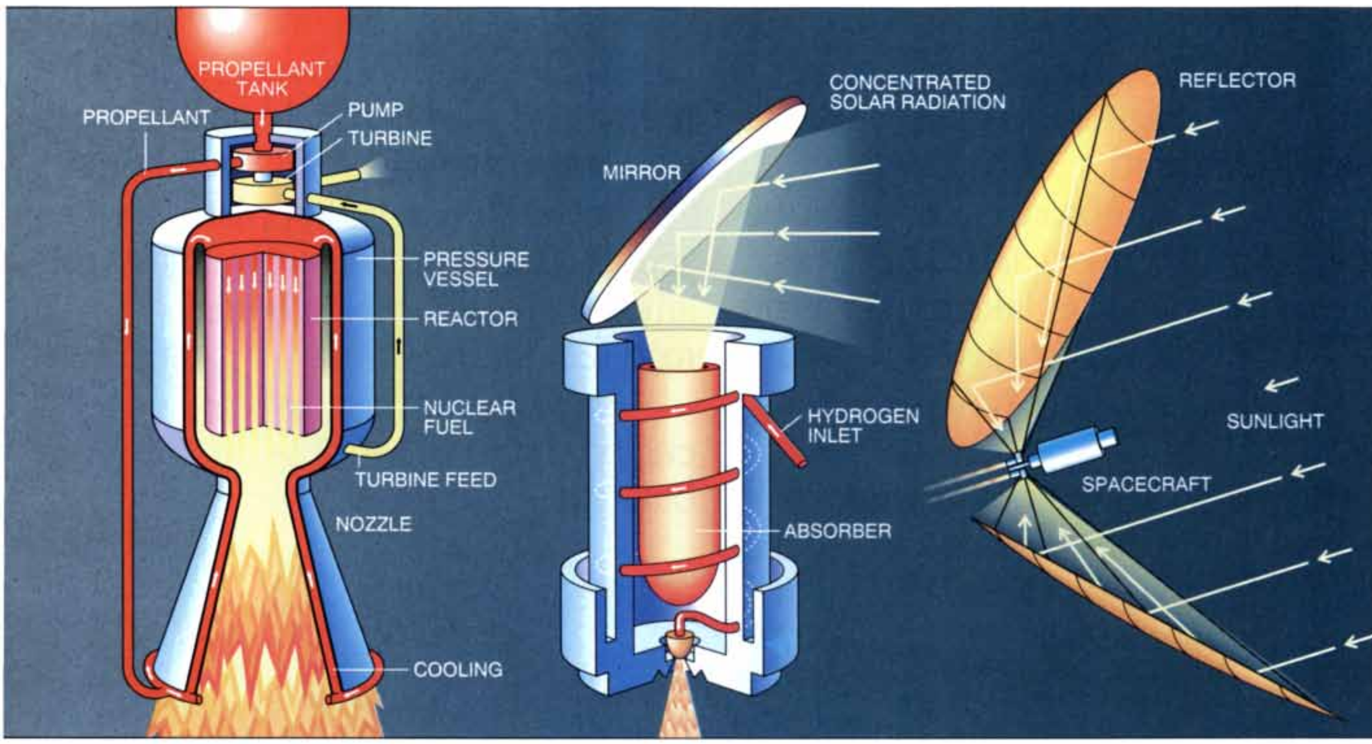
the 1970's, Japan agreed to launch only Japanese payloads. But a new rocket, the H-II, which is set for its first launch in 1993, is based entirely on Japanese design. This will free Japan to launch other nations' satellites.

Although Japanese officials maintain that their interest in promoting a commercial H-II program is slight, 13 H-II contractors, led by Mitsubishi, are due to establish formally a new company late this month to develop the rocket for NASDA. "Japan began development of the H-II with the knowledge that commercial launches would occur in the future," says Koji Sato, a member of Mitsubishi's coordination office for the new company.

Hurdles remain. Fishermen working near the Tanegashima Space Center convinced the government to limit rocket launches to two 45-day periods every year, so Japan is watching with interest the development of a commercial launching center in Australia. Moreover, the appreciation of the yen has helped push the estimated launching costs for the H-II out of the commercial arena. "If cost reductions cannot be achieved, this company will remain at a

NUCLEAR MOTOR

SOLAR MOTOR



...and to get through space

Once payloads are in orbit, they have to maneuver. Small rockets, called thrusters, keep satellites in orbit. Other propulsion systems are required to put satellites into geosynchronous orbit—or to guide them to distant points.

With no gravity to fight, low thrust suffices. But such rockets must still produce high specific impulse. Orbital adjustment thrusters heat propellant, which is squirted out of a nozzle. The hotter the gas, the faster it shoots out of the nozzle and the higher the specific impulse. An improved type of thruster, called an arcjet, uses an electrical arc to heat a propellant gas to several thousand degrees. Increasing the specific impulse of the system could add years to a satellite's life. GE Astro-Space's *Telstar 4* satellite, slated for launch in June, 1993, will be the first to use arcjets.

Renewed interest in sending humans to Mars has prompted NASA to reexamine nuclear propulsion, which offers even higher specific impulse and could cut trip time by two thirds. Two schemes bear watching. In one, a nuclear reactor heats a gas, such as hydrogen, to produce thrust. NASA built and test-fired such a rocket, called NERVA, in the late 1960's. In the second, a reactor produces electricity for driving an electrical thruster, such as an ion thruster.

Ion thrusters, which use electrostatic fields to accelerate ions, were first tested nearly 20 years ago. Japan will inaugurate a functional ion thruster on its *ETS-6* satellite, set for launch in 1994. An alternative type of electrical propulsion, called magnetoplasmadynamic, uses magnetic fields to help accelerate the ionized gas. The Soviet Union has performed several flight tests based on this technology, which works best at high power.

For maneuvers nearer to the earth, the U.S. Air Force is investigating lightweight solar propulsion. Reflectors measuring 120 meters in diameter would focus the sun's rays and heat propulsion units to more than 3,500 degrees Celsius. The hot units, in turn, heat on-board hydrogen, which produces small amounts of thrust. It could take as long as 40 days for a solar drive to ferry a satellite from low earth to geostationary orbit, and so solar drives are most valuable for nonurgent payloads. But because a drive is light, a smaller booster—say, a Delta instead of a Titan—could initially launch the payload and so save $50 million.

In the offing are some very speculative concepts. One would utilize the energy that binds atoms of hydrogen into molecules. Recombining split hydrogen molecules releases enormous energy. If 15 percent of a sample of molecules could be split and stored as single atoms—a distant goal—atomic hydrogen might be a viable fuel. Another technique would use the energy from laser beams to power a spacecraft—and one day perhaps boost payloads into space on a column of laser light. Some investigators are even thinking about antimatter engines.

Send in t

There's a good reason why more cars are looking more and more alike. It's called the Accord.

The Honda Accord has become the benchmark for other car makers. They take it apart. Study it. Take it apart again.

Now while they're busy doing that, we would like to move on to something bigger and better. The new Accord.

The new Accord was redesigned to create a new car. A car attuned to the driver. One that improves the quality of driving.

There is more room inside, the ride is more comfortable and the car is more responsive under all driving conditions.

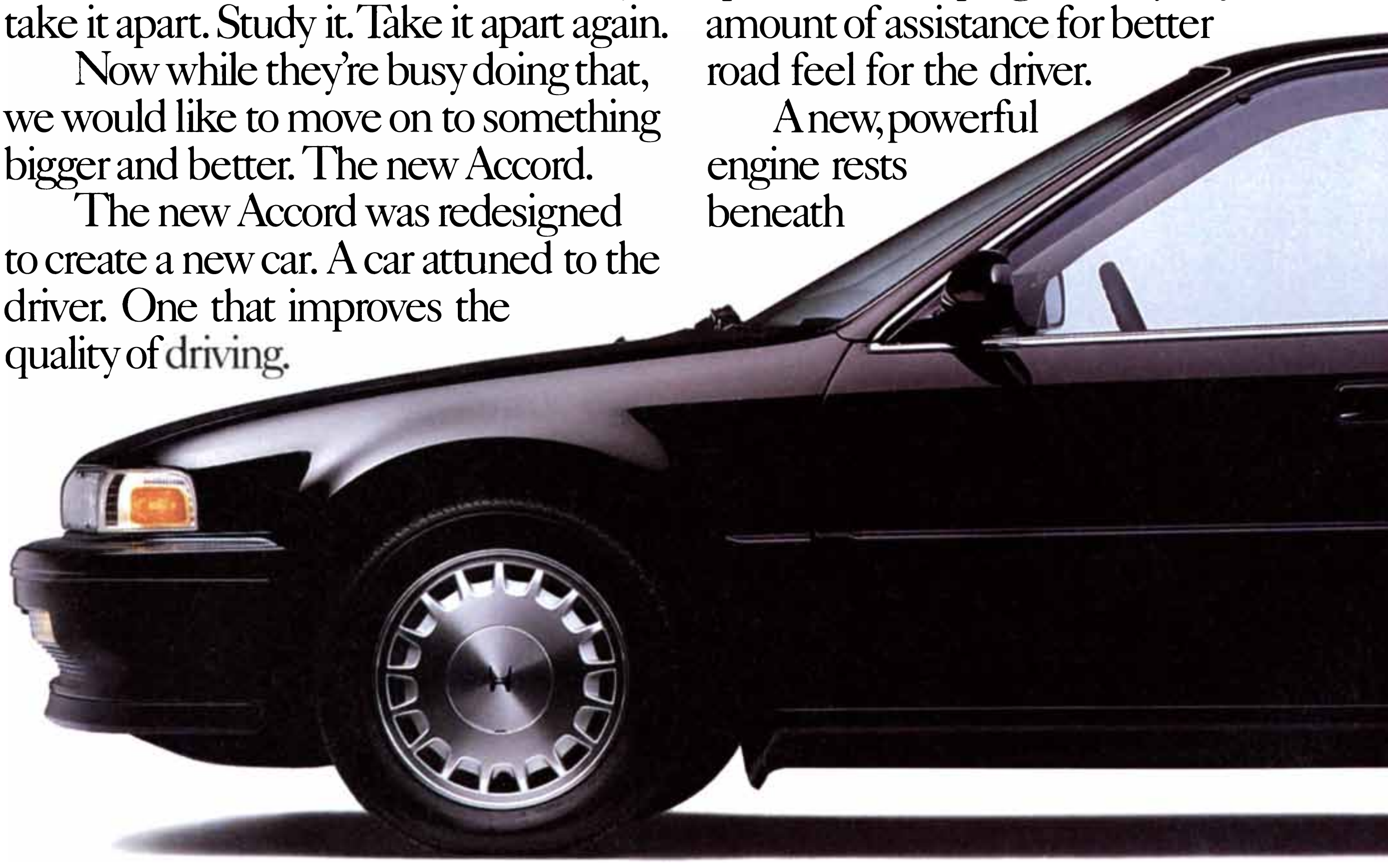
You can see the smooth, flowing lines and the substantial increase in glass area. But the greater refinements are inside the car, where they matter most.

The body is built with thicker steel panels. So it's stiffer and more solid. The redesigned 4-wheel double wishbone suspension system works even better.

Power steering now operates at all speeds. And it progressively adjusts the amount of assistance for better road feel for the driver.

A new, powerful engine rests beneath the low, sloping hood. It's fuel injected. Unique balance shafts cancel inertial forces that cause other engines to twist and shake. You'll be amazed at just how smoothly and quietly it runs.

There is a new 4-speed automatic transmission available that's compact, smooth shifting and even more efficient.



he clones.

Once seated inside the car, you are welcomed by a bright, airy environment that immediately puts you at ease. The controls and instruments are logically placed. Everything falls into your line of

sight and is easily within your reach.

The front seats are larger and offer greater support. Rear seats contour to your shape for better comfort.

The interior is designed to give the appearance of smooth, seamless space. Panels are moulded from one piece so there is nothing to rattle. The carpeting is thick and plush and helps to absorb noise and vibrations.

A new ventilation system further assures your comfort. It's quiet, efficient and the controls are easy to use.

When designing the new Accord, Honda engineers kept one thought foremost in mind. To build a car that will make you more pleasantly aware of the enjoyment of driving. You will have to drive it to believe it.

And you should keep this in mind. If you are thinking about buying a new Accord, make sure it's the one built by Honda.

HONDA

The New Accord

low profile, just selling hardware to the government," Sato says.

Marginally reducing the cost of flying will not change the private sector's disinterest in working in space in the near term, however. It will simply help current users reduce their initial capital expenditure. "Without a valid business purpose, no one will go to space at any price," says Albert D. Wheelon, retired chief executive of Hughes Aircraft. Expensive rockets are bought only by governments and commercial satellite companies. The U.S. government spent about $25 billion on space goods and services in 1988. That same year the commercial sector spent about $1.8 billion (and $2.7 billion in 1989), more than 90 percent of which is attributable to satellite communications.

From a corporate perspective, most space ventures must return revenues that are at least double their costs because of the risky nature of the business. Last February, for example, a failed Ariane 4 launch deposited two Japanese satellites worth $200 million into the ocean. In March, a commercial Titan put a $150-million *Intelsat 6* satellite into a useless orbit. Consider space from the vantage of GTE Spacenet Corporation in McLean, Va. To replace three of its aging satellites, GTE has budgeted almost $200 million to build the spacecraft, at least $210 million to launch them and another $70 million for insurance. That puts GTE's total cost for one mission at roughly $160 million. On the other side of the ledger, however, GTE expects the satellites to recover their costs in less than three years and to continue generating earnings for another seven.

Satellite owners would doubtless switch their business to cheaper but equally reliable flights. "If you're chairman of an American satellite company and your choice is $80 million at Cape Canaveral or $15 million with Great Wall [China's national launching company], you put up the $15 million and travel to China," says Joseph P. Allen, president of Space Industries International in Webster, Tex., and a former astronaut. "It's a no-brainer."

Profit Squeeze

Even so, prices are far from low enough to turn other space activities into lucrative ventures. Remote sensing, for example, "is a good way to lose money, hand over fist," says John E. Pike, a space-policy analyst at the Federation of American Scientists. Neither the U.S. Landsat program nor the French SPOT Image is a self-supporting business. The largest planned remote-sensing program over the next few years will be funded by NASA.

Although corporate interest in manufacturing in space flared in the early 1980's, it was doused by the irregularity of launches and by new— and far less costly—earth-bound techniques for accomplishing the same task. "There was a lot of hype," recollects Earl L. Cook, who directs the 3M Corporation's space research. 3M remains one of the few large U.S. companies that still sends experiments into orbit on board the space shuttle. And NASA does not charge for these flights.

"It's almost too bad that communications satellites succeeded so fast," says H. Guyford Stever, science adviser to presidents Nixon and Ford. "It gave us a false impression. It got people thinking there must be a lot of nuggets up there," he says. They may be up there, he adds; business has not yet learned how to squeeze from them enough revenues to justify going into space.

So national space goals will continue to drive advances in rocket technology. During the next few decades a range of mammoth government projects—free-flying orbiting platforms (which astronauts would visit occasionally), manned space stations and possible missions to the moon and to Mars—will require hauling huge amounts of materials into low earth orbit.

As a result, some space-vehicle orga-

Launch of Japan's H-I rocket from the Tanegashima Space Center.

nizations are straining to increase the payload capacity and cut the costs of launching their next generation of vehicles. ESA's development of Arianespace's next family of vehicles, the Ariane 5, may be the most aggressive project. Slated for its first launch in 1995, Ariane 5 will have two stages and rely on a main engine fueled with liquid hydrogen and liquid oxygen, supplemented by solid boosters. This should enable Ariane 5 to ship 1.5 times the payload weight handled by its predecessor into geotransfer orbit. Ariane 5 will have an option for its second stage: replacing the unmanned cargo bay with the *Hermes* manned spaceplane.

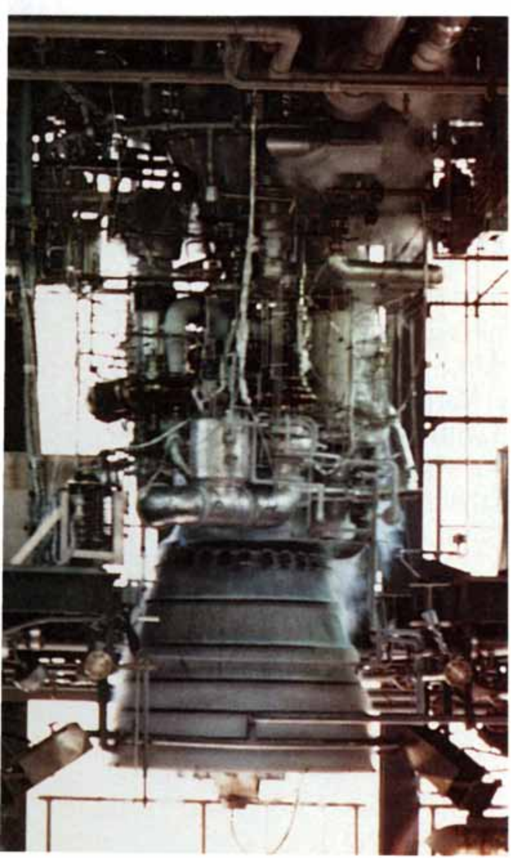
Although Arianespace relies heavily on proved technologies, innovations are working their way into Ariane 5. The solid-booster nozzles are lined with carbon-carbon, an advanced composite of woven carbon fibers impregnated with resins and then sintered. Ongoing tests are evaluating a carbon-silicon carbide nozzle for the liquid engine. To lower the operating costs of the larger vehicle, Arianespace is planning a mobile, ground-control system that will trim the time between launches and reduce the crew needed. As a result, Ariane 5's operating costs may be only 85 percent of those of Ariane 4—about $12,000 to $15,000 per kilogram to geotransfer orbit and half that to low earth orbit, says Arianespace's Heydon.

Japan's H-II will employ a complex liquid hydrogen-oxygen system in an engine called the LE-7. The LE-7 "is very efficient, but its development is very difficult," concedes Masafumi Miyazawa, who directs NASDA's propulsion-systems group. Two disastrous tests of the engine last year pushed back the H-II development schedule some 12 months. Still, Japan has already successfully built a highly reliable second-stage engine—the LE-5A—that can reignite in orbit. The cost of developing the H-II will be more than ¥ 250 billion ($1.7 billion), $470 million of which is devoted to the LE-7.

The future U.S. launching picture looks bleak in comparison. Because of the policy of relying on the shuttle, the U.S., although long the world's leader in rocket innovations, spent virtually nothing on other advanced propulsion and launch systems between 1972 and 1986, points out John M. Logsdon of George Washington University. The Aerospace Industries Association is seeking redress by proposing that the government spend $5.5 billion over 10 years on rocket research—and is invoking the specter of foreign domination of the launch market to buttress its demands.

Current U.S. aerospace materials research aims to satisfy the demands of the National Aero-Space Plane (NASP), an air-breathing vehicle that would take off and land like a plane but travel at an unprecedented 25 times the speed of sound and go into orbit. A prototype is unlikely to be built before the year 2000. Still, NASA hopes that high-performance materials being developed for the plane will benefit rocket engines. These include copper reinforced with fibers of graphite, tungsten or niobium for conducting heat. Silicon carbide, silicon nitride and aluminum oxide fibers are being used to strengthen ceramics. Even more heat-resistant ceramics, based on zirconium and hafnium, are in the works.

Until recently the Advanced Launch System (ALS) program, initially proposed for the Strategic Defense Initiative Organization, was the principal U.S. rocket-technology program. In theory, the ALS would be a family of reliable rockets to boost as much as 200,000 kilograms into low earth orbit for as little as $660 per kilogram. The effort, originally expected to go on through the 1990's and cost up to $12 billion, was soon derided by critics as too ambitious, too expensive and unnecessary. Last December, under pressure from Congress to cut its budget, the Air Force dropped the effort.



Test firing of Japan's LE-7 engine, the first stage of the H-II rocket.

New Elements, Old Vehicles

The ALS project aimed to trim launch costs by designing components for easy manufacturing and by employing more automation for testing parts. Martin Marietta is developing new liquid hydrogen and oxygen tanks made from an aluminum-lithium alloy that could be extruded or cast instead of milled. These might be used on future Titans. Rocketdyne has designed a fuel pump with only one weld; comparable pumps on the shuttle have 150.

The government also hoped the ALS contractors would incorporate such innovations into their current vehicles. To some extent, they are. Martin Marietta says the ALS technologies will help reduce Titan IV launch costs by one third to about $4,400 per kilogram to low earth orbit. Engineers note, however, that only limited efficiencies can be gained by trying to fit new elements into old vehicles.

Although the full ALS agenda has been canceled, NASA and the defense department are still funding rocket manufacturers Aerojet and Rocketdyne to develop a prototype engine to the tune of $107 million in fiscal 1990. The engine "is supposed to be a cheaper, expendable liquid oxygen and hydrogen engine," says James R. Thompson, Jr., deputy administrator at NASA. But because the work is funded annually by Congress, its future is uncertain.

In lieu of a new vehicle, NASA is half-heartedly considering refitting the liquid-fuel engines and solid-fuel rocket boosters from the shuttle onto an unmanned cargo vehicle dubbed *Shuttle C.* The program does not hold many technology advances for the U.S., however. Even these tentative plans are in limbo. "Frankly, it will all come back into focus once we all agree—and this is going to take several years—on a schedule to implement the space exploration initiative"—the manned missions to the moon and to Mars, Thompson says.

While NASA sets its sights on Mars, tiny U.S. companies are betting they can earn a living by putting payloads of a few hundred kilograms into low earth orbit. But first they have to improve their track record. The earliest space start-up, Space Services, conducted a suborbital test flight of its Conestoga solid-fuel rocket in 1982 but has yet to run a successful orbital flight. An attempt by the American Rocket Company in Camarillo, Calif., to fly an innovative hybrid-fuel vehicle late last year ended in a smoldering heap. E'Prime

Aerospace Corporation in Titusville, Fla., hopes to piece together sections of Peacekeeper missiles.

At the front of the pack is Orbital Sciences Corporation in Fairfax, Va., with its Pegasus vehicle. Pegasus flaunts a triangular wing and an unusual launch technique: it rides beneath the wing of a large aircraft such as a B-52 or 747 to about 12,000 meters and takes off horizontally from there. In April, Pegasus lofted its first cargo into a polar orbit about 600 kilometers high—two experiments weighing 220 kilograms in all.

Orbital Sciences adopted the strategy that has worked well for the major rocket companies; it has secured the U.S. government as an anchor client. The Defense Advanced Research Projects Agency largely sponsored that first flight and is also helping support Orbital Sciences' Taurus, scheduled for flight next summer. This vehicle, which will take off with only 72 hours of preparation, will eventually loft more than 1,360 kilograms into low earth orbit.

For now, the start-ups are pinning their hopes on lightsats, a new class of small satellites that weigh only a few hundred kilograms. From the vantage of the Defense Department, which is funding much of this work, lightsats could quickly replace satellites destroyed or jammed during a military crisis. Lightsats will augment large military satellites, says Terry A. Higbee, program manager for lightsats at the Ball Aerospace Systems Group, who predicts demand in the U.S. could reach 10 per year by the mid- to late 1990's. The small launch companies hope there will be a commercial market as well. Last February, Orbital Sciences requested permission from the

Just call Alexander I. Dunayev. He presides over Glavkosmos, the Soviet Union's civilian space organization, from a makeshift office lodged in a dingy, yellow-brick 12-story block of residential apartments in Moscow. Outside, pensioners shuffle by as children play noisily.

Why does Dunayev, who is spearheading one of the Soviet Union's most successful endeavors, put up with these conditions? "In the Soviet Union," a Glavkosmos official explains, "it is easier to build a launch pad than find suitable offices."

For the past decade, the Soviet Union has launched about 90 rockets every year. Its factories crank out boosters with an efficiency that would make Henry Ford sigh. After more than 1,300 launches of its Soyuz-class rockets and 117 of the more powerful Protons, Soviet launchers are the cheapest and most reliable in the world. "The U.S.S.R. still makes the same boosters it used to launch the Sputnik in 1957," says Phillip S. Clark, a rocket specialist at Commercial Space Technologies in London.

Soviet leaders are now hoping to capitalize on those rockets to bolster their sputtering economy. Glavkosmos began advertising its launching services to the West for as little as $25 million a shot in 1986—a bargain compared with the $70- to $100-million rides on comparable Western rockets. The Soviets calculate that they could devote at least 10 launches a year to foreign payloads. In addition, Glavkosmos offers the four-year-old *Mir* space station and cosmonauts with more than 16 man-years of experience in conducting space experiments.

But the Soviet Union faces significant obstacles before it can win many Western customers. U.S. trade restrictions prohibit most high technology from crossing the Soviet border. In 1987 Glavkosmos lost a contract to launch a Hughes satellite when the U.S. company was not granted an export license. Hughes will not comment on the case. Negotiations with India recently broke down for similar reasons. Even though Glavkosmos has launched

Federal Communications Commission to construct its own constellation of 20 lightsats for two-way commercial alphanumeric communications.

Eventually those companies may be positioned to capitalize on renewed interest in microgravity. "Some day there'll be a lot of commercial activity in space," says Laurence J. Adams, retired president of Martin Marietta. "But we need to do a lot of experimentation with all the ideas that people have."

Europe seems to be trying to do just that. ESA is devoting more of its money to microgravity experiments than NASA is: in 1988 the European agency spent $98 million, or 2.6 percent of its budget, on microgravity. By comparison, NASA spent $63 million (.7 percent). Japanese, West German and Italian companies are also investigating the effects of microgravity on materials.

Since the advent of the space shuttle, the U.S. has offered limited opportunities for prolonged experimentation in microgravity. One reason: NASA mothballed its 20-year program in returnable reentry vehicles (space capsules) and only recently began backing this technology again. Western researchers have made some use of the available alternatives, namely, returnable Chinese and Soviet capsules. Payload Systems has also flown materials experiments on board the *Mir* space station. But trade restrictions still limit these options.

Small European and American companies and university researchers are also trying to revive returnable capsule technology. To foster longer-term experiments, ESA is building a free-floating platform called *Eureca* and an autonomous laboratory dubbed *Columbus* that will be serviced by the *Hermes* spaceplane every six months. The purpose: to "validate a large quantity of results in order to persuade industry to commit itself to operational uses," according to Euroconsult. (Two other *Columbus* modules are in the works as well: a unit for the space station *Freedom* and an orbiting polar platform.

In contrast, NASA is placing its bets on *Freedom*. Private industry schemes such as the *Leasecraft*, proposed by the Fairchild Space Company, or the *Industrial Space Facility*, promoted by Space Industries, have been stranded in a political and financial no-man's-land. Any proposed platform would need a well-heeled anchor tenant—most likely the government—to help recover costs and to convince other investors that the project is real. But because supporting a research platform could weaken NASA's argument favoring a space station, the agency has been skittish about funding platforms.

Space is no longer a match between the U.S. and the Soviet Union. Neither has it become a commercial free-for-all. That will change only when entrepreneurs and researchers demonstrate the commercial potential of space.

They cannot do so on their own, however. Governments must reduce the risks borne by space-bound entrepreneurs—and can do so by supporting research and building accessible infrastructure in space. It is a challenge that has some urgency; as the force of military necessity wanes, governments must prove their interest in commercial space goes beyond rhetoric. Those nations that succeed will see their team pull ahead. Those that do not will be left in the dust.

several payloads of experiments for Western companies, experts say that none but the technologically simplest devices would pass muster with current trade laws.

Dunayev, leaning back in his chair with his hands clasped across his belly and his thin mouth drawn in a perpetual wry grin, blames the embargo on protectionism rather than on national security concerns. "It is clear," he says, "that U.S. manufacturers of launch vehicles do not want to see the Soviet Union in the market." But Dunayev is optimistic that the trade restrictions will not tether the Soviets' vehicles for long. The U.S. is under pressure from its allies to ease its restrictions, he points out. And Glavkosmos may circumvent the trade imbroglio by launching its midsize Zenit rockets from Australia or Brazil. Both those nations are trying to develop commercial "spaceports."

Ironically, the political upheaval under way in the Soviet Union may spell more problems for Glavkosmos at home. With the economy faltering, the Soviet Parliament has fiercely criticized the approximately 6.9 billion rubles ($12 billion at official exchange rates) spent on the civilian and military space programs. Military cutbacks this year of 10 percent could trim funds for the space program. (In fact, the number of launches dropped to 74 last year.) More draconian cuts could increase Glavkosmos's vulnerability.

To enable government agencies to reduce their expenses, President Mikhail S. Gorbachev freed the organizations on January 1 to pursue contracts with foreign or Soviet groups outside the framework of central planning. That freedom is spurring the agencies to demand better service from one another. For instance, although the Institute for Space Research (IKI) already uses Glavkosmos to arrange space flights for its scientific missions, IKI has sometimes been dissatisfied with the results, says Vyacheslav Balibanov, deputy director of the institute. Now, he promises, "we would demand better service from Glavkosmos."

The institute could turn to another launching service if Glavkosmos does not perform up to standard. Although there is no official Soviet alternative, rumors have circulated at international space-industry conferences that the factories, or design bureaus, that manufacture rockets are keen to sell the boosters as well.

In spite of the capitalistic bravado, old habits die hard. Glavkosmos and other Soviet agencies are, for instance, often reluctant to disclose details about their products or operations. Consider Star City, the Soviets' training center for cosmonauts. General Yuri Nikolaevich Glazkov, a former cosmonaut and now a deputy director at Star City, gives an enthusiastic plug for the center's facilities. "We've got Americans, French, Canadians, and they all want to train their cosmonauts here," he boasts.

Q: How much does it cost to train a cosmonaut?
A: That's a secret.
Q: How can it be a secret if you're trying to sell your services?
A: Oh. More than $10 million. [Then he adds:] My secrets are leaking out!
Q: What else are you doing to make money?
A: We've begun to let tourists onto the premises [about 18 months ago], and we charge them.

Asked about the goals of Glavkosmos's commercial activities, Dunayev, too, lapses into bureaucrat-speak. "Glavkosmos is responsible for the whole civil space program," he says. "We coordinate its many different facets and also try and channel in some Western currency." Anders Hansson, a senior space specialist at Commercial Space Systems in London, puts it more succinctly. "Glavkosmos does not need to make a profit. It only needs to bring in enough Western currency to justify its existence. And that," he adds, "is more a question of politics than economics."—*Frederick Guterl, Moscow*

With reporting by Hervé This (Paris), Fred Guterl (Moscow), Tom Koppel (Tokyo), Wang Shide (Sichuan) and contributions from the editors of *Pour la Science* (Paris), *Spektrum der Wissenschaft* (Heidelberg), *V Mire Nauki* (Moscow), *Saiensu* (Tokyo) and *Le Scienze* (Milan).

SCIENCE AND BUSINESS

Controlled cruising, distant diagnosis, synthetic speech, East bloc economics

Driving While Automated

Planning smart highways for tomorrow's smart cars

Last month 25 lucky commuters who travel the Santa Monica Freeway in Los Angeles began receiving warning of traffic delays ahead via a screen mounted on the dashboard and a computer-synthesized voice. Thus alerted, they can exit to city streets and bypass a snarl, presumably arriving at the office ahead of their less fortunate colleagues.

This small project, costing less than $1 million, is the first U.S. test of IVHS, an inelegant acronym for intelligent vehicle/highway systems. IVHS is a battery of electronic technologies that allow car and road to communicate. If its proponents are right, IVHS could unclog highways at least temporarily, save lives and gasoline, and cut pollution. Smart roads and cars could save the U.S. tens of billions of dollars annually, advocates concluded at a recent meeting in Dallas, Tex.

The Los Angeles experiment, called Pathfinder, is crude compared with some navigation projects now under way in Europe and Japan. Toyota, Nissan and Mazda are already selling cars with navigation systems in Japan at premiums of up to $4,000. Combined sales are estimated to be about 1,000 units a month. In the U.S. such features will probably not be offered until the 1994 model year. Yet the U.S. is likely to step up the pace soon.

ON THE ROAD TO SMART HIGHWAYS

Road/Communications Systems, Tokyo

Data on road conditions and navigation is transmitted to dashboard map displays by 91 roadside beacons covering a 350 kilometer area. Project, which began in 1987, was set up by the Ministry of Construction with 25 Japanese companies.

Advanced Mobile Traffic Information and Communication, Tokyo

Similar to the above except that it uses central transmitters to communicate with vehicles. Sponsored by Japan's National Police Agency, the Ministry of Posts and Information and 59 companies. Plans call for expanding the project to 74 cities.

Leitung und Information System, West Berlin

System started in June, 1989, with 700 cars transmitting and receiving data from 240 beacons. The $11-million cost is shared by government, Robert Bosch and Siemens. Drivers receive data on a dashboard display and by synthesized voice.

Pathfinder, Los Angeles

Begun in June, 1990, the project involves 25 specially equipped Oldsmobiles that ply 12 miles of the Santa Monica Freeway. A map display highlights congestion ahead. Text displays and digitized voice offer alternative routes.

Autoguide, London

Beginning in 1991, the project will exploit technology similar to that used in Berlin and involve as many as 1,000 vehicles. The $16-million test is funded by a consortium headed by General Electric Co., which aims to offer a commercial service.

TravTek, Orlando

The $8-million test will begin in 1992 with 100 cars equipped with display terminals and navigation systems. GM, the American Automobile Association, the Federal Highway Administration and state agencies are funding the project.

Driving the growing interest in IVHS is the stark fact that throwing concrete at traffic problems just is not possible anymore. Although travel in the U.S. has been growing at 5 percent a year, at the same time, highway budgets are being trimmed and vociferous residents frequently delay projects indefinitely. "In some places we have built our last mile of road," says G. Sadler Bridges, associate director of the Texas Transportation Institute, a university highway-research consortium.

So transportation planners and automakers are eager to throw silicon at traffic congestion instead. Steady advances in semiconductor technology have made a remarkable array of driving aids not just possible but nearly affordable. Among the innovations:

- Route guidance systems, which would direct drivers to their destination around the worst traffic jams.
- Vision enhancement, which pierces fog with infrared sensors and shows the road ahead on a so-called heads-up windshield display.
- Adaptive cruise control, which automatically cuts back on the gas when a car is closing too fast on the vehicle ahead. That's just a short step from collision avoidance—automatic braking and swerving.

Still lacking, however, is a widely agreed on scenario for implementing IVHS technology. At a seemingly endless series of meetings in the past several years, researchers from government, universities and private industry have attempted, with little success, to devise a plan for the orderly introduction of IVHS. The imponderables include the level of federal government support for an IVHS infrastructure and adoption of standards. Agreement on such steps, IVHS boosters note ruefully, is easier in Europe and Japan, where government and industry work closely in centralized planning efforts.

Yet some IVHS technology is already on U.S. roads. Thousands of long-haul trucks already have voice and data links to dispatchers. Navigation and route guidance are next. "Truckers can measure the savings, and they can be substantial," says David K. Willis, senior vice president of the American Trucking Associations Foundation. Among the benefits is reduction—by as much as 20 percent—in unnecessary mileage. Willis also says there is a ready market for devices that can help truckers avoid congestion.

Other potential buyers for navigation systems include taxis, emergency vehicles and rental cars. Budget Rent a Car seems the most enthusiastic. The company is planning a major test of navigation-equipped cars, possibly next year. Budget is evaluating navigation systems made by Diesel Kiki, a Japanese parts maker, and by Ford Motor Company, which owns virtually all of Beech Holdings, Budget's parent.

Although such navigation systems will show drivers the shortest route home, they will not be able to help avoid traffic jams until linked to real-time communications. Traffic engineers favor the heavy use of sensors embedded in the road to gauge traffic flow. The information would be collected in a central computer that would then relay instructions back to vehicles via radio frequency or infrared signals. That scheme would be an extension of existing systems that use electronic signs to warn drivers of problems ahead.

A less costly proposal would use the cars themselves to collect the traffic data. The cars would relay signals to the central computer, allowing it to calculate the time required to traverse stretches of road. This might dramatically reduce the need for sensors and for computing power. "We shouldn't commit to heavy infrastructure spending until we've done enough testing," says Randy M. Doi, director of IVHS at Motorola.

No one has dared put a number on the eventual size of the IVHS marketplace, but it clearly could be a bonanza for automakers and suppliers of both equipment and services. Automakers in the U.S. could sell a package that would include navigation and other advanced features for as much as $1,000 a car. That's a price General Motors says customers quizzed in clinics would be willing to pay.

GM is by far the IVHS leader among U.S. automakers, spending an estimated $10 million a year. Some 65 GM scientists and engineers are working on aspects of the technology. The company, which has licensed a route-planning system from Etak, is providing the 25 cars for Pathfinder. It is also a leading force in TravTek, a recently announced $8-million experiment in Orlando, Fla.

IVHS proponents hope the technology will get a boost when Congress begins to work in earnest on the highway reauthorization bill, scheduled for next year. The IVHS "lobby," a loosely knit confederation that may constitute itself into a new trade group, has already picked up support in Congress, some of it unexpected.

Senator Daniel Patrick Moynihan of New York, a strong advocate of high-speed magnetic levitation rail systems, recently introduced a bill that would appropriate $10 million for the Federal Highway Administration to come up with a comprehensive IVHS plan and an additional $10 million to fund a demonstration project. The administration, particularly Department of Transportation Secretary Samuel K. Skinner, seems solidly behind IVHS.

IVHS will face plenty of skeptics in Congress. The mass-transit and concrete lobbies are not on board. Rural legislators may balk at a program that will benefit mainly urban areas. In addition, there are doubts about whether IVHS will, in fact, clear traffic jams or make the highways safer. The National Highway Traffic Safety Administration already is concerned that truck drivers could be distracted by the welter of information provided by instrumentation. And as drivers hand over more and more control of their vehicles to computers, the question of who is liable in an accident looms large. *—Michael G. Sheldrick*

Do You See What I See?

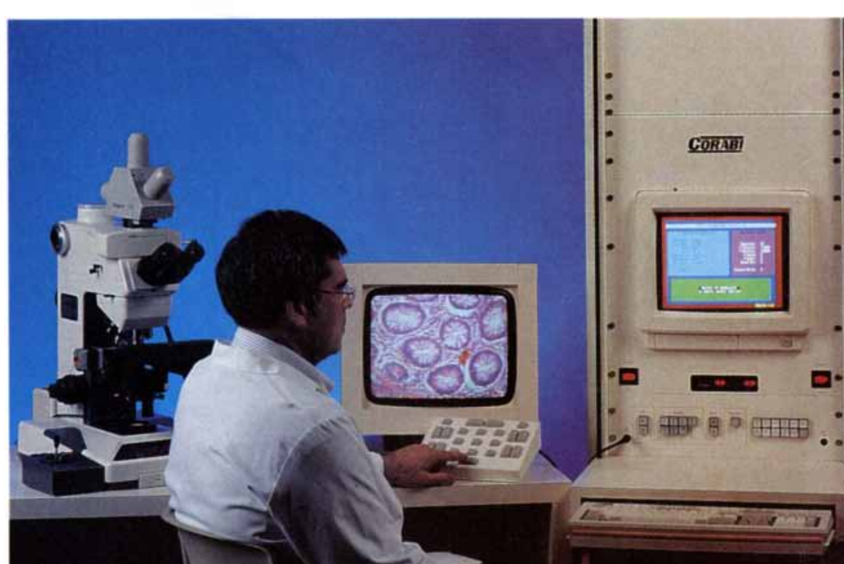
Pathologists lead the way for long-distance diagnosis

Final confirmation of many diseases often comes from the pathologist. This type of scientist laboriously scans microscope slides of tissue samples, looking for cells with aberrant color, shape or arrangement. Pathologists frequently seek a second opinion before making the critical decision; this consulting can be the most difficult part of the process. It can take precious days to send slides or tissue samples to other experts.

With a new approach, however, that may change. Now pathologists can analyze tissue samples hundreds of miles away with a "telepathology" system that combines a robotically controlled microscope, custom software and high-definition television (HDTV) cameras and monitors. The first such set-up in the U.S. was installed this past March by Corabi International Telemetrics, located in Alexandria, Va. It connects the pathology departments of the Emory University Hospital and Grady Memorial Hospital in Atlanta, Ga. The interactive network lets pathologists with individual areas of expertise make initial diagnoses or render second opinions without shuttling specimens or doctors in between.

When a pathologist puts a slide on the microscope stage, the image is transmitted automatically. The recipient uses a keypad to control the focus, position and illumination of the enlarged image on a high-resolution screen. The real-time images, which are sent by broadband microwave in the Atlanta installation, can be manipulated as if the receiving pathologist were sitting directly over the transmitting microscope. Images for a second planned telepathology installation, for a network of 10 hospitals in the Washington, D.C., area, will travel through fiber-optic cables.

If telepathology catches on, remote areas around the world could have ex-



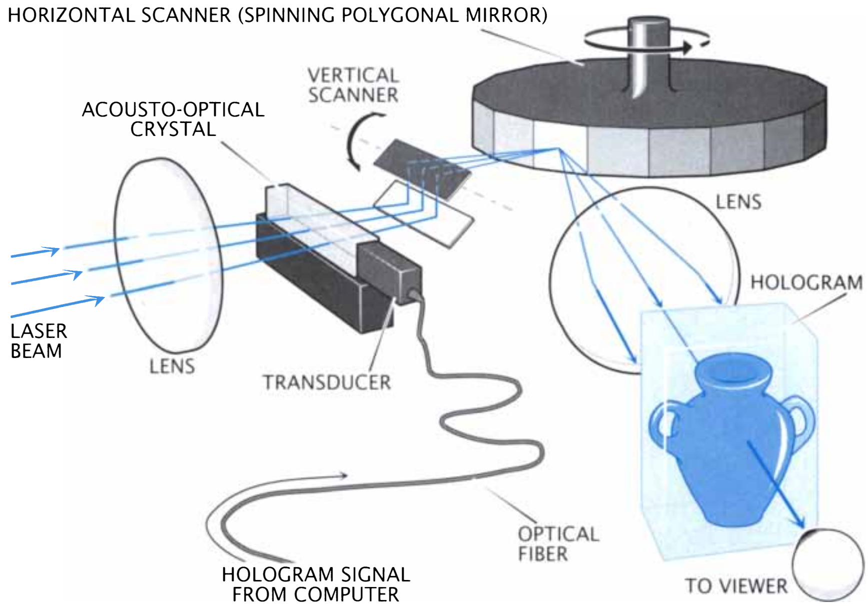
TELEPATHOLOGY SYSTEM sends high-resolution images of tissue samples from a remote-controlled microscope, enabling long-distance diagnosis.

pert diagnoses in an instant. That is the hope of the system's developer, pathologist Ronald S. Weinstein. He founded Corabi in 1985 with his sister, Beth W. Newburger, "to do good as well as do well." The team is not alone in thinking about using telecommunications to improve health care, although it is the first to commercialize the concept of "telemedicine" for a field aside from radiology. Putting together the technology was no simple matter. "We knew what we were looking for. The trouble was getting someone to build the equipment," Newburger recalls.

Corabi first approached four large and then half a dozen midsize U.S. companies, all working on some kind of video project. Weinstein and Newburger explained pathologists' persnickety technological requirements: absolutely accurate color, high resolution and the ability to look at different areas of the image under varying magnification and illumination—all live. They got a less than enthusiastic response. "Everybody told us that if we could get anybody to build it, it would be a minimum of five, maybe eight, years before anything was delivered," Newburger says.

The pair paid a visit to Japan in 1986. But without introductions, they failed to win any interest. Then COMSAT, the U.S. satellite communications company, gave Corabi $1.5 million in funding. COMSAT soon introduced its new partner to Sumitomo Electric, with which it operates a joint office in Japan. Corabi returned to Japan in 1987 with a dance card arranged by Sumitomo; 32 companies stepped forward to woo it. The system was built and tested in less than two years. Nikon agreed to build a computer-controlled microscope for Corabi at their first meeting. The HDTV system was built to order by Ikegami Tushinki. Unlike the Sony Corporation's broadcast-quality monitor, Corabi's product does not compress data to give an unnaturally sharp appearance. Pathologists want the ragged edges of cells to show.

With its first project out the door, Corabi may soon find itself getting the squeeze from the likes of Du Pont, General Electric and Kodak, all of which already sell remote systems for reading X rays and other diagnostic data. Faced with giant competitors, Corabi will have to rely on its head start. And so far the prognosis is good. Robert R. Pascal, Emory's director of anatomical pathology, says, "It had to have a pathologist's input to make it so easy to use." —*Deborah Erickson*



MOVING HOLOGRAMS are created by computer at M.I.T.'s Media Lab. A laser beam and data converted into sound waves pass through a crystal. The light that emerges is organized by mirrors into a rapidly changing 3-D image.

Move Over, Mr. Spock

Sound-sensitive crystals bring computerized holograms to life

In *Star Wars* and other science-fiction films, holograms are astonishingly lifelike, moving images that appear on computer screens and futuristic videophones. In reality, most holograms are blurry, still pictures that look only a little bit three-dimensional and appear on credit cards and costume jewelry. Yet investigators at the Media Laboratory of the Massachusetts Institute of Technology are catching up with Hollywood: they have produced computer-generated moving holograms.

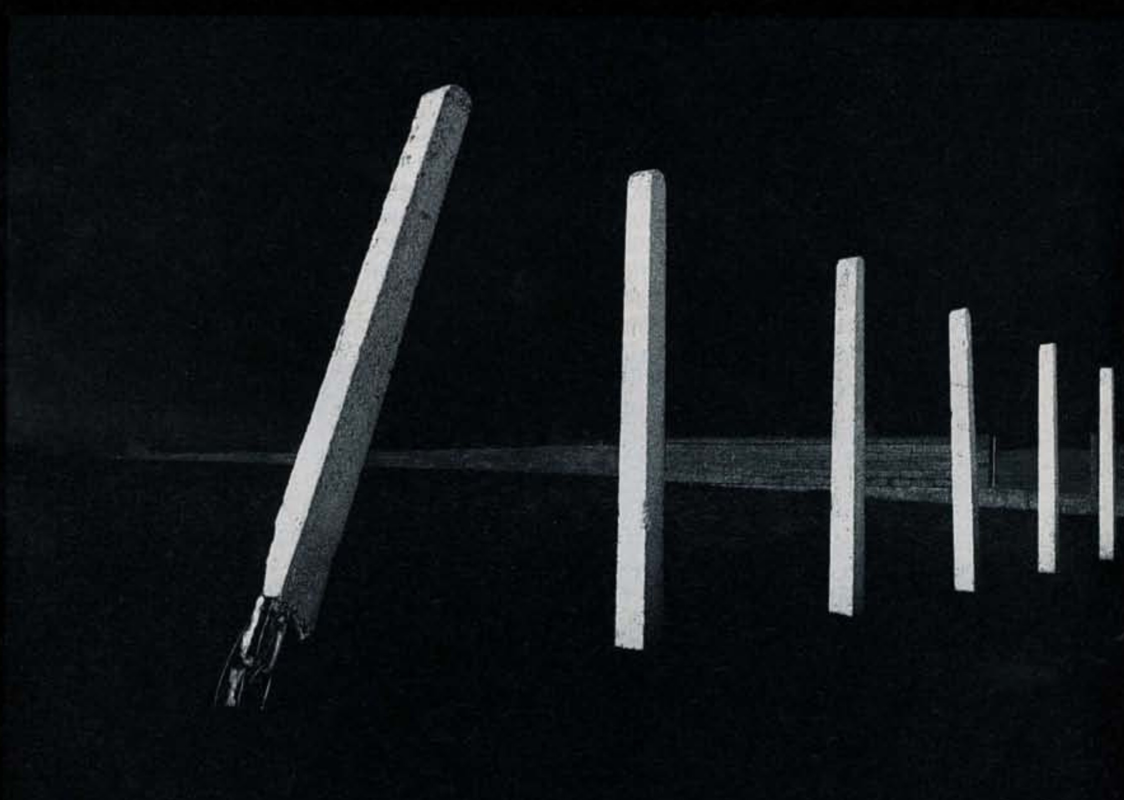
A demonstration sequence prepared by the researchers shows a tiny starship *Enterprise* from the television show *Star Trek* orbiting a planet. More advanced systems could one day serve as 3-D displays on workstations and other computers. Air-traffic controllers, for example, might be able to see the true relative positions of airplanes in crowded flight patterns. Physicians may base their diagnoses on three-dimensional images of organs scanned by computed tomography instead of relying on flat cross-sectional views.

The holograms that Stephen A. Benton, Pierre St. Hilaire, Mark Lucente and their colleagues have produced so far are small—about four centimeters on a side—and seem to hover in front of the 35-millimeter camera lens that focuses the image. The images exhibit parallax: as a viewer's head moves to the left or right, different parts of an image become visible. The moving hologram is animated at a speed of 15 to 20 frames per second, which matches the animation quality of some cartoons shown on television today.

But turning computer data into a holographic display is tricky. Holograms are usually made photographically by recording on film the interactions between a laser beam and its reflection from an object. Although a computer can calculate how a hologram should appear from different angles, rendering that information as a visible image has been difficult. Benton and his colleagues have accomplished this task by first converting computer data into sound and then into light. The key to their success is an acousto-optic crystal, which responds to sounds by bending light in different directions.

In the M.I.T. system, a crystal is simultaneously exposed to a laser beam and to sound waves that convey a computer's description of a hologram. The crystal refracts the laser beam in such a way that, at each instant, the beam takes on the appearance of a different part of the hologram. Mirrors then organize the beam into a still picture. By changing the picture slightly 20 times per second, Benton's group creates the illusion of a moving image.

Generating a hologram in this way requires vast number-crunching power. To keep the prototype system feasi-



Michael Kenna, *Tilted Poles*

"The field cannot well be seen from within the field."

Ralph Waldo Emerson

To know what is possible tomorrow you must be willing to step outside of what is possible today.

The long view is all.

At Toshiba, we've been taking that view to help create a tangible future for over one hundred years.

Sustaining it now with our commitment to manufacturing in the U.S. And with a wide diversity of products of the highest quality.

All of which have helped make us one of the leading electrical and electronic manufacturers in the world.

A nice perspective to have. When you work in as many fields as we do.

In Touch with Tomorrow

TOSHIBA

Color TVs & VCRs

Computers & Printers

Copiers

CT Scanners

Digital Instrumentation

Disk Drives

Facsimile Systems

Home Appliances

Industrial Motors & Controls

Liquid Crystal Displays

Medical Tubes

MR Imaging

Nuclear Medicine

OPTO Electronic Devices

Picture Tubes

Portable Audio

Professional Video Systems

Room Air Conditioners

Semiconductors

Telecommunication Systems

Turbines & Generators

Ultrasound

Uninterruptible Power Supply

Variable Speed Drives

X-Ray Systems

ble, the M.I.T. group keeps its holograms small, narrowing the viewing angle and eliminating vertical parallax. Still, two million bytes are needed to create a single holographic frame. Most computers would take hours or days to make the necessary calculations. Even the M.I.T. group's $4-million supercomputer needs about five seconds.

At this rate the system is about 100 times too slow to create new frames fast enough for real-time moving displays. To show the *Enterprise* moving, the investigators must make all the calculations in advance and store the results. Because the computer can store only about 50 frames' worth of information, the moving hologram sequence can last only a few seconds.

Within 10 years, Lucente guesses, technology may have advanced sufficiently for holographic displays to be available on some business machines. There are $1-million parallel-processing computers that could do the job now, he points out. But, laughing, Lucente adds a disclaimer that could serve as the Media Lab's unofficial slogan: "We're only inventing the future, not predicting it." —*John Rennie*

Now Hear This

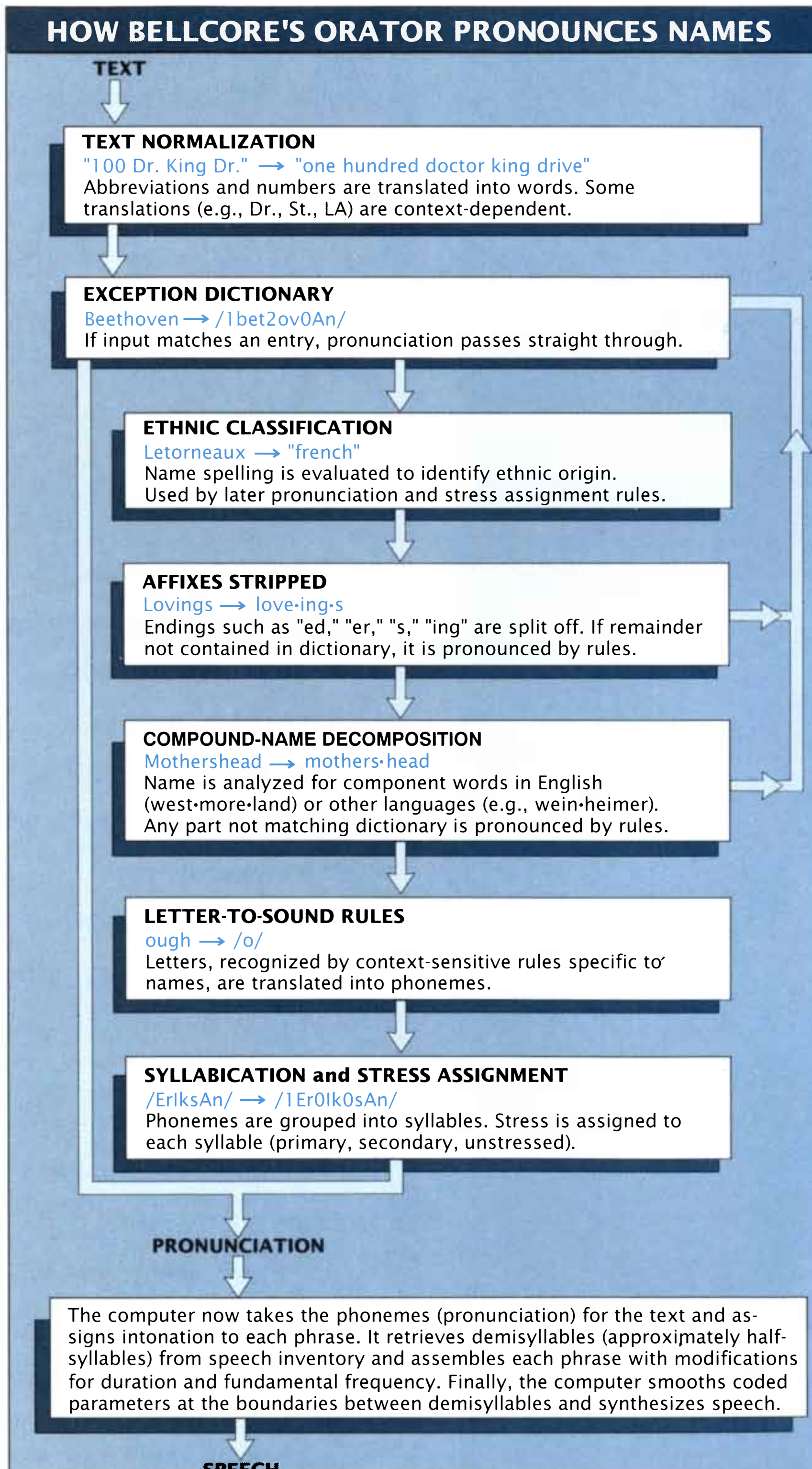
Bellcore's speech synthesizer says something out of nothing

When the voice on the telephone sounds like a Midwestern male robot, ORATOR is on the line. This new speech synthesizer, developed by Bell Communications Research (Bellcore), pronounces words with a certain buzzing, staccato ring. The sound is even less pleasing than the perfectly bland digitized female voices that intone, "The new number is...." But ORATOR does have a special talent that current operator services lack—it can pronounce speech from computerized text. And ORATOR's odd voice is nonetheless drawing praise from potential users.

Until now, digitized voice has been the only speech technology to automate operator services. Such devices take snippets of recorded speech and rearrange them as needed. The technology works fine for numbers and a few pat phrases. But it cannot provide information that has not been recorded previously by a human.

ORATOR can. It is the first speech synthesizer for commercial use to break down words into demisyllables. Older synthesizers tried to glue together smaller portions of words—for instance, discrete sounds like "e" or "a" or "ch." But "ORATOR looks at larger units, so it can incorporate the interacting effects that neighboring sounds have on each other," explains Murray Spiegel, who developed the technology with Marian Macchi.

On the way to being pronounced by the computer, each word is first checked against a dictionary and, if not found there, whisked through a series of analyses. Ethnic classification is just one measure that makes ORATOR particularly adept at pronouncing names. It says the "J" in "Jean-Claude" with a soft French "zh" and the American "Jean" as in blue jeans. "But the real guts of the system are the letter-



HOW BELLCORE'S ORATOR PRONOUNCES NAMES

to-sound rules," Spiegel notes. These are complete enough to let ORATOR pronounce most names accurately but not to read random text.

After ORATOR retrieves the demisyllables from an extensive inventory of computerized speech, it then puts them back together and smooths the boundaries between sounds. But not too much. The staccato rhythm enhances clarity, Spiegel explains.

Telephone companies hope to have ORATOR on the job in the near future. When customers call up to question a bill, the computer—instead of a human operator—could tell them the name and address that match a given number. It could also dispatch repair crews. Workers could simply call in to a central number and with proper access codes receive instructions to "Go to Joe's house on St. John's Place in Brooklyn." (ORATOR knows to pronounce "St." before a word as "saint," not "street.")

Eventually ORATOR's developers predict that the system will provide a range of new services to telephone customers. It might allow businesses to personalize their information listings. For example: "Henry's Hardware has moved, and you can reach it at 345-9276," one customer's message might say. "The phone companies are not now allowed to offer Henry the option of saying 'Our hours are this, and our special today is nails,'" says Jeanne Canteen, general manager of operator services at New York Telephone. "But we will be looking at that."

In a similar fashion, medical insurers might provide customers with a number to hear a list of approved doctors in a region. A voice synthesizer could also personalize messages given by voice mail—or even automated-teller machines—by reacting to user codes. Spiegel says: "Everyone is complaining about society becoming numerized. This way, if it pronounces your name correctly, you're able to be treated as a human again."

In some applications, however, ORATOR raises questions. "You can't deploy technology just because it's there. We have to find out the impact on the employee, the user, the regulatory climate and the cost," Canteen observes. Yet Spiegel is convinced that more advanced versions of ORATOR will one day see widespread use. "With each incremental improvement to the technology, more applications will be found," he declares. Thoughtful companies will consider the implications of this technology before it becomes pervasive. *—D.E.*

THE ANALYTICAL ECONOMIST

On the eve of the East bloc's morning after

In Berlin, the border between Eastern and Western Europe brings to mind nothing quite so strongly as a tutorial in basic economics. After so many years of controlled markets, the laws of supply and demand are now reigning sovereign.

Buses from Poland stretch for a kilometer along the West Berlin avenue that leads to the Brandenburg Gate and to the East bloc. Young Polish men stream from the Kurfürstendamm, Berlin's main shopping district, and crowd onto the buses, carrying boxes labeled "Aiwa" and "Casio." Older men and women push shopping carts laden with canned apricots, orange juice or beer. This is inelastic demand in action: almost without regard for price, Poles are buying Western goods. Many hope to sell the products back home at even higher prices.

The inelastic supply of ostmarks, meanwhile, is advertised on every street corner near the center of town. At almost any price, East Germans are trying to sell their currency. To them, one deutsche mark may be worth three ostmarks—maybe even four or six. Fueling their eagerness to sell is the knowledge that ostmarks will soon have little value. As of July 2, banks will swap deutsche marks for ostmarks—but the one-for-one exchange is guaranteed only for the first 4,000 ostmarks an adult converts. (Senior citizens can trade in 6,000 ostmarks at the favorable rate.)

The lust for Western goods has a flip side: few Easterners buy goods from the East. In some cases, the superior quality of Western goods clearly warrants the higher prices. Eastern Europeans are no longer interested in the noxious little two-cycle Trabant cars. In other cases, the faltering demand is less logical. The director of an East German agricultural commune near Berlin complains there is no point in sending pigs or cattle to the slaughterhouse. No one will spend their ostmarks to buy the meat. The commune's dilemma may be temporary, however. This year export restrictions between East and West will fall, allowing the commune to sell its animals to West German slaughterhouses.

East Germany is also struggling with the economic lesson of marginal return on investment. Will a deutsche mark spent renovating an old factory there return as much to investors as money spent expanding a plant in West Germany? Probably not. Throughout the East bloc, factories are crumbling, often saddled with ancient equipment and dumps of hazardous wastes. Even the air in the brown-coal district near Leipzig smells of burnt flint and sulfur. East German industry may be inflicting as much as 30 billion deutsche marks of damage on the environment every year, according to a West Berlin research institute. It may cost the country about 100 billion deutsche marks just to clean up fouled groundwater. That figure could double if the government tries to bring the energy-producing plants up to the environmental standards of the West.

Finding funds to rebuild the industrial infrastructure will be difficult. The East German government estimates private national savings to be 167 billion ostmarks—roughly 10,000 ostmarks per person. Although some of that will be converted to deutsche marks at a favorable rate, the rest may be exchanged for far fewer deutsche marks or spent on Western goods, thus depleting the country's pool of savings.

There is little to draw Western money now. Not only are the dilapidated factories riddled with environmental liabilities, many employ unnecessarily large work forces. When ostmarks are changed to deutsche marks, East German salaries will be converted one-for-one into deutsche marks. Food and shelter, which have been heavily subsidized, are likely to rise about threefold in price as the government withdraws its support from the now free market. Western producers may consequently decide to serve the East from their established plants in West Germany.

Some Westerners nonetheless speak with hope of "market forces" that will eventually force salaries and prices in the East into a proper equilibrium. Other Westerners grumble that they will have to pay—through higher taxes and inflation—to support their poor cousins. Along with the flood of new goods, Easterners are facing the possible demise of their extensive social welfare system and the likelihood that unemployment will befall four million of them in the wake of unification.

Centrally planned economies have clearly succumbed in Eastern Europe. But it may be some time before the abrupt change to capitalism in East Germany begins to improve the lives of people there. *—Paul Wallich, Berlin*

When their systems aren't integrated,

some businesses

waste time and effort

running from

department

to department

to department

until finally

Pyrotechnics

The secrets behind dazzling fireworks displays are yielding to scientific snooping. Similar principles are at work in devices ranging from the space shuttle to safety matches

by John A. Conkling

A distant, muffled "whoomp" resounds, and a trail of yellow-orange sparks tumbles into the night sky, culminating in a circular burst of brilliant blue and green streaks. Another explosion shoots out a ragged arc of red streamers, followed by a shower of white and gold sparks. A third firework produces a staccato barrage of bright flashes of white light and thunderous noise.

These effects have been a familiar part of major celebrations for centuries. For most of that time, the design and composition of fireworks was a craft, not a science. Only in recent decades have researchers begun to unravel the physical processes that underlie the production of dramatic colors and special effects. As a result of these investigations, a true discipline of pyrotechnics—the "science of fire"—has emerged. Pyrotechnics embraces not only fireworks but a whole range of devices that use similar materials, including hazard flares, safety matches and even the solid-fuel rocket boosters of the space shuttle.

One of the oldest pyrotechnic compositions, black powder, serves as both the propellant and explosive charge in modern firework shells. The Chinese developed black powder (the original gunpowder) more than 1,000 years ago for use in crude missiles and firecrackers. Awareness of black powder traveled west during the Middle Ages. The English monk Roger Bacon disclosed a formula for the explosive mixture in 1242 as part of his defense against accusations of witchcraft. He considered it such a dangerous material that he wrote about it in code. As the formula became more widely known, black powder revolutionized quarrying and construction. Weapons such as muskets and cannons, developed during the 14th century, exploited black powder as a propellant.

The basic formula for black powder has persisted essentially unchanged throughout the centuries: an intimate blend of potassium nitrate (commonly known as saltpeter), charcoal and sulfur in a 75:15:10 ratio by weight. It may in fact be the only chemical product that is produced today using the same ingredients, the same proportions and the same manufacturing process as were used in the time of Columbus. This constancy reflects the fact that black powder is a nearly ideal pyrotechnic substance. It consists of abundant, inexpensive chemicals that are relatively nontoxic and environmentally safe. The mixture is so stable that it can be stored for decades without deteriorating, if kept dry. Black powder is easily ignited by means of a moderate jolt of energy, such as a spark or a small burning fuse.

Historically, only a handful of families have dominated the fireworks industry in the West. Details such as chemical recipes and mixing procedures were cloaked in secrecy and passed down from one generation to the next. Families remain an important force in the industry. In the U.S., for instance, there are the Gruccis of Bellport, N.Y., the Zambellis of New Castle, Pa., the Rozzis of Loveland, Ohio, and the Souzas of Rialto, Calif. One effect of familial secretiveness is that, until recent decades, basic pyrotechnic research was rarely performed, and even when it was, the results were not generally reported in scientific journals.

In principle, the pyrotechnic process is not unlike normal combustion. A pyrotechnic composition contains an oxygen source (oxidizer) and a reducing agent (fuel). They are usually separate, solid chemicals that must be physically mixed together. When heat is applied, an electron-transfer, or oxidation-reduction (redox), reaction takes place.

Atoms in the fuel lose electrons to atoms in the oxidizer. In the process

JOHN A. CONKLING is executive director of the American Pyrotechnics Association and adjunct professor of chemistry at Washington College in Chestertown, Md. Every summer he teaches several one-week seminars on pyrotechnics at the college; these are the only pyrotechnics courses offered in the U.S. Conkling received a Ph.D. in chemistry from Johns Hopkins University in 1969. He has written extensively on the chemistry of pyrotechnics and on fireworks safety, and he serves on both the explosives and pyrotechnics technical committees of the National Fire Protection Association. In his more peaceful moments, Conkling enjoys fishing.

the fuel atoms bond to the oxygen atoms that are liberated from the oxidizer and form stable reaction products. The new chemical bonds are more stable, and so energy is released in the form of heat; the same process occurs in combustion. In combustion, however, the oxygen comes from the air. In a pyrotechnic mixture the oxygen is self-contained, and the heat is much more closely confined.

As long as a pyrotechnic mixture remains cool and dry, it is generally quite stable. A solid mixture experiences only a very slow surface reaction controlled by diffusion. When the composition is ignited, it begins to liquefy and vaporize in the resulting pyrotechnic flame, and the fuel and oxidizer closely intermingle. This proximity leads to faster chemical reactions and, in turn, still more rapid energy release.

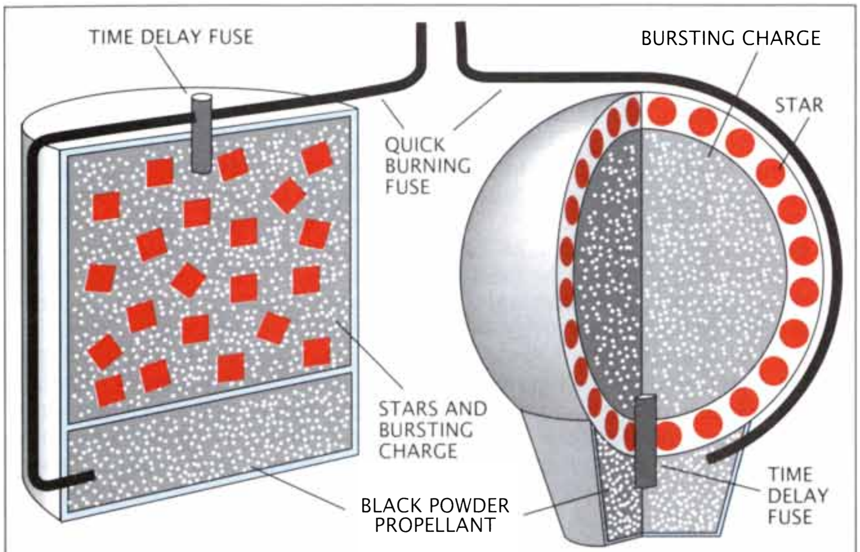
Pyrotechnics makes use of a variety of fuels. Many mixtures incorporate organic (carbon-containing) materials, such as charcoal (used in fireworks and gunpowder) or sugar (in smoke grenades). Other common fuels include nonmetallic elements, such as sulfur, silicon and boron. Silicon and boron release a large amount of energy when oxidized, and they do not produce gas in the process. They are used in delay fuses to ignite other compositions at a desired time. Chemically active metal fuels—most often aluminum, magnesium and titanium—burn at high temperatures and emit bright light. They came into use in the 19th century and dramatically improved the brilliance of pyrotechnic explosions.

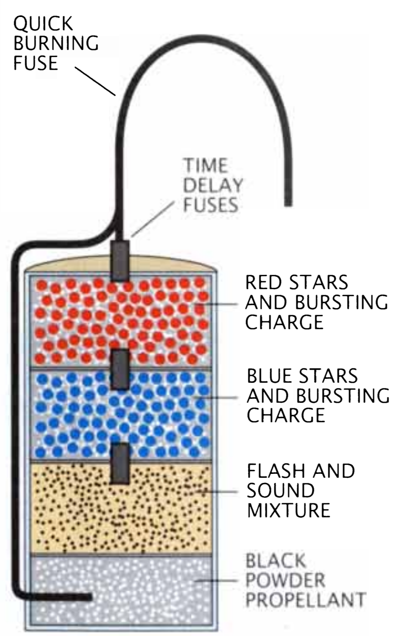
The spectacular splashes of light produced by fireworks are the most famous pyrotechnic phenomenon. The color of the light depends on its wavelength. Visible light consists of electromagnetic radiation having wavelengths between 380 and 780 nanometers (a nanometer is one billionth of a meter). The longest visible rays appear red, the shortest, violet. A glowing object appears white if it radiates throughout the visible spectrum. If most of the light is emitted in a narrow portion of the spectrum, it takes on the color of that portion.

Pyrotechnic compositions emit light by three basic processes: incandescence (blackbody radiation), atomic emission and molecular emission. Incandescence occurs when solid or liquid particles in the pyrotechnic flame are heated to a high temperature. The hot particles emit a broad spectrum of radiation as they attempt to shed their excess energy. The higher the temperature, the shorter the wavelength at which the most radiation is emitted. The intensity of the emission is proportional to the fourth power of the flame temperature, so a moderate increase in temperature drastically brightens the flame.

White-light flares contain a reactive metal, such as magnesium, as a fuel. Solid metal oxide particles, created when the fuel is oxidized, are heated to more than 3,000 degrees Celsius; at these temperatures, their incandescent glow is white-hot. A mixture of potassium perchlorate and fine aluminum or magnesium powder produces a powerful explosion along with a burst of

SPLENDOR OF FIREWORKS bursts forth in this photograph taken of New York Harbor during the centennial celebration of the Statue of Liberty. Crude fireworks have existed for more than a millennium, but vivid color-emitting compounds were not developed until the 19th century. Firework compositions were often closely held family secrets; only in recent decades have researchers begun to decipher the chemical processes behind the flashes and booms.





DESIGN OF FIREWORK shells controls their visual effect. All contain black-powder propellant and a quick fuse that ignites the powder, expelling the shell from a launch tube. A time-delay fuse sets off a bursting charge after the shell is far above the ground. The charge activates pellets (stars) of light-emitting compounds. American-European–style shells contain a random mixture of stars and bursting charge (*top left*). They produce irregular splashes of color. In Japanese-style chrysanthemum shells, the stars surround the charge (*top right*), creating symmetric explosions. Multiple-burst shells have several compartments, each containing a different composition, that are joined by time-delay fuses (*left*). The loud bang that traditionally ends a firework is produced by a compartment containing aluminum-powder mixture instead of stars and black powder.

white light. Such "photoflash" or "flash and sound" compositions have a wide range of uses, from firecrackers to special effects for rock concerts to bursts of light for nighttime photography. These compositions produce the bright flash that traditionally terminates a firework explosion.

Larger metal particles retain their heat longer than powders and can continue to burn by drawing on the oxygen in the air. They create white sparks rather than an instantaneous flash. The bigger the particles, the longer the sparks last. Charcoal and iron particles do not become as hot as active metal particles—only about 1,500 degrees C—and so they produce dimmer, gold-colored sparks.

The brilliant colors seen in modern fireworks displays are generated either by atoms or by molecules present in a vapor form in the pyrotechnic flame. In the former case, the heat of the flame excites an electron in an atom and bumps it from its normal, ground-state orbital to a higher-energy one. The electron rapidly returns to its ground state and emits the excess energy as a photon (a single particle, or unit, of radiation) of a specific wavelength.

Sodium is one of the most potent atomic light emitters. Sodium atoms heated above 1,800 degrees C give off yellow-orange light having a wavelength of 589 nanometers. The process is so efficient that it tends to overwhelm any other atomic or molecular light sources in a pyrotechnic flame. Even small amounts of sodium-containing impurities can ruin efforts to produce a flame of any other color.

In other applications, sodium's prodigious light emission can be helpful. Sodium nitrate oxidizer combined with magnesium metal fuel is the principal composition used by the U.S. military to illuminate nighttime operations. The magnesium is oxidized by the sodium nitrate when the mixture is ignited; the resulting hot magnesium oxide particles shine with a white incandescent glow. High temperatures (3,600 degrees C) in the magnesium flame also broaden the range of wavelengths emitted by the sodium atoms. The result is an intense white light.

As with atomic emission, molecular emission involves a transition from a ground state to an excited one. The molecule must be in gaseous form in the pyrotechnic flame, and it must be heated to a temperature high enough to reach the excited state that causes it to radiate. If the flame is too hot, however, the molecule disintegrates into its constituent atoms and no light is emitted. Moreover, the molecules must be sufficiently concentrated in the flame to generate intensely colored light, but the production of solid or liquid particles must be kept to a minimum because they give off incandescent radiation that washes out the color.

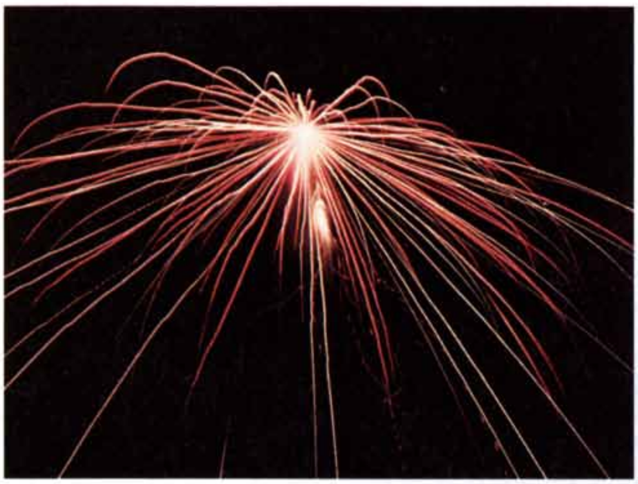
In the absence of theoretical understanding, colors were generated in fireworks by a trial-and-error process. Over the past several decades Bernard E. Douda and Henry A. Webster III of the Naval Weapons Support Center in Crane, Ind., and David R. Dillehay of the Longhorn Division of the Thiokol Corporation in Marshall, Tex., have conducted research that has helped identify the principal colored emitters in pyrotechnics. Takeo Shimizu of the Koa Fireworks Company in Japan also has contributed to this area.

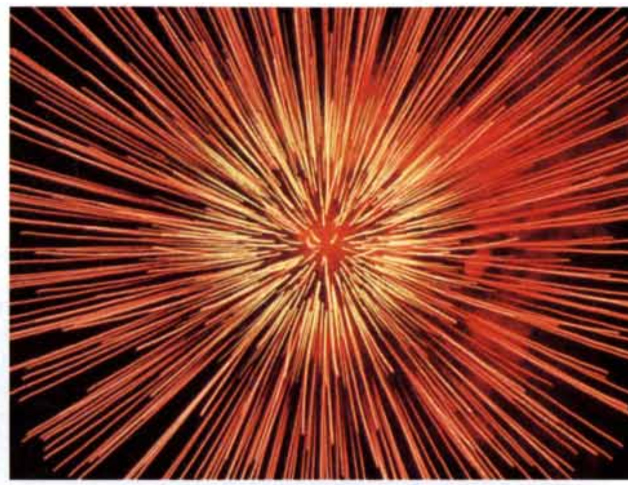
A few groups of molecules are responsible for nearly all the colors in fireworks. Compounds of the element strontium produce the reds: stronti-um hydroxide ($SrOH$) and strontium chloride ($SrCl$) emit red light at wavelengths between 605 and 682 nanometers. Molecules containing barium create the greens. Barium chloride ($BaCl$), for instance, emits green light at wavelengths between 507 and 532 nanometers.

These molecules are so fragile that they are unstable at room temperature; consequently, they cannot be packed directly into a firework. Instead they are synthesized in rapid reactions in the flame. Manufacturers add such compounds as chlorinated rubber, polyvinyl chloride (a chlorine-containing plastic) or perchlorate or chlorate oxidizers (containing a chlorine atom and four or three atoms of oxygen, respectively). These compounds decompose at high temperatures and release free chlorine. The chlorine atoms combine with barium or strontium, briefly creating the desired light-producing molecules.

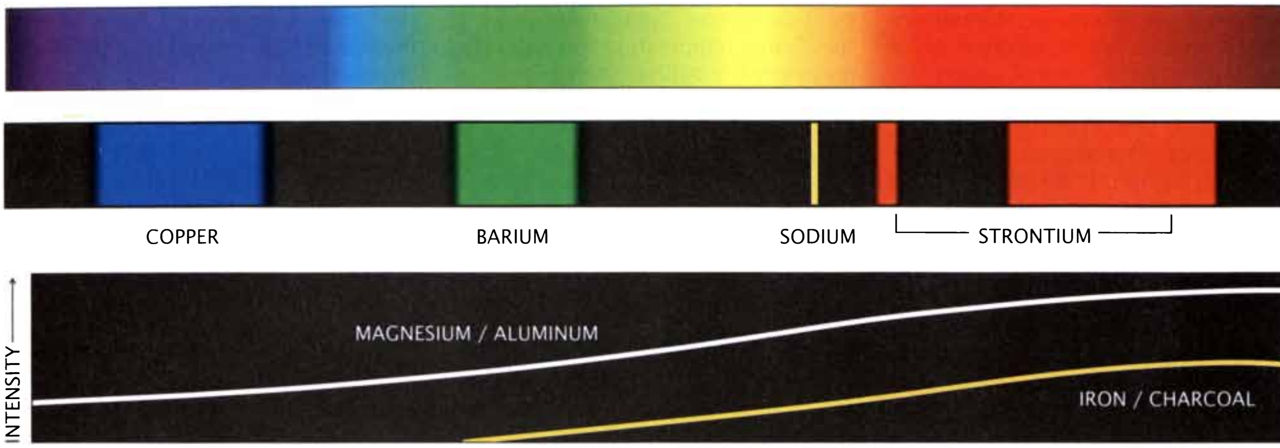
A rich blue flame is perhaps the ultimate challenge to the pyrotechnician. The best blue emitter yet identified—copper chloride (CuCl)—is unstable at the elevated temperatures needed to produce intense light in fireworks. If the flame temperature exceeds that necessary for optimal molecular emission, the molecules disintegrate rapidly. Distinctly blue fireworks therefore demand especially precise control over the relative proportions and particle size of the necessary chemicals. The same holds true for purple or violet colors, which are created by combined emission from strontium chloride and copper chloride formed in the flame. I pay close attention to flame colors when I view a fireworks display; if a decent blue color appears, I am always impressed and curious to know what chemical mixture was used.

Color-generating compounds combined with the appropriate fuels and oxidizers can produce special effects. Red sparklers derive their color from the combination of strontium carbonate (which emits red light) and aluminum granules (which provide the sparks). These ingredients are mixed with fuels, binders and an oxidizer to create a thick slurry; wires are dipped into the slurry and allowed to harden





SHAPES AND COLORS of firework bursts reflect the type of shell and the composition of the stars. A ragged spray of light is the signature of an American-European shell (*top left*). Round, flower-shaped explosions result from the aptly named chrysanthemum shell (*top right*). Multiple-burst shells or volleys of single-burst shells can produce dramatic effects by incorporating a variety of pyrotechnic compositions that emit different colors and sounds when ignited (*bottom*).



FIREWORKS GLOW by incandescent, atomic, and molecular emission. The color depends on where the emission falls on the visible spectrum (*top*). Atoms and molecules become excited at the high temperatures in a pyrotechnic flame and release their excess energy as light. Atomic sodium is an intense yellow-orange emitter. Fragile compounds containing copper, barium and strontium produce blue, green and red colors, respectively (*middle*). These unstable molecules are created only briefly in the hot flame. Aluminum or magnesium particles appear as white-hot incandescent sparks when ignited; iron or charcoal particles do not become so hot and therefore radiate dimmer, gold-tinged light (*bottom*).

to create sparklers. Another strontium compound, strontium nitrate, is blended with potassium perchlorate (an oxidizer and chlorine source) and various fuels to create the distinctive red glow of roadside hazard flares.

The structure of a firework is also an intricate brew of craft and engineering. There are two kinds of firework "shells." Cylindrical American-European–style shells, typically seven to 30 centimeters in diameter, are launched from metal, cardboard or plastic mortar tubes. A portion of black powder in the bottom of the shell is ignited, which propels the tube a few hundred meters into the air. A time-delay fuse begins burning when the shell is set off; some seconds later, when the shell is far above the ground, a bursting charge of black powder breaks the shell open and ignites pellets of color composition (called stars), which are irregularly packed into the shell. The stars are expelled in a random pattern of light and color. This type of firework may also contain several ounces of flash-and-sound powder rather than stars and a bursting charge. Such a shell, called a salute, produces a flash of light and a loud boom instead of a burst of colors.

Round Japanese-style chrysanthemum shells are similar in diameter to American-European shells, and they, too, are launched from mortar tubes. In chrysanthemum shells the stars are arranged in a sphere about a central black-powder bursting charge. When the charge explodes, it ignites the numerous stars and distributes them in a round, symmetric pattern. Depending on the size and chemical composition of the stars, the result can vary from a quick flash to an extended trail. The trail may even change color if the stars contain more than one layer of color-producing composition.

Some American-European shells contain several compartments, each with its own bursting charge and stars (or flash-and-sound powder). When one compartment explodes, it ignites a time-delay fuse that leads to the next compartment. In this way, a single shell can produce multiple bursts. Incredibly, the protective barriers that separate the explosive compartments are fabricated from nothing more exotic than cardboard.

In addition to light, pyrotechnics is often exploited for its production of heat. The best-known heat-generating pyrotechnic—the safety match—contains an energetic blend of potassium chlorate oxidizer and sulfur, along with a gluelike fuel and binder.

Calcium silicide fuel mixed with iron oxide generates a moderate amount of heat but no gas. During World War II small pyrotechnic devices containing this composition and a fuse were built into cans of rations so that they could be warmed in the absence of a stove. Time-delay mixtures, usually pressed columns containing boron, tungsten or silicon fuel, produce a controlled dose of heat for a specific length of time that can then set off a larger reaction. Such mixtures are used to control the timing sequence in various aerospace devices, including the exploding bolts that rapidly jettison emergency-exit hatches and spent rocket stages; similar time delays prevent hand grenades from exploding as soon as the pin is pulled and the lever released. Decoy compositions have been developed to protect aircraft from enemy heat-seeking missiles. These compositions emit infrared radiation that emulates the thermal signature of a jet engine.

Heat production is often associated with the emission of smoke and gas. Colored smoke grenades, used for signaling and for daytime displays, contain a mixture of potassium chlorate oxidizer and sugar fuel that, when activated, vaporizes organic dyes to create a richly hued smoke cloud. Sugar is used because it burns at a low temperature; a hotter flame would disintegrate the dyes.

Solid-fuel rockets are in essence giant pyrotechnic devices designed to optimize gas production. Each space-shuttle booster rocket contains half a million kilograms of a propellant consisting of energetic pulverized aluminum fuel and ammonium perchlorate oxidizer; the mixture also includes a special fuel and binder called polybutadiene–acrylic acid–acrylonitrile terpolymer (PBAN). When oxidized, PBAN releases copious quantities of carbon monoxide and carbon dioxide gas and steam that help loft the shuttle into space. Ammonium perchlorate is well suited for this application because its decomposition products are all gases, and so it enhances the rockets' thrust.

Gas generation on a smaller scale creates the whistle effect heard in some fireworks. Compositions containing potassium perchlorate oxidizer and an organic acid salt (such as sodium salicylate—a chemical cousin of aspirin) burn one layer at a time and emit gas in spurts. When such compositions are pressed into narrow tubes, the rapid pulses of escaping gas create a whistling sound.

The most appropriate application for a particular pyrotechnic mixture is largely determined by the reactivity of its oxidizer and its fuel. The reactivity of the fuel is closely related to the amount of energy (heat of combustion) liberated when it combines with oxygen. Metals release large amounts of energy when oxidized; sugar releases relatively little. Charcoal and natural materials such as red gum—a tree secretion—produce the moderate heat needed to activate the color-producing compounds in a firework.

The reactivity of an oxidizer depends on two main factors: decomposition temperature and heat of decomposition. At decomposition temperature, the oxidizer begins to release oxygen at a significant rate. Heat of decomposition, as the term implies, is the amount of heat required to decompose the oxidizer in order to release the oxygen. This amount can be positive (endothermic), in which case decomposition absorbs heat, or negative (exothermic), in which case it generates heat.

Potassium chlorate decomposes at a low 360 degrees C and is exothermic; it is used in smoke grenades and household matches because it is energetic and easily activated. At the other extreme, iron oxide decomposes at nearly 1,500 degrees C and is strongly endothermic. It can be activated only by an energetic metallic fuel such as aluminum.

Packaging and the homogeneity of a pyrotechnic mixture also affect its reaction rate. As every maker of a pipe bomb knows, confinement significantly speeds up the pyrotechnic process by concentrating heat and hot gas near the reaction site. A mixture that burns at a controlled rate in the open can explode violently if confined. In general, the greater the homogeneity of the fuel and oxidizer, the faster the burning rate.

Once, while conducting a seminar, I was asked why liquids are not widely used as pyrotechnics, since they should mix more thoroughly and so produce more reactive compositions than solids. The answer, I realized, is that liquids intermingle too well. Liquid compositions would be extremely homogeneous and therefore highly reactive and sensitive to ignition. Liquids also could settle out during storage, thereby upsetting the chemical balance. Early versions of dynamite, which consisted of porous materials (such as sawdust) soaked in liquid nitroglycerin, were extremely unstable precisely for this reason.

Reactivity is greatest when the oxidizer and fuel are blended at the atomic level and the electron-accepting oxidizer is located immediately adjacent to the atom or ion of fuel that donates the electrons when the pyrotechnic reaction is initiated. Such energetic atomic mixtures are, strictly speaking, explosives not pyrotechnics, but the principles underlying their behavior are similar. Nitroglycerin, for instance, has the molecular formula $C_3N_3H_5O_9$. A small disturbance (heat or impact, for instance) causes it to decompose into carbon dioxide (CO_2), water (H_2O), nitrogen (N_2) and a little excess oxygen (O_2). In the process, nitrogen-oxygen atomic bonds are replaced by far more stable carbon-oxygen, hydrogen-oxygen and nitrogen-nitrogen bonds; the result is a violent release of energy.

A less familiar, but increasingly important, material of this kind is sodium azide, the active component in automotive airbags. This compound consists of interpenetrating lattices of ions of sodium and azide (a group of three chemically bound nitrogen atoms). An energetic impact disrupts the lattice structure. The sodium combines with oxygen while the nitrogen atoms regroup into pairs to form a large quantity of nitrogen gas.

The history of the pyrotechnics industry, both in the U.S. and abroad, has been filled with tragic accidents that have occurred during the manufacturing process, such as the devastating explosion that destroyed the Grucci plant in Long Island, N.Y., in 1983. Improving safety demands a detailed understanding of the phenomenon of ignition.

Ignition begins when energy from

HOUSEHOLD MATCH is a specialized pyrotechnic device. A reaction between potassium chlorate oxidizer in the match and red phosphorus in the striker produces a flame that ignites the mixture of potassium chlorate and gluelike fuel in the match head. All the pyrotechnic effects—heat, smoke, light, gas and sound—are present. Some early matches were lit by dipping a potassium chlorate composition in sulfuric acid. The combination of sulfur and potassium chlorate is so sensitive to ignition that mixtures of the two were outlawed in England in 1875. Use of potassium chlorate is also restricted in fireworks in the U.S.

some source—a flame, friction, impact, spark, elevated temperature or even laser beam—breaks the chemical bonds in a pyrotechnic mixture. As a result, more stable bonds are formed and energy is released. If the energy is adequate to activate the next layer of the mixture, the reaction continues; if the energy is absorbed by the surrounding material or is insufficient to activate the next layer, the reaction dies out.

Fred L. McIntyre and his co-workers at the John C. Stennis Space Center's Hazard Test Range in Mississippi have analyzed a series of pyrotechnic compositions to determine how sensitive each is to ignition by various energy sources. These studies demonstrated that critical factors controlling sensitivity are the amount of heat generated by the reaction and the ignition temperature—the minimum temperature necessary to induce a rapid, self-propagating reaction. Ignition sensitivity is also affected by the particle size of the chemical components and by the grain size of the blended composition; fine grains ignite more readily than large ones. Ease of ignition by friction depends on the presence of abrasive materials in the mixture. Adding a lubricant, such as wax, can significantly reduce the likelihood of friction-induced ignition.

Safety is also important for the users of pyrotechnics—often children celebrating Independence Day, Bastille Day, Guy Fawkes Day, New Year's or other holidays traditionally accompanied by fireworks. In 1976 the U.S. Consumer Product Safety Commission enacted strict federal standards for consumer fireworks. The European Community nations have a wide range of safety standards that they are attempting to incorporate into a single code as part of their legal and economic unification.

CELEBRATIONS have made use of fireworks for centuries. A chromolithograph of the opening in 1883 of the Brooklyn Bridge spanning the East River recorded a particularly imaginative display (*top*). The bridge's centennial was greeted by another dazzling demonstration of pyrotechnic art (*bottom*). The 1983 display was arranged by the Gruccis, a major fireworks-producing family in the U.S.

There is no question—based on numerous conversations I have had with researchers in many fields—that many science careers were stimulated by early experiments in pyrotechnics. The need to come up with a quick explanation to parents concerning a basement filled with smoke from a successful experiment probably has also sparked more than a few legal careers.

The family-business nature of the civilian pyrotechnics industry and the classified nature of much defense-related work have made academic training in the field difficult to acquire. The only academic pyrotechnics courses in the U.S. that I am aware of are several annual, one-week seminars at Washington College in Maryland, taught by several colleagues and me. Fortunately, a number of organizations actively encourage treating pyrotechnics as a science and publish research in the field. These groups include the International Pyrotechnics Society and the more amateur-oriented Pyrotechnics Guild International. An occasional journal called *Pyrotechnica* also contains articles on current work.

The most active supporter of the civilian industry continues to be the viewing public. Since 1976 the annual consumption of fireworks in the U.S. by private individuals for family and neighborhood celebrations has doubled. The centuries-old tradition of fireworks displays is still fascinating, despite competition from rock concerts, music videos and other eye-dazzling and ear-filling forms of entertainment. Evidently modern technology has yet to find a match for the excitement one feels when the work of a master pyrotechnician explodes in the night sky with a clap of thunder and a brilliant shower of color.

FURTHER READING

PYROTECHNICS FROM THE VIEWPOINT OF SOLID STATE CHEMISTRY. Joseph H. McLain. The Franklin Institute Press, 1980.

THE CHEMISTRY OF PYROTECHNICS. John A. Conkling. Marcel Dekker, Inc., 1985.

FIREWORKS: THE ART, SCIENCE AND TECHNIQUE. Second Edition. Takeo Shimizu. Pyrotechnica Publications, 1988.

FIREWORKS: A HISTORY AND CELEBRATION. George Plimpton. Doubleday & Co., Inc., 1989.


The shorebirds who found a new wetland.
There's a saltwater wetland in Mississippi at the edge of a pine forest. It's the kind of place nature might have made. But it actually was created a few years ago by people who work in the nearby refinery. They not only agreed to replace an existing wetland but made it even more valuable as a wildlife habitat. Now herons and whimbrels, egrets and plovers are among more than a dozen kinds of birds that congregate there to rest and feed. Do people need to create new places so that nature can spread its wings?
People Do.
Chevron
For more information, write: People Do-W, P.O. Box 7753, San Francisco, CA 94120


Begin a voyage of discovery that takes you from quarks to the cosmos, and beyond.
SLEEP
THE SCIENCE OF STRUCTURES AND MATERIALS
ANIMAL NAVIGATION
TALBOT H WATERMAN
FROM QUARKS TO THE COSMOS
Tools of Discovery

Join the Scientific American Library now by examining *From Quarks to the Cosmos* for 15 days risk-free.

You are invited to embark upon a journey to the newest frontiers of science. From the mysteries of the brain . . . to the most spectacular phenomena in the universe.

Join us and you will explore the most exciting discoveries on the scientific horizon today as a member of the SCIENTIFIC AMERICAN LIBRARY.

From Molecules, *a polarized light view of light heating oil; courtesy Manfred Kage/Peter Arnold, Inc.*

Written by leading scientists in accessible, compelling prose.

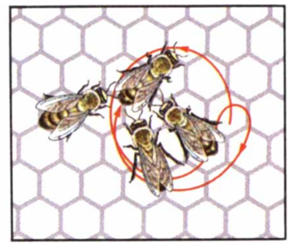
SCIENTIFIC AMERICAN LIBRARY books are created in a unique collaboration between our editors and some of the world's most renowned scientists (including many Nobel laureates). The result is a series of books that are readable *and* authoritative. Each volume . . .

- Covers a topic of major interest with immediate relevance to ongoing research.
- Carries the authority of an author who has made significant contributions to the field.
- Is written in clear, compelling prose that's uniquely accessible even to the non-expert.

No other book club delivers the literature of science with such uncompromising editorial quality.

If science is your passion, reward yourself. Begin your membership in the SCIENTIFIC AMERICAN LIBRARY today.

Beautiful, richly illustrated volumes that are made to endure.



From The Honey Bee, *foragers dance to indicate a food source; drawing by John Hatzakis, adapted from K. von Frisch,* Tanzsprache und Orientierung der Bienen.

From the moment you receive your first book, you'll see that SCIENTIFIC AMERICAN LIBRARY volumes are produced to exacting standards. Each work:

- Features abundant photographs, paintings and drawings in full, vibrant color.
- Includes many helpful diagrams, charts, and maps to graphically illustrate the text.
- Is impeccably printed on high quality paper and clothbound for uncommon durability.

The books in the SCIENTIFIC AMERICAN LIBRARY are fascinating reading *and* they make a handsome addition to your library.

Join now by sending for your premier volume without risk or obligation.

Join the SCIENTIFIC AMERICAN LIBRARY now by previewing *FROM QUARKS TO THE COSMOS,* written by 1988 Nobel Prize-winning physicist Leon Lederman and cosmologist David Schramm. This intriguing book explores the revolutionary discoveries made through the powerful merger of particle physics and cosmology.

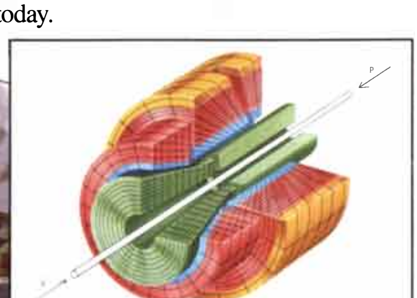
From Eye, Brain, and Vision, *the neurobiology group at Harvard Medical School, 1963, including Nobel Prize winners David Hubel, standing right, and Torsten Wiesel, seated right. Photograph by Joseph Gagliardi.*

This is just the first of many fascinating volumes you'll receive. In months to come you'll discover books covering a wide range of subjects, including *Molecules, Sexual Selection,* and *Eye, Brain, and Vision.*

As a member, you'll enjoy:

- A NEW BOOK EVERY TWO MONTHS or so, shipped direct to your doorstep.
- 15-DAY FREE EXAMINATION. Review each book without risk. You may keep it or return it for full credit.
- A LOW, MEMBERS-ONLY PRICE. Pay just $24.95 (25% off the retail price) for each book you keep.
- NO PREPAYMENT OR MINIMUM PURCHASE. Pay only for the books you choose to keep.
- OUR GUARANTEE that you'll be fully satisfied or *you may cancel at any time without penalty.*

TO JOIN, WRITE TO SCIENTIFIC AMERICAN LIBRARY, P.O. BOX 646, HOLMES, PA 19043, OR CALL TOLL-FREE 1-800-345-8112.



◄ *From* From Quarks to the Cosmos, *a schematic of a particle collider detector; drawing by George Kelvin.*

SCIENTIFIC AMERICAN LIBRARY

AN EXCITING NEW GENRE IN THE LITERATURE OF SCIENCE

Chestnut Blight

A fungus that has ravaged the American chestnut is itself coming under biological attack. Plant pathologists hope the new parasite will eventually control the old one

by Joseph R. Newhouse

A hundred years ago the American chestnut (*Castanea dentata*) could have been called the all-American tree. The tall, regal species made up a quarter of the hardwood trees in its natural range, a nearly unbroken chain of forests stretching from Maine to Georgia. It grew faster than most other hardwoods, sometimes reaching a stature of 100 feet and a girth of seven feet. Its straight-grained timber split easily, burned almost smokelessly and resisted rot by virtue of a generous endowment of tannin, a bitter substance also found in tea leaves. Such resistance made the wood ideal for fence posts, telegraph poles, ship masts and railroad ties; the tannin supported a thriving tanning industry wherever the tree was abundant. The nuts—far more flavorful than those of other chestnut species—were a staple of many animals and an ingredient in many confections.

Today the chestnut survives mainly as a minor understory shrub sprouting from the root systems of declining and dead trees. Most mature chestnut trees—from three to four billion—were killed during the first half of this century by a fungal disease thought to have been introduced around 1900 with the importation of chestnut specimens from Asia. Oaks and other hardwoods have filled the considerable gaps left by the chestnut's demise.

Persistent efforts to breed more resistant varieties of the tree have so far been disappointing. Recent advances in biotechnology, however, offer hope that other weapons may be turned against the fungus. Some fungal strains are known to be too weak to kill their host, and the genetic mechanism responsible for the weakness has been partially unraveled. If plant pathologists learn to transmit such weakness to all the other strains, then mature chestnuts may once again adorn the American landscape.

The blight was first noticed in 1904 by Herman W. Merkel of the New York Zoological Park in the Bronx. Alarmed by the appearance of cankers—necrotic depressions—on the park's trees, he asked for and received an appropriation to prune the diseased branches. But within a year—despite a valiant effort—most of the trees became infected, and many of the smaller specimens died. Although the larger trees lingered as long as a decade, they, too, eventually succumbed.

In 1906 William A. Murrill of the New York Botanical Garden, having established that the fungus enters the chestnut through wounds in the bark, called the pathogen *Diaporthe parasitica*. In 1912 taxonomic considerations led mycologists to rename it *Endothia parasitica,* the designation by which the pathogen became widely known. In 1978, again for taxonomic reasons, the fungus received its current name, *Cryphonectria parasitica.*

Unimpeded by natural barriers and faced with a completely defenseless and relatively homogenous host population, *C. parasitica* spread rapidly. By 1950 it occupied the entire natural range of the chestnut [*see illustration on page 108*]. How was it able to spread so rapidly? Like most fungi, it reproduces sexually and asexually. First it scatters sexual spores, called ascospores, on the wind to establish distant colonies. Then the colonies fan out by producing sticky masses of asexual spores, called conidia, which can be splashed about by rain or dispersed from tree to tree on the feet of birds and insects.

Foresters tried to slow the blight's advance by cutting swaths as wide as a mile in forests not yet affected by the blight and by destroying diseased trees, but to no avail. Attempts to replant a region after the dead chestnuts had been removed also failed: the fungus can survive the death of its host by subsisting as a weak parasite on other tree species or as a saprophyte on dead organic matter.

Many plant scientists felt the American chestnut was doomed, but the tree's capacity to sprout repeatedly from the roots has saved it from extinction. Although the sprouts, too, eventually fall to the blight, usually before reproductive age, they provide cuttings by which the breeding stock can be maintained in sheltered environments. Another ray of hope is offered by the stands of mature chestnut trees planted west of the tree's natural range.

The first step toward a cure was to reconstruct the history of the epidemic. If researchers knew the host on which the fungus had originally lived, they might discover how the host had defended itself. The suspicion of earlier investigators that the fungus had entered the U.S. on imported Asian chestnut trees found confirmation in several developments. First, in 1938, blight attacked the European chestnut (*Castanea sativa*) in regions of Italy where experimental nurseries of the Japanese chestnut (*Castanea crenata*) had been established about a decade earlier. Second, cultured isolates of fungus taken from chestnuts in China,

AMERICAN CHESTNUT, 75 feet tall, escaped the blight's attack because it was planted in Lake County, Mich., some 200 miles west of the species' natural range. Such rare survivors are a national treasure, valuable as breeding stock.

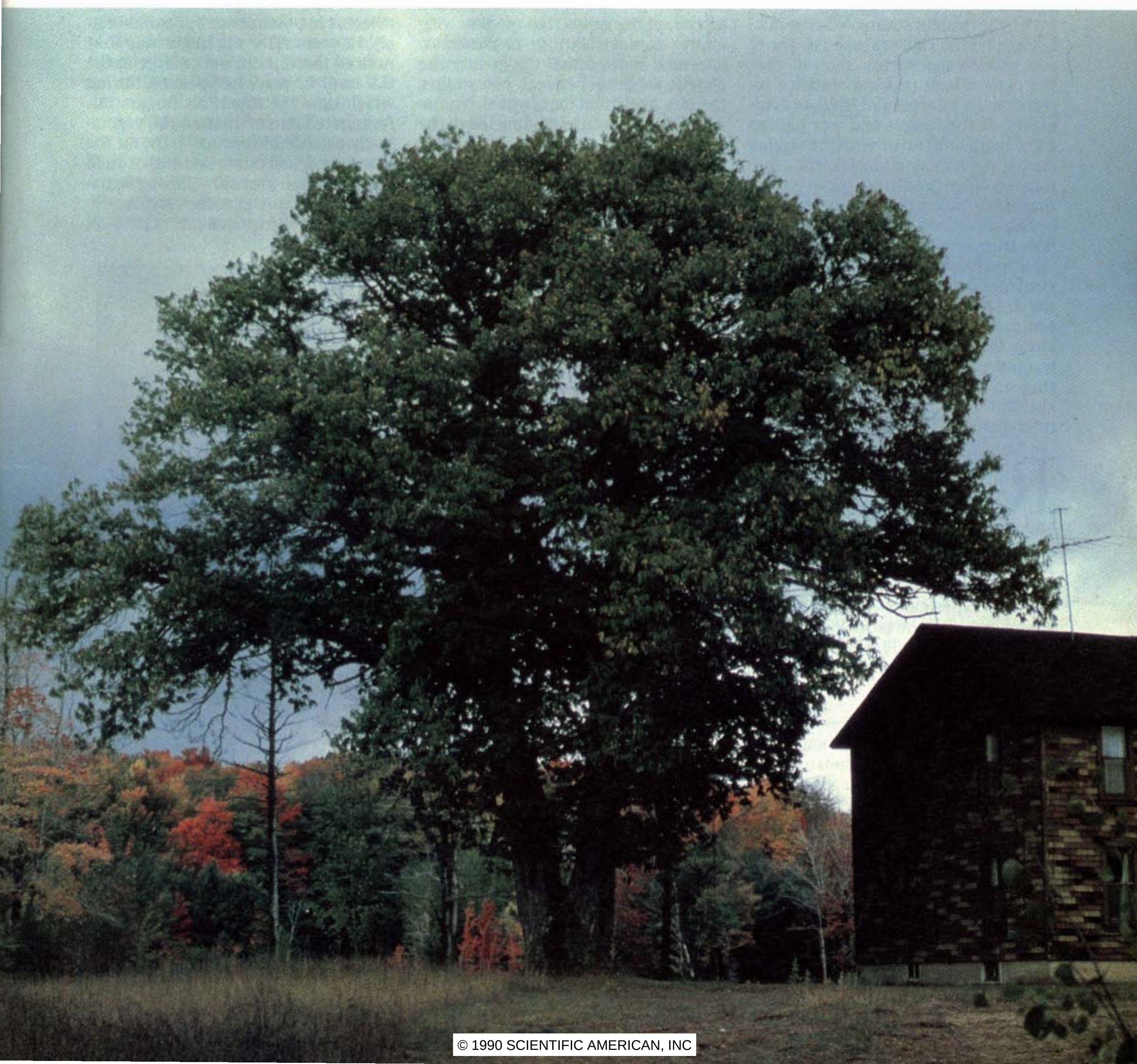
JOSEPH R. NEWHOUSE studies the molecular biology and cytology of fungi that attack plants. Currently he is a postdoctoral fellow at the U.S. Department of Agriculture's Foreign Disease-Weed Science Research Unit in Frederick, Md. He earned a B.S. in biology from Saint Vincent College in Latrobe, Pa., in 1977 and a Ph.D. in plant pathology from West Virginia University in Morgantown in 1988. In his spare time he enjoys photography, computer recreations and golf.

Japan, Europe and the U.S. resembled one another in their virulence and in the morphologies they displayed in artificial culture. Third, as fossil evidence showed, *Castanea* grew in China and Japan long before it appeared elsewhere in the world. Finally, the Asian species proved to be resistant to the fungus. This finding suggests that the Asian trees coevolved with the blight fungus and that resistant trees were favored by natural selection.

The disease progressed more slowly in Europe than in the U.S., partly because the European species is more resistant to the fungus. Moreover, the European chestnut is a small, orchard tree that normally grows in widely separated clumps. The separations—together with such barriers as the Alps—slowed the spread of the blight. Nevertheless, within 25 years the European tree was decimated in Italy and coming under attack in France.

Then, in 1950, an encouraging sign appeared: a number of European chestnuts growing near Genoa were found to possess cankers that had stopped enlarging [*see illustration on page 111*]. Antonio Biraghi of the Forestry Institute of Florence found the trees had walled off some diseased areas with callus tissue, a form of bark. Yet the same trees had other cankers that continued to spread. How could this be? Biraghi suggested that some fungal strains might have lost their virulence. Strangely, more than a dozen years passed before anyone attempted to test his hypothesis. Finally, in 1964, Jean Grente of the National Institute for Agricultural Research in Clermont-Ferrand confirmed that Biraghi had been right.

Grente extracted the fungus from closing cankers and noticed it lacked the orange pigment and most of the pycnidia, or asexual fruiting structures, that characterize the fungus in spreading cankers. Although the unpigmented strains could produce a



canker when inoculated into a European chestnut, the canker grew so slowly that the tree had time to contain it. Because the new strains could not sustain an effective infection, Grente termed them hypovirulent.

Most remarkable of all, Grente and his colleagues discovered that the trait for hypovirulence was transmissible. They found that when agar plugs from hypovirulent and virulent cultures were placed next to one another in petri dishes, the resulting colonies initially resembled those from which the plugs had been taken. After several days, however, new growth of the virulent colony resembled the hypovirulent colony. This discovery encouraged the workers to treat cankers with hypovirulent strains, which quickly stopped the cankers' spread. These and other experiments showed that the determinant of hypovirulence resided in the intracellular fluid, or cytoplasm, of the fungus and was passed from hypovirulent to virulent strains during hyphal anastomosis. (In anastomosis a bridge forms between the threadlike stalks, or hyphae, of neighboring fungi, enabling cytoplasm to flow from one hypha to another.)

Hypovirulence had triumphed over virulence both in the laboratory and in the field. Grente began a biocontrol program in France in 1966 that has since met with great success. In Italy, on the other hand, naturally occurring hypovirulence has led to the chestnut's recovery without human intervention.

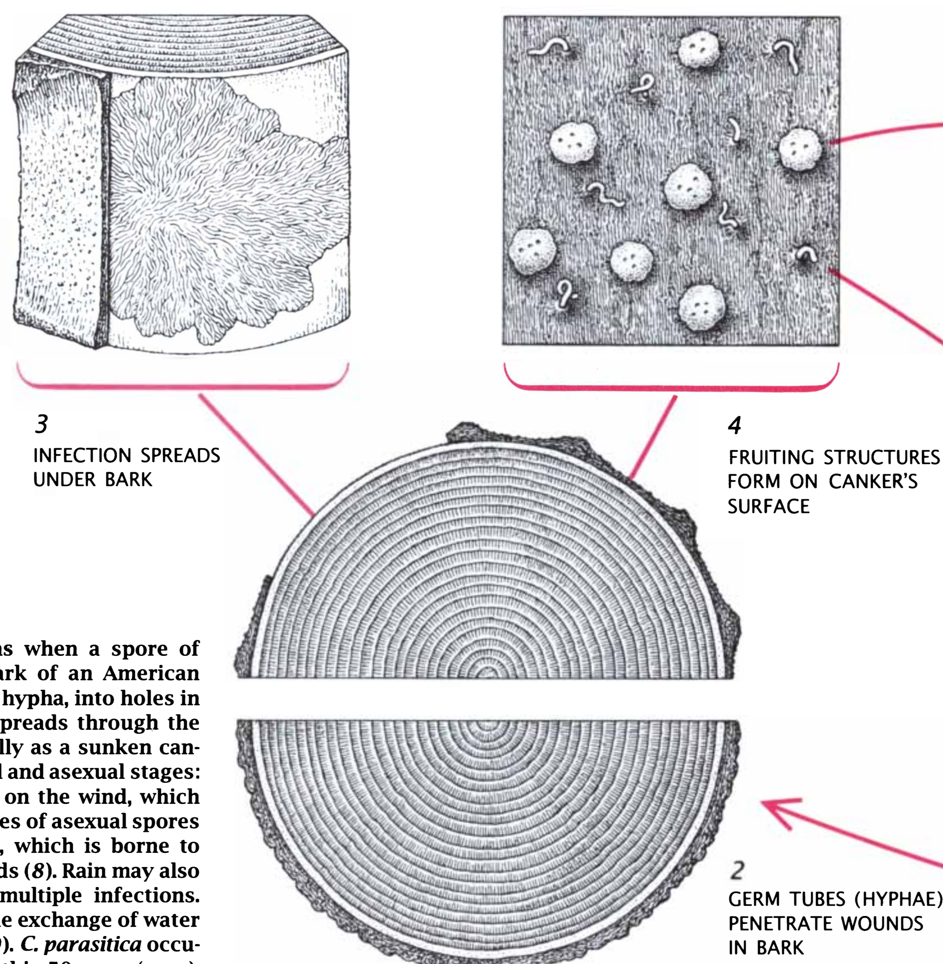
Researchers thought it would be possible to reproduce these splendid results in the U.S., and indeed, early successes generated great excitement. In 1972 Richard A. Jaynes, Neal K. Van Alfen and their colleagues at the Connecticut Agricultural Experiment Station in New Haven impeded the growth of cankers in greenhouse seedlings by inoculating them with European hypovirulent strains supplied by Grente. A year later, after receiving permission from the U.S. Department of Agriculture to undertake field studies, they inoculated a number of trees with European hypovirulent strains and again stopped the growth of cankers. The inoculations had successfully converted most fungi to the hypovirulent form.

But further studies showed inoculation did not always prevent a canker from spreading. An unknown factor was preventing some of the hypovirulent strains from passing their benign traits to virulent strains during hyphal anastomosis. Such selective blockages were the result of what is called vegetative incompatibility. Sandra L. Anagnostakis of the Connecticut station identified the genes that mediate vegetative compatibility in *C. parasitica* and then showed that the greater the genetic variation between two strains, the less likely it is that hyphal anastomosis will occur. Her finding led to the use of compatibility testing to classify virulent fungi into vegetatively compatible groups. Hypovirulent strains known to undergo anastomosis with strains in those groups were then applied to the cankers, a treatment that usually produced better results.

Meanwhile, in 1976, indigenous hypovirulent strains were discovered in a stand of American chestnuts in western Michigan, outside the tree's natural range, and later in states inside the range. Many plant pathologists wondered if the American chestnut would recover on its own, as the European species had done in Italy. But neither inoculation nor natural hypovirulence ended the blight. No American chestnut tree has yet been nursed to maturity in regions infested by *C. parasitica*.

Why has the American chestnut fared worse than the European chestnut? It may be a question of timing: whereas hypovirulence appeared in Italy 12 years after the blight was first noticed there, it did not emerge in the U.S. until 72 years had passed, during which time the American fungus differentiated into more than 100 vegetatively compatible groups. Italy, on the other hand, had only a handful of such groups, which probably allowed hypovirulence to spread more quickly.

Although the fungus has eluded all



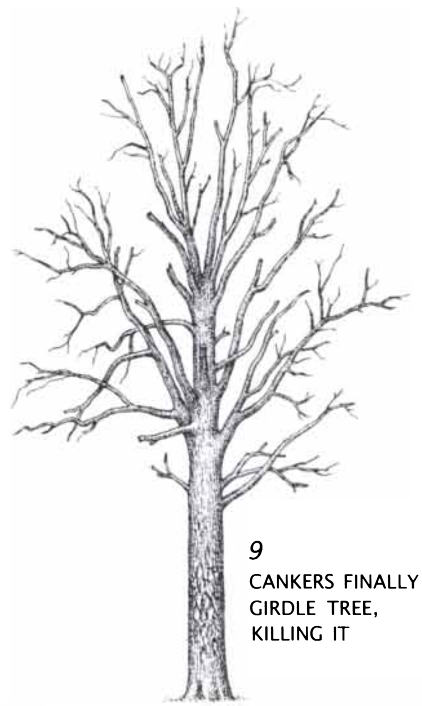
DISEASE CYCLE of chestnut blight begins when a spore of *Cryphonectria parasitica* lands on the bark of an American chestnut (*1*). The spore extends a tube, or hypha, into holes in the bark (*2*), creating an infection that spreads through the cambium (*3*) and manifests itself externally as a sunken canker. Reproduction (*4–6*) proceeds in sexual and asexual stages: sexual spores (ascospores) are scattered on the wind, which carries them great distances (*7*), and masses of asexual spores (conidia) are secreted in a sticky matrix, which is borne to nearby trees on the feet of insects and birds (*8*). Rain may also wash conidia down the tree, triggering multiple infections. Cankers finally kill the host by blocking the exchange of water and nutrients between leaves and roots (*9*). *C. parasitica* occupied the chestnut's entire natural range within 50 years (*map*).

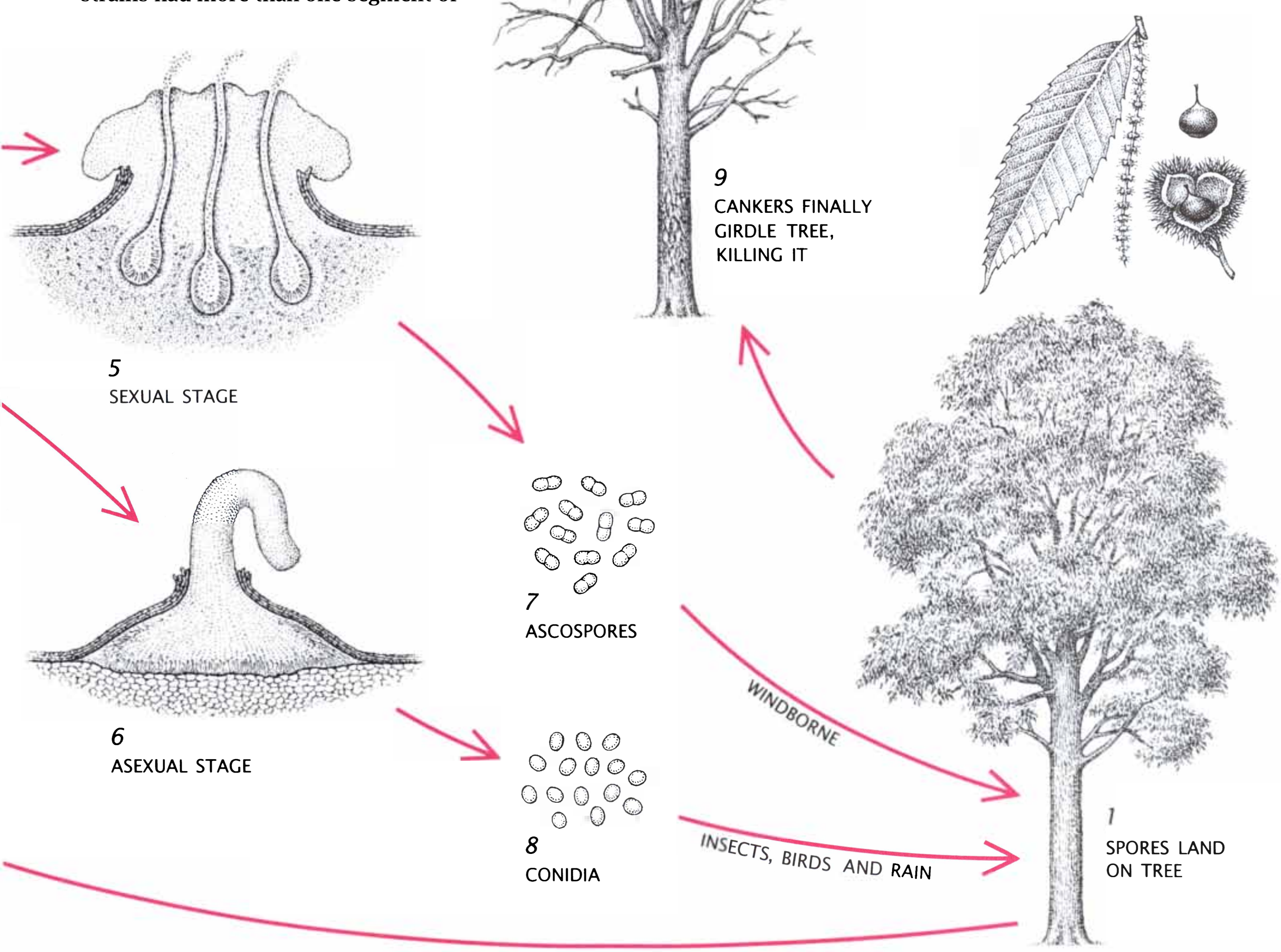
attempts at eradication thus far, recent advances in biotechnology suggest the war may yet be won. The first breakthrough came in 1975, when Van Alfen and his colleagues proved that nuclear material is not passed from a hypovirulent to a virulent strain during hyphal anastomosis and concluded that the agent responsible for hypovirulence must therefore reside in the cytoplasm.

Their finding was followed by another important discovery later the same year. Richard M. Lister of Purdue University and his graduate student Eileen M. Moffitt detected double-strand RNA (dsRNA), a nucleic acid that encodes genetic information in many fungal viruses, in the cytoplasm of two hypovirulent strains from Europe, but they failed to find it in virulent strains from Europe or the U.S. Two years later Peter R. Day and his co-workers at the Connecticut station confirmed the results of Lister and Moffitt and found dsRNA in North American hypovirulent strains. They also showed that some hypovirulent strains had more than one segment of dsRNA. Since then, workers have identified hypovirulent strains possessing as many as 10 segments of different lengths, but no one has yet linked any segment to hypovirulence.

Nevertheless, the role of dsRNA in causing hypovirulence is clear: the transmission of hypovirulence always involves the transfer of dsRNA, and the removal of dsRNA always causes hypovirulent strains to revert to virulence. The cause and effect were clear, but the mechanism by which they were linked was not.

As the first step in identifying the mechanism, plant pathologists sought the origins of the dsRNA. Viruses were the most likely source of the dsRNA, because most viruses known to infect fungi have their genetic information encoded on dsRNA. In 1978 James A. Dodds, then at the Connecticut sta-

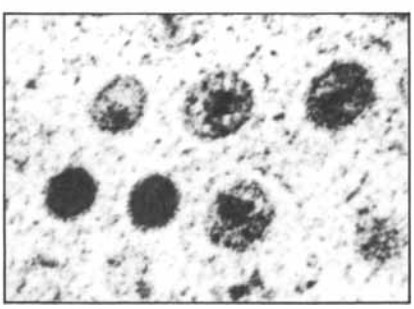




tion, first presented evidence for the existence of a viruslike particle (VLP) in a hypovirulent strain. He extracted club-shaped particles 100 nanometers long from a European hypovirulent strain and found dsRNA associated with them.

Similar particles were later taken from other European hypovirulent strains. None, however, has been extracted from virulent strains from either the U.S. or Europe or from North American hypovirulent strains. Yet, as I shall presently show, there is evidence that VLP's are always present wherever hypovirulence is found.

Clearly, much could be learned about VLP's from their position inside the fungal cytoplasm, but for many years efforts to observe them in situ were fruitless because delicate subcellular structures were poorly preserved by conventional chemical fixation. In 1981 it occurred to me that snapshots of microscopic portions of the fungus could be taken with the help of freeze substitution. The technique enables one to fix cellular structures permanently by first flash-freezing and then soaking them in chemicals that leach out their water and stabilize the cytoplasm and its contents. The tissue is then embedded in plastic, sectioned, stained and imaged by a transmission electron microscope.



I found VLP's in the hyphae and conidia (asexual spores) of European and North American hypovirulent strains but not in genetically identical virulent strains that were free of dsRNA. Although the spherical, membrane-bounded particles were smaller than those Dodds had isolated and of a different shape, they probably are identical. To begin with, the transmission electron microscope always makes particles appear larger in isolation than they do in a tissue section. Second, cytochemical tests found that the particles contain nucleic acid and are surrounded by a lipid membrane, a membrane highly susceptible to distortion. Thus, the shape of the particles could have been distorted by Dodds's extraction procedure.

In European hypovirulent strains, VLP's formed aggregates surrounded by rough endoplasmic reticulum and other elements of the cell's machinery of synthesis. In North American strains, VLP's were sparsely scattered throughout the cytoplasm, which may account for previous failures to extract them from hyphae.

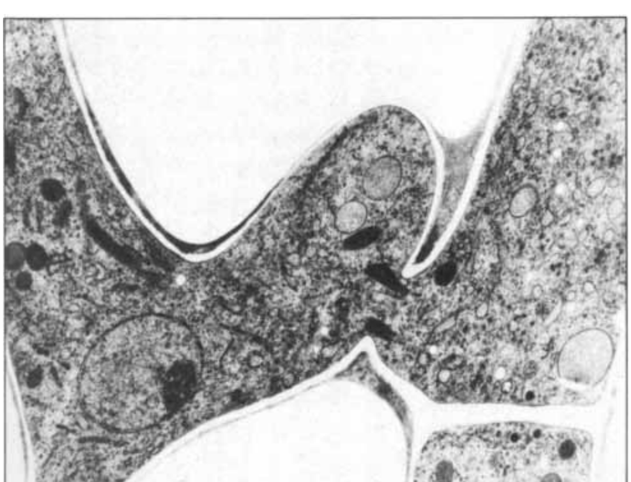
VLP's resemble only one cellular component normally found in *C. parasitica*—the apical vesicles, which help to produce cell membranes and walls at the growing tips of hyphae. This resemblance leads me to suspect that the mechanism of growth has been turned here to another purpose—that of protecting the fungal cell against foreign nucleic acid, that is, dsRNA.

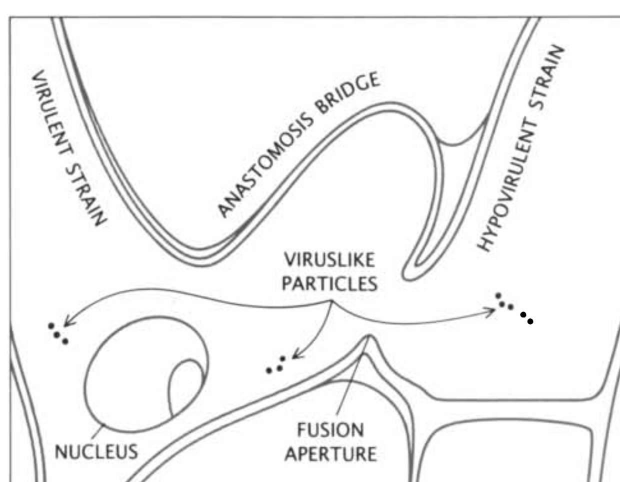
In this scenario, VLP's are end products, formed when foreign nucleic acids are sealed in structures made from cell-wall precursors. In the final phase, VLP's would migrate to the cell membrane and fuse with it, thus depositing their contents in the cell wall. In this way, dsRNA is purged from the cell.

Yet if the dsRNA in VLP's is in fact neutralized, then what causes the morphological aberrations in strains of *C. parasitica* infected with dsRNA? Perhaps most of the dsRNA in hypovirulent strains never gets trapped in VLP's but rather floats freely in the cytoplasm, as viroids do. Van Alfen suggests the dsRNA may represent the genome of a defective virus, one that has lost the ability to produce a capsid, or protein coat. In any case, the dsRNA associated with hypovirulence is unique among the biological agents known to infect fungi.

The freeze-substitution technique also allowed me to study hyphal anastomosis between various strains of *C. parasitica*. I was able to confirm, for the first time, that cytoplasmic continuity does exist between vegetatively compatible strains. In addition, I found VLP's in the anastomosis bridge and in the virulent strain only four hours after anastomosis began [*see illustration below*]. Although incompatible strains also start anastomosis, their cytoplasm degenerates soon after fusion, preventing the passage of dsRNA, and hence the trait for hypovirulence, from one strain to another.

The tools of molecular biology have enabled workers to discover some secrets of hypovirulence. A group led





ANASTOMOSIS, four hours after it began, is seen in this transmission electron micrograph of two fungal strains whose structures were flash-frozen and chemically fixed. Hyphae from each strain, one virulent and the other hypovirulent, have formed a bridge through which they can exchange cellular fluid (*drawing*). Viruslike particles (*close-up*) are present in the bridge and the virulent strain's hypha, indicating that double-strand RNA has already been passed into the virulent strain.

by Robert E. Rhoads of the University of Kentucky used a technique called hybridization, by which they tried to bind radioactively labeled dsRNA from a hypovirulent strain to dsRNA from other hypovirulent strains. Successful binding indicates that the probe and its target are closely related. Rhoads's group found, however, that hybridization occurs only when the probe and its target come from strains that originated on the same continent—either Europe or North America. This result proved that hypovirulence arose independently in the two continents and that it involves two distinct groups of dsRNA.

Additional hybridization studies found two hypovirulent strains in Michigan whose dsRNA would not bind with dsRNA from many other strains around the country. There may therefore be many unrelated species of dsRNA in hypovirulent *C. parasitica,* at least in the U.S. But other hybridization studies have shown that in some hypovirulent strains, the multiple segments of dsRNA are related to one another. This finding led Donald L. Nuss of the Roche Institute of Molecular Biology in Nutley, N.J., to speculate that the dsRNA may be genetically less complicated than it appears. If so, identification of the genes responsible for hypovirulence could be simplified.

To pinpoint the dsRNA segment or combination of segments that cause hypovirulence, one must first determine how dsRNA affects the cell. Molecular biologists at the Roche Institute have begun to study this problem by sequencing one strand of dsRNA from a European hypovirulent strain. They translated a portion of the strand in vitro to produce a protein having a molecular weight of 29 kilodaltons. This protein was the first product synthesized from dsRNA associated with hypovirulence.

To determine how dsRNA's protein products induce hypovirulence in *C. parasitica,* researchers are now using an enzyme called reverse transcriptase to make a DNA copy of dsRNA genes that specify the production of individual proteins. The workers then insert the DNA into virulent strains, a process called transformation, in an attempt to discover what each protein does to the fungus.

Armed with this information, plant pathologists will be able to determine whether the mechanisms responsible for hypovirulence necessarily doom the fungus to reproductive weakness. If not, then genetic engineering may be able to create strains too weak to kill chestnut trees but robust enough to compete with virulent strains in the forest. Or perhaps the genetic determinants of vegetative incompatibility may be unraveled, so that hypovirulence can be passed on to all strains of *C. parasitica*. That objective may be achieved by studying recently discovered hypovirulent strains called multiconverters, which can pass hypovirulence to members of several vegetatively compatible groups.

SPREADING CANKER caused by virulent *C. parasitica* girdles a chestnut and will kill it (*left*). When such a canker is treated with hypovirulent strains, the virulent fungus weakens, allowing the tree to wall off the infection with callus tissue (*right*).

Biologists have another way to try to save the American chestnut: they can breed a more resistant tree even as they work to debilitate its parasite. Plant geneticists are crossing the chestnut with its more resistant relatives in Asia, then breeding the resulting hybrids with pure American stock in an effort to retain the desired qualities: the beauty of the American tree and the blight resistance of the Asian tree. Other geneticists have followed countless leads to find chestnut trees that survived the blight, notably in North Carolina and Virginia. They graft cuttings taken from the survivors onto rootstock of blight-resistant specimens. Even so, such breeding is slow and its results uncertain.

Here, too, molecular biology may one day make a contribution. If the genes that confer resistance can be identified, then eventually it may be possible to incorporate them into germlings from which blight-resistant American chestnuts might grow. Neither breeding nor gene-splicing can return the chestnut to its former glory overnight, for other species have long since filled the ecological gaps created by the tree's destruction. But restoration of the chestnut to the yards of those who love it would be a sweet victory. Then might Robert Frost's prophecy come true:

Will the blight end the chestnut?
The farmers rather guess not.
It keeps smoldering at the roots
And sending up new shoots
Till another parasite
Shall come to end the blight.

FURTHER READING

DEATH OF THE AMERICAN CHESTNUT. George H. Hepting in *Journal of Forest History,* Vol. 18, No. 3, pages 60-67; July, 1974.

PROCEEDINGS OF THE AMERICAN CHESTNUT SYMPOSIUM. West Virginia University, Morgantown, W. Va., January 4-5, 1978. Edited by William L. MacDonald et al. West Virginia University Books, 1978.

CHESTNUT BLIGHT, OTHER ENDOTHIA DISEASE, AND THE GENUS *ENDOTHIA*. Martha K. Roane, Gary J. Griffin and John R. Elkins. American Phytopathological Society Monograph Series. St. Paul, Minn., APS Press, 1986.

VIRUS-LIKE PARTICLES IN HYPHAE AND CONIDIA OF EUROPEAN HYPOVIRULENT (DSRNA-CONTAINING) STRAINS OF *CRYPHONECTRIA PARASITICA*. Joseph R. Newhouse, William L. MacDonald and Harvey C. Hoch in *Canadian Journal of Botany,* Vol. 68, No. 1, pages 90-101; January, 1990.

Ramsey Theory

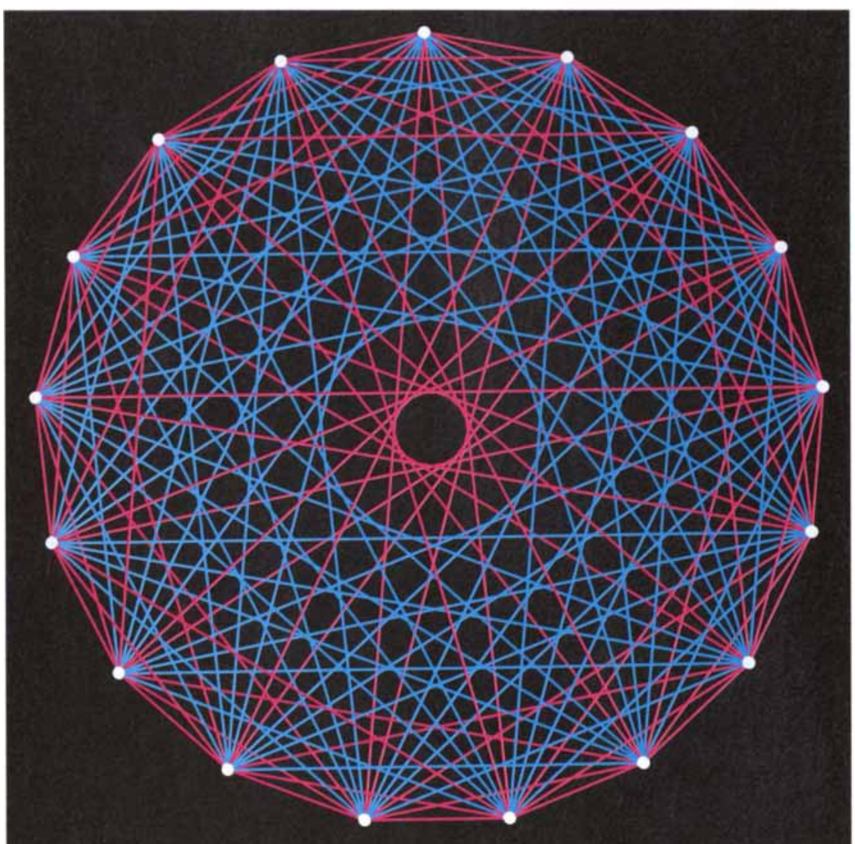
The brilliant mathematician Frank Plumpton Ramsey proved that complete disorder is an impossibility. Every large set of numbers, points or objects necessarily contains a highly regular pattern

by Ronald L. Graham and Joel H. Spencer

According to a 3,500-year-old cuneiform text, an ancient Sumerian scholar once looked to the stars in the heavens and saw a lion, a bull and a scorpion. A modern astronomer would be more likely to describe a constellation as a temporary collection of stars, which we earthlings observe from the edge of an ordinary galaxy. Yet most stargazers would agree that the night sky appears to be filled with constellations in the shape of straight lines, rectangles and pentagons. Could it be that such geometric patterns arise from unknown forces in the cosmos?

Mathematics provides a much more plausible explanation. In 1928 Frank Plumpton Ramsey, an English mathematician, philosopher and economist, proved that such patterns are actually implicit in any large structure, whether it is a group of stars, an array of pebbles or a series of numbers generated by throws of a die. Given enough stars, for instance, one can always find a group that very nearly forms a particular pattern: a straight line, a rectangle or, for that matter, a big dipper. In fact, Ramsey theory states that any structure will necessarily contain an orderly substructure. As the late American mathematician Theodore S. Motzkin first proclaimed some 25 years ago, Ramsey theory implies that complete disorder is an impossibility.

Ramsey theorists struggle to figure out just how many stars, numbers or figures are required to guarantee a certain desired substructure. Such problems often take decades to solve and yield to only the most ingenious and delicate reasoning. As Ramsey theorists search for solutions, they assist engineers attempting to build better communications networks as well as information transmission and retrieval systems. Ramsey theorists have also discovered some of the mathematical tools that will guide scientists in the next century. Perhaps most important, Ramsey theorists are probing the ultimate structure of mathematics, a structure that transcends the universe.



PARTY PUZZLE typifies the problems that Ramsey theory addresses. How many people does it take to form a group that always contains either four mutual acquaintances or four mutual strangers? In the diagram, points represent people. A red edge connects people who are mutual acquaintances, and a blue edge joins people who are mutual strangers. In the group of 17 points above, there are no four points whose network of edges are either completely red or completely blue. Therefore, it takes more than 17 people to guarantee that there will always be four people who are either acquaintances or strangers. In fact, in any group of 18 people, there are always either four mutual acquaintances or four mutual strangers.

Unlike many branches of mathematics that interest professionals today, Ramsey theory can be presented intuitively. Indeed, the charm of Ramsey theory is derived in part from the simplicity with which the problems can be stated. For example, if six people are chosen at random (say, Alfred, Betty, Calvin, Deborah, Edward and Frances), is it true that either three of them mutually know one another or three of them mutually do not know one another?

We can solve the "party puzzle" in many ways. We could list all conceiv-

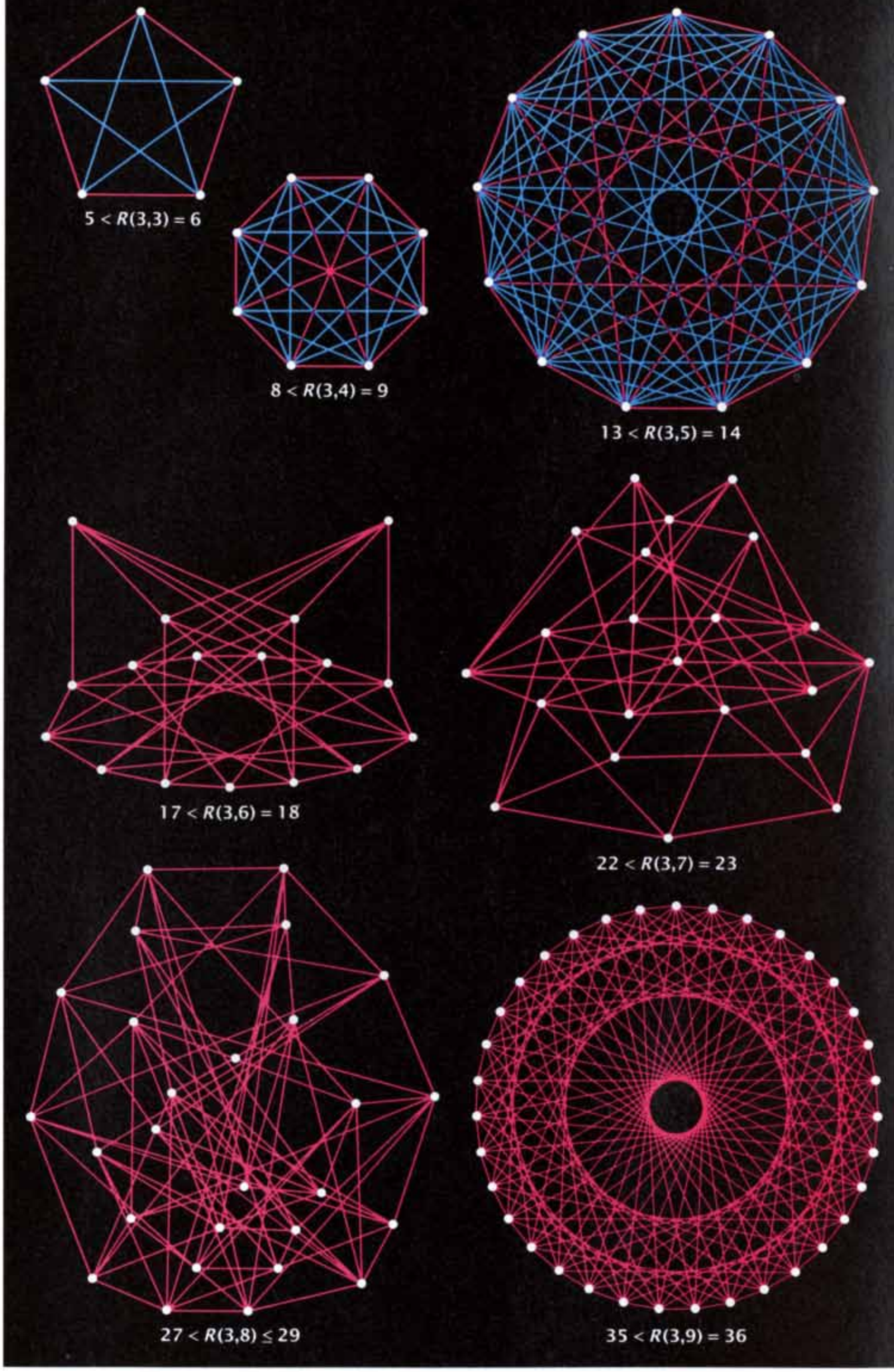
RONALD L. GRAHAM and JOEL H. SPENCER have written the definitive text on Ramsey theory with Bruce L. Rothschild of the University of California, Los Angeles. Graham is adjunct director for research of the information sciences division at AT&T Bell Laboratories and professor of mathematics at Rutgers University. In 1962 he received his Ph.D. in mathematics from the University of California, Berkeley. This is Graham's third article for *Scientific American*. Spencer is professor of mathematics and computer science at New York University's Courant Institute of Mathematical Sciences. In 1970 he earned his Ph.D. in mathematics from Harvard University. Spencer has a particular interest in the history of Hungarian mathematics, such as the work of Paul Erdős, who pioneered much of Ramsey theory.

able combinations and check each one for an acquainted or unacquainted group of three. But because we would have to check 32,768 (or 2^{15}) combinations, this brute-force method is neither practical nor insightful.

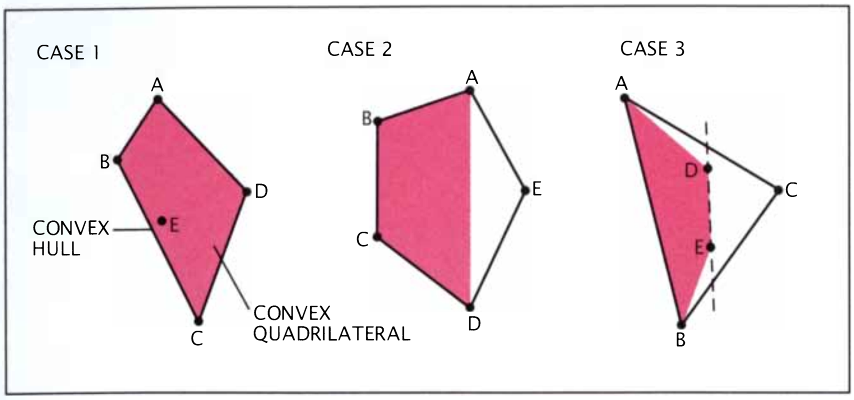
Fortunately, we can find the answer by considering two simple cases. In the first case, suppose Alfred knows three (or more) of the others, say, Betty, Calvin and Deborah. If either Betty and Calvin or Betty and Deborah or Calvin and Deborah are mutual acquaintances, then Alfred and the acquainted pair make three people who know one another. Otherwise Betty, Calvin and Deborah are mutual strangers. In the second case, suppose Alfred knows only two (or fewer) of the others, say, Betty and Calvin. If either Deborah and Edward or Deborah and Frances or Edward and Frances are strangers, then Alfred and the unacquainted pair make three people who do not know one another. Otherwise Deborah, Edward and Frances are mutual acquaintances. In just six sentences, we have shown why any party of six people must include at least three mutual acquaintances or three mutual strangers. More to the point, the solution to the party puzzle is a special case of Ramsey theory.

By generalizing this special case, we can give the full theorem. Instead of considering six people in the problem, we can have any number of people or, for that matter, any number of objects. We need not restrict ourselves to two relationships, acquaintances and strangers. We can have any number of mutually exclusive relationships—for instance, friends, foes and neutral parties.

We can then describe the full Ramsey theorem. If the number of objects in a set is sufficiently large and each pair of



RAMSEY NUMBERS are defined as the smallest value of *n* such that in a group of *n* points either a group of *j* points forms a complete network of red edges or a group of *k* points forms a complete network of blue edges. The diagrams above indicate how large a particular Ramsey number should be. The first diagram shows five points connected by red and blue edges in such a way that no three points form either a red or a blue complete network. Hence, the first diagram implies that the Ramsey number for red three and blue three must be greater than five. In a similar manner, one can argue that the second diagram suggests that the Ramsey number for red three and blue four is greater than eight. By other more complicated techniques, it can be demonstrated that the Ramsey number for red three and blue three is six and that the number for red three and blue four is nine. All of the exact Ramsey numbers that are known are given above, except the Ramsey number for red four and blue four, whose diagram is shown on the opposite page. (In some of the diagrams the blue edges are omitted for simplicity.) The Ramsey number for red three and blue eight has been proved to be greater than 27 and less than or equal to 29. Recently it was shown (but not yet verified) to be 28.



RAMSEY THEORY was rediscovered in 1933 when a young student, Esther Klein, introduced a geometric problem: if five points lie in a plane so that no three points form a straight line, prove that four of the points will always form a convex quadrilateral. All cases of the problem are variations of the three above. The simplest case occurs when the convex hull—the convex polygon enclosing all the points—is a quadrilateral. If the convex hull is a pentagon, then any four points can be connected to form a quadrilateral. A triangular convex hull will always contain two points, here *D* and *E*. The line *DE* splits the triangle so that two points, *A* and *B*, are on one side. The four points *ABDE* must form a convex quadrilateral.

objects has one of a number of relations, then there is always a subset containing a certain number of objects where each pair has the same relation.

Frank Ramsey, who first proved this statement in 1928, grew up in Cambridge, England. His father, Arthur S. Ramsey, was professor of mathematics and president of Magdalene College at Cambridge. In 1925 young Ramsey graduated as the university's top mathematics student. Although philosophy and mathematical logic chiefly engaged him, he also contributed to economics, probability, decision theory, cognitive psychology and semantics.

Shortly after graduation, he joined a group of economists headed by John Maynard Keynes. Ramsey wrote only two papers on mathematical economics, but both are still widely cited. In philosophy his inspiration came from George E. Moore, Ludwig Wittgenstein and Bertrand Russell. Moore wrote, "He was an extraordinarily clear thinker: no one could avoid more easily than he the sort of confusions of thought to which even the best philosophers are liable." Then, tragically, in 1930, at the age of 26, Ramsey took ill and died of complications from abdominal surgery.

An irony is attached to the story of how, two years before his death, Ramsey derived his eponymous theory. He came on the central idea while attempting to prove a premise put forward by Russell and Alfred North Whitehead in their masterwork, *Principia Mathematica*. They proposed that all mathematical truths can be deduced from a concise set of axioms. Expanding on their idea, the German mathematician David Hilbert had suggested that there must be a procedure to decide whether or not a given proposition follows from a particular set of axioms. Ramsey showed that there was such a decision procedure for a special case. (A few years later Kurt Gödel, followed by the English mathematician Alan M. Turing and others, showed conclusively that for the general case, there was no such decision procedure.)

Ramsey proved his theorem as a first step in his attempt to demonstrate the special case. As it turned out, he could have accomplished the same task by other means. Ramsey had proved a theorem that was superfluous to an argument, which he could never have proved in the general case.

There matters lay until 1933, when two Hungarian mathematicians, Paul Erdős and George Szekeres, rediscovered Ramsey theory. They are largely responsible for its popularization in the mathematics community. At the time, Erdős was a 19-year-old student at the University of Budapest, and Szekeres had recently earned a degree in chemical engineering from the Technical University of Budapest. They and a group of fellow students would meet almost every Sunday in a park or at school, mainly to discuss mathematics.

At a meeting during the winter of 1933, one of the students, Esther Klein, challenged the group to solve a curious problem: if five points lie in a plane so that no three points form a straight line, prove that four of the points will always form a convex quadrilateral. (The term convex suggests a bulging geometric figure such as a hexagon but not a five-pointed star. More specifically, a polygon is convex if every line segment drawn between its vertices lies inside the polygon.)

After allowing her friends to contemplate the problem, Klein presented a proof [*see illustration at left*]. Erdős and Klein quickly came up with a generalization of the problem. They realized that five of nine points in a plane will always form a convex pentagon. They then offered a new problem: if the number of points that lie in a plane is equal to $1 + 2^{k-2}$ where k is 3 or 4 or 5 and so on, can one always select k points so that they form a convex k-sided polygon?

In a memoir Szekeres recalled the scene: "We soon realized that a simple-minded argument would not do and there was a feeling of excitement that a new type of geometrical problem emerged from our circle." Szekeres eagerly demonstrated that there always exists a number n such that if n points lie in the plane so that no three form a straight line, it is possible to select k points that form a convex k-sided polygon. In other words, given enough points, one can always find a set that will form a particular polygon. In proving this, Szekeres had rediscovered Ramsey's theorem, although no one in the group knew it at the time.

In 1934 Erdős and Szekeres reported their results, but neither they nor anyone else to this day has been able to prove Erdős's conjecture that precisely $n = 1 + 2^{k-2}$ points suffice. Erdős often refers to their joint publication as the "Happy End Paper," because soon after publication Szekeres and Klein married. Erdős became the most prolific mathematician of this century.

Erdős was intrigued by Ramsey's idea that any sufficiently large structure must contain a regular substructure of a given size. But he wondered exactly how large the structure must be to guarantee a certain substructure. So Erdős began work on a version of the party puzzle.

In this version the six people are represented as six points. For convenience, the points are drawn on a plane so that no three are in a line. The points are connected by an edge, which is colored to represent the relationship of the corresponding two people. A red edge means the people mutually know one another, and a blue edge means they do not know one another.

Hence, if three people are mutual acquaintances, the edges between the points will form a red triangle, and if three people are mutual strangers, a blue triangle will be formed. The party puzzle can be rephrased as follows: Is it true that if each edge between six points is arbitrarily colored either red or blue, there will always be either a red or a blue triangle?

The problem that Erdős studied is a general version of this one. He defined a complete network as a number of points that are all connected by edges. He then asked what the smallest complete network is that when arbitrarily colored red and blue will guarantee either a red or a blue complete network of three points. The answer is a complete network of six points. The problem and solution can be more conveniently expressed as follows: the Ramsey number for red three and blue three is equal to six.

But what about the Ramsey number for red five and blue three? In other words, what is the smallest complete network that when arbitrarily colored red and blue will guarantee either a red network of five points or a blue network of three points? The Ramsey number for red five and blue three is 14, which was not proved until 1955 by Robert E. Greenwood of the University of Texas at Austin and Andrew M. Gleason of Harvard University.

Ramsey numbers are notoriously difficult to calculate. The efforts of generations of mathematicians and computers have succeeded in finding only seven Ramsey numbers, which are given in the illustration on page 113.

To express the difficulty of calculating Ramsey numbers, Erdős often tells the following anecdote. Aliens invade the earth and threaten to obliterate it in a year's time unless human beings can find the Ramsey number for red five and blue five. We could marshall the world's best minds and fastest computers, and within a year we could probably calculate the value. If the aliens demanded the Ramsey number for red six and blue six, however, we would have no choice but to launch a preemptive attack.

Erdős did find a way, however, to get some idea of how large a Ramsey number must be. What if he could find a red and blue coloring of a large complete network that did *not* form either a red or a blue network of three points? Such a coloring of the complete network of five points is shown on page 113. It follows that the Ramsey number for red three and blue three must be greater than five. Five is a lower bound for that Ramsey number.

In 1947 Erdős proposed an unusual method for finding the lower bound of any Ramsey number: flip a coin. He performed a thought experiment in which each edge of a complete network of, say, a million points was colored according to the flip of a fair coin. The edge would be red for "tails" or blue for "heads." He then tried to show that the Ramsey number for, say, red 34 and blue 34 is greater than a million. The experiment was a success if there was neither a red nor a blue network of 34 points.

How could he ensure success? Any 34 points must have 561 edges between them. If the first coin flip specifies blue for the first edge, then the next 560 flips must also specify blue in order to produce a blue network. The probability of this occurring is one half to the power of 561. A red network has an equal probability of occurring so the total probability is double, or approximately 2.6×10^{-169}.

Now the number of sets of 34 points in a million points is equal to $(1{,}000{,}000 \times 999{,}999 \times \ldots \times 999{,}967)$ divided by $(34 \times 33 \times \ldots \times 2 \times 1)$, which is about 3.4×10^{165}. Out of all possible complete networks of 34 points, therefore, one would expect $(3.4 \times 10^{165}) \times (2.6 \times 10^{-169})$, or roughly .001, to be monochromatic. Thus, 99.9 percent of the time the thought experiment is a success with no 34-point monochromatic sets created.

Erdős then employed a subtle reductio ad absurdum. He hypothesized that no coloring scheme was successful. Then the thought experiment would have zero probability of success, which he knew was not the case. The hypothesis must be incorrect; there *must* be a successful coloring (not just with 99.9 percent certainty but with absolute certainty). The existence of a coloring implies that one million is a lower bound for the Ramsey number for red 34 and blue 34.

This argument, which is known as the probabilistic method, has given the best available lower bounds for Ramsey numbers. The probabilistic method gives no clues, however, concerning the actual "construction" of the desired coloring. In an attempt to produce these colorings, workers have applied a range of techniques from number theory, set theory and other branches of mathematics. The results, although interesting, do not yet approach the bounds obtained by flipping a coin.

Although much of the early work in Ramsey theory focused on sets of points and lines, many of the first problems involved sets of numbers. In fact, the Dutch mathematician Bartel L. van der Waerden began solving such problems before Ramsey proved his theorem.

In 1926 van der Waerden learned

Ramsey Theory and Arithmetic Progressions

An arithmetic progression is a sequence of numbers in which the difference between successive terms remains constant. For instance, 7, 10, 13, 16 is an arithmetic progression in which the difference between successive terms is three. The following statement about arithmetic progressions follows from Ramsey theory: if each number 1 through 9 is colored either red or blue, either three red or three blue numbers will form an arithmetic progression.

To prove the conjecture, we might check every one of the 512 ways that nine numbers can be colored. Yet we can prove the conjecture by considering only two cases. We start with the case in which 4 and 6 are the same color, say, blue.

1 2 3 4 5 6 7 8 9

To avoid the red arithmetic progression 4, 5, 6, we color 5 red.

1 2 3 4 5 6 7 8 9

To avoid the blue arithmetic progressions 2, 4, 6 and 4, 6, 8, we color 2 and 8 red.

1 2 3 4 5 6 7 8 9

But that leaves the red arithmetic progression 2, 5, 8. Hence, if 4 and 6 are the same color, there is always either a red or a blue arithmetic progression. Next we consider the case in which 4 and 6 are different colors. We can color 5 either red or blue without forming an arithmetic progression, so we arbitrarily choose to color 5 red.

1 2 3 4 5 6 7 8 9

We continue to color the numbers as follows:

3 to avoid **3 4 5**
7 to avoid **5 7 9**
2 to avoid **2 5 8**
9 to avoid **3 6 9**
8 to avoid **6 7 8**
1 to avoid **1 2 3**

This coloring yields the sequence

1 2 3 4 5 6 7 8 9

Yet we are still left with the red arithmetic progression 1, 5, 9. Therefore, regardless of whether 4 and 6 are the same color or different, there is always either a red or a blue arithmetic progression.

about a curious problem involving arithmetic progressions. As the phrase implies, an arithmetic progression is a sequence of numbers in which the difference between successive terms remains constant. For example, the sequence 3, 5, 7 is a three-term arithmetic progression in which the difference between successive terms is two. A special case of the problem that caught van der Waerden's interest was the following: If each integer from 1 through 9 is printed on a page in one of two colors, either red or blue, is it always true that either red three or blue three numbers will form an arithmetic progression? The answer is given in the box on the preceding page.

Van der Waerden challenged himself with the following generalization: if n is a sufficiently large integer and if each integer from 1 through n is printed arbitrarily in one of two colors, then there is always a monochromatic arithmetic progression with a certain number of terms. One can think of this statement as Ramsey's theorem for arithmetic progressions, although it is generally known as van der Waerden's theorem.

Van der Waerden enlisted the help of his colleagues Emil Artin and Otto Schreier. He later wrote: "We went into Artin's office in the Mathematics Department of the University of Hamburg, and tried to find a proof. We drew some diagrams on the blackboard. We had what the Germans call 'Einfälle': sudden ideas that flash into one's mind. Several times such new ideas gave the discussion a new turn, and one of the ideas finally led to the solution." It turned out, however, that van der Waerden could not demonstrate the result for two colors without simultaneously demonstrating it for an arbitrary number of colors.

For his proof, van der Waerden employed a special form of mathematical induction. The usual form, known as single induction, has two steps. First, one shows that the result holds for some small value, such as two. Second, one proves that if the result holds for any value then it holds for the next larger value. This implies that it holds for three, four and so on. The results fall like an infinite set of dominoes.

Van der Waerden employed a more subtle, double induction to prove Ramsey's theorem for arithmetic progressions. He assumed that for any fixed number of colors there was a number n such that if each integer from 1 through n were printed in one of these colors, then there would be a monochromatic arithmetic progression of, say, 10 terms. He could then deduce that for any fixed number of colors, there was a number m such that if each integer from 1 through m were printed in one of these colors, then there would be a monochromatic arithmetic progression of 11 terms. In general, he showed that knowing the result for k terms and all numbers of colors implies the result for $k+1$ terms and all numbers of colors.

Once van der Waerden had arrived at that stage in the proof, he had only to demonstrate that the result does hold for some small value of k. If the number of integers is one more than the number of colors, then there are always two integers that have the same color. These two integers form an arithmetic progression of two terms. Thus, if the number of integers is one more than the number of colors, there is always a monochromatic arithmetic progression of two terms. The infinite set of dominoes with two terms now pushes over the infinite set with three terms, which in turn pushes over the infinite set with four terms and so on [*see box on this page*].

Having proved Ramsey's theorem for arithmetic progressions, van der Waerden applied his knowledge to the following problem: What is the smallest value of n that will guarantee a monochromatic arithmetic progression of, say, 10 terms if each integer from 1 through n is printed arbitrarily in one of two colors? The best answer that van der Waerden could find was so large that it cannot be written in conventional notation. It was larger than a billion, larger than 10 to the power of a billion.

In fact, in order to express his result, mathematicians rely on a sequence of functions known as the Ackermann hierarchy. The first function in the hierarchy is simply called DOUBLE (x). As the name implies, the function doubles the number x. Therefore, DOUBLE(1) equals 2, and DOUBLE(50) equals 100. The second function, EXPONENT (x), can be expressed as 2 to the power of x, and therefore, EXPONENT(3) equals 8. We also can describe EXPONENT in terms of DOUBLE. To find EXPONENT (3), for instance, we double 1, then double the result, then double the result again. In fact, each function in the Ackermann hierarchy is defined in terms of its predecessor.

Hence, the third function in the hierarchy, TOWER (x), can be expressed using EXPONENT. TOWER(3), for example, is 2 to the power of 2 to power of 2, which equals 2 to the power of 4, or 16. TOWER (x) is sometimes written as a tower of exponents

$$2^{2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}}$$

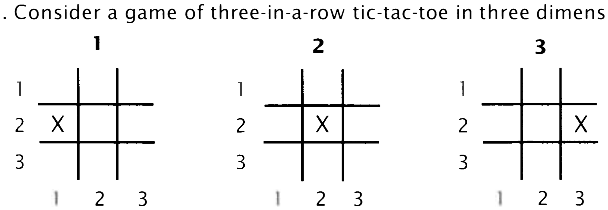
where there are x number of 2's in the

Ramsey Theory and Tic-Tac-Toe

In 1926 Bartel L. van der Waerden proved that if n is a sufficiently large integer and if each integer 1 through n is printed arbitrarily in one of two colors, then there is always a monochromatic progression with a certain number of terms. In 1963 Alfred W. Hales and Robert I. Jewett found what has proved to be the essence of van der Waerden's theorem while investigating the game tic-tac-toe. Although the classic three-on-a-side tic-tac-toe can get tiresome, four-on-a-side tic-tac-toe in three dimensions is quite challenging. The board for the three-dimensional game has 64 cells arranged in a cube. Players alternately fill the cells with naughts and crosses until one player wins by occupying four cells in a line. Two- and three-dimensional tic-tac-toe sometimes end in a tie. But what about higher-dimensional games? Is a player ever guaranteed to win in some n-dimensional k-in-a-row version of tic-tac-toe?

Hales and Jewett showed that if the dimension n is large enough, one can always find a k-in-a-row version that *never* ends in a tie. For instance, no matter how the naughts and crosses are arranged on a three-dimensional three-in-a-row version, either three naughts will occupy a line or three crosses will occupy a line.

Van der Waerden's theorem can be derived from the Hales-Jewett result by employing a transformation that converts lines of tic-tac-toe into arithmetic progressions. Consider a game of three-in-a-row tic-tac-toe in three dimensions.



The coordinates for the crosses in this winning combination are 121, 222 and 323, which form an arithmetic progression. It can be shown that any winning combination transformed by this method will yield an arithmetic progression.

tower. Yet even the TOWER(x) function does not increase rapidly enough to describe van der Waerden's result.

The next function, informally known as WOW(x), is found by beginning at 1 and applying the TOWER function x times. Therefore,

WOW(1) = TOWER(1) = 2
WOW(2) = TOWER(2) = 4
WOW(3) = TOWER(4) = 65,536

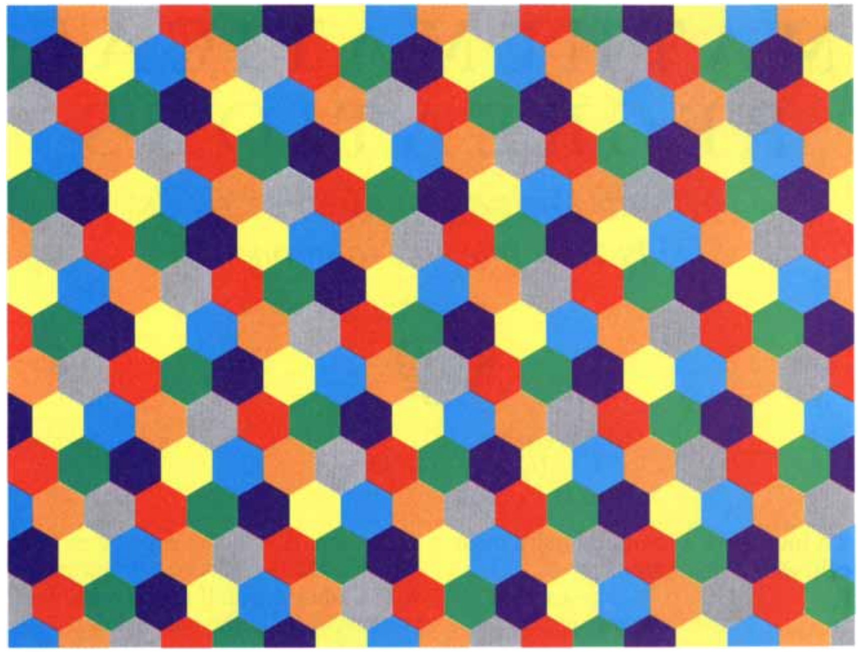
To find WOW(4), we need to compute TOWER(65,536). To do this, we begin at 1 and apply EXPONENT 65,536 times. Even applying EXPONENT just five times gives $2^{65,536}$, a number whose digits would fill two pages of this magazine. In fact, if a number filled every page of every book and every memory bank of every computer, it would still be incomparable to WOW(4).

Yet to give van der Waerden's result, we must define a function that grows even faster. The function ACKERMANN (x) is defined by the sequence DOUBLE (1), EXPONENT(2), TOWER(3), WOW(4) and so on. ACKERMANN(x) eventually dominates all of the functions of the hierarchy. Van der Waerden's proof gave the following quantitative result: if the integers 1, 2,..., ACKERMANN(k) are two-colored, then there is always a monochromatic arithmetic progression of k terms.

It seemed preposterous that such enormous numbers could come out of such an innocent statement involving only arithmetic progressions. Over the years many mathematicians attempted to improve the proof of van der Waerden. As the failures mounted, the idea began to gain support that a double induction and the corresponding ACKERMANN function were necessary features in any proof of van der Waerden's theorem. Increasingly, logicians tried to supply arguments that this indeed was so.

In 1987, however, Israeli logician Saharon Shelah of the Hebrew University in Jerusalem achieved a major breakthrough. Shelah is widely regarded as one of the most powerful problem solvers in modern mathematics. He broke through the ACKERMANN barrier to show the following: if the integers 1, 2,..., WOW(k) are two-colored, then there must always be a monochromatic arithmetic progression of k terms.

Despite his background, Shelah's proof uses no tools from mathematical logic whatsoever. His proof employs only elementary (but highly ingenious) mathematical ideas. Written out in full, the proof is perhaps four pages long, and most experts consider it clearer than van der Waerden's original proof.



CONCEPTS in Ramsey theory can be applied to problems in geometry such as this hexagon puzzle. If the sides of the hexagons are all .45 unit long (the unit is arbitrary), then two points within a hexagon are at most .9 unit apart. Each hexagon is shaded with one of seven colors so that no two hexagons of the same color are less than 1.19 units apart. No two points of the same color are precisely one unit apart. No one has been able to determine whether or not the plane can be shaded with six colors so that no two points of the same color are precisely one unit apart.

Most important, he avoids the double induction. He fixes the number of colors at two (or any particular number) and then proves a simple induction: if the result holds for progressions of k terms then it also holds for progressions of $(k+1)$ terms.

Mathematicians are now poring over Shelah's proof to see if it can in fact be further improved to give a TOWER or even an EXPONENT function for van der Waerden's theorem. One of us (Graham) has offered a reward of $1,000 for a proof (or disproof) that for each number k, if the numbers 1, 2,..., TOWER(k) are two-colored, then a monochromatic arithmetic progression of k terms must be formed.

The work of Ramsey, Erdős, van der Waerden and many others established the fundamentals of Ramsey theory. Yet workers have only begun to explore the implications of the theory. It suggests that much of the essential structure of mathematics consists of extremely large numbers and sets, objects so huge that they are difficult to express, much less understand.

As we learn to handle these large numbers, we may find mathematical relations that help engineers to build large communications networks or help scientists to recognize patterns in large-scale physical systems. Today we can easily recognize the constellations in the night sky as a consequence of Ramsey theory. What patterns may we find in sets that are ACKERMANN(9) times larger?

FURTHER READING

COMBINATORIAL RELATIONS AND CHROMATIC GRAPHS. A. M. Gleason and R. E. Greenwood in *Canadian Journal of Mathematics,* Vol. 7, No. 1, pages 1-7; 1955.

HOW THE PROOF OF BAUDET'S CONJECTURE WAS FOUND. B. L. van der Waerden in *Studies in Pure Mathematics.* Edited by L. Mirsky. Academic Press, Inc., 1971.

PAUL ERDŐS: THE ART OF COUNTING: SELECTED WRITINGS. Edited by Joel Spencer. The MIT Press, 1973.

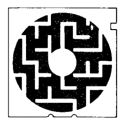
THE MAN WHO LOVES ONLY NUMBERS. Paul Hoffman in *Atlantic Monthly,* Vol. 260, No. 5, pages 60-74; November, 1987.

NUMBERS IN RAMSEY THEORY, IN SURVEYS AND IN COMBINATORICS. R. L. Graham and V. Rödl. London Mathematical Society Lecture Notes Series, No. 123, pages 111-153; 1987.

RAMSEY THEORY. Ronald L. Graham, Bruce L. Rothschild and Joel H. Spencer. Second Edition. John Wiley & Sons, Inc., 1990.

MATHEMATICAL RECREATIONS

An odd journey along even roads leads to home in Golygon City



by A. K. Dewdney

"A journey of a thousand miles must begin with a single step."

—LAO-TZU,
The Way of Lao-tzu

Allow me to start you on a journey in Golygon City. You can take a similar trip in New York, Kyoto or almost any large city whose streets form a grid of squares. Here are your directions. Stroll down a city block, and at the end turn left or right. Walk two more blocks, turn left or right, then walk another three blocks, turn once more and so on. Each time you turn, you must walk straight one block farther than before. If after a number of turns you arrive at your starting point, you have traced a golygon. If you do not need the physical exercise, you can easily simulate the journey by moving a pencil along a piece of graph paper with a square grid. If you become lost, you may refer to the map below.

A golygon consists of straight-line segments that have lengths (measured in miles, meters or whatever unit you prefer) of one, two, three and so on, up to some finite number. Every segment connects at a right angle to the segment that is one unit larger—except the longest segment, which meets the shortest segment at a right angle. Golygons are more than just a geometric curiosity. They have inspired some delightful puzzles as well as some intriguing problems for research. What better way to develop insight into the research process than to take a recreational journey?

A map of Golygon City

It seems that golygons were first conceived by Lee Sallows, a redoubtable engineer of the Catholic University Nijmegen in the Netherlands. I last featured Sallows's work in October, 1984, when I described his search for pangrams, sentences containing each letter of the alphabet. Since then Sallows has invented many new recreations but none so engaging, as far as this department is concerned, as golygons.

In the fall of 1988 Sallows began his search for golygons. It did not take him long to find an eight-sided golygon. Yet he could find no such objects that had fewer than eight sides. Nor did he discover a golygon with nine sides, nor 10, nor 11,...until at last he stumbled on a 16-sided golygon.

Wondering whether he had missed any, Sallows wrote a computer program to automate the search. It spewed out no less than 28 different 16-sided golygons [*see illustration on opposite page*] before moving on to higher orders. The program did not find golygons with from 17 up to 23 sides, but it generated numerous 24-sided golygons—2,108 to be exact.

By then Sallows had a hunch that the number of sides in a golygon must be a multiple of eight. Yet his program, which was already laboring in the 20's, was unequal to the task of discovering any 32-sided golygons. Frustrated, Sallows appealed to Martin Gardner, the supreme authority in matters of recreational mathematics. Could Gardner prove that the number of sides in a golygon must be a multiple of eight?

We can creep up on Gardner's proof by starting with the simpler task of proving that the number of sides must be a multiple of two. To explain, allow me to take you back to Golygon City. We start our journey by walking one block to the north. (I choose north only for convenience.) Hence, the first side of the golygon will run north for one block. Next we can turn either left or right, that is, west or east. Therefore, the second side of the golygon will run either two blocks to the west or two blocks to the east. It follows that all odd sides in the golygon (the first, third, fifth, etc.) are an odd number of blocks long and run north or south; all even sides (the second, fourth, sixth, etc.) are an even number of blocks long and run east or west. Because the last side of the golygon meets the first at right angles, the last side must run east or west. Therefore, the last block is an even side of the golygon, and the number of sides in a golygon must be a multiple of two.

To show that the number of sides must be a multiple of four, we begin by calculating the total distance we traveled to the north of the starting point. To do this, we simply add the number of blocks we walked to the north and subtract the number of blocks we traveled to the south. (A net negative distance to the north simply means that we are south of the starting point.)

Because all north and south sides are an odd number of blocks long, we are essentially adding and subtracting consecutive odd numbers. It takes only a little tinkering to convince ourselves that the result is always even when we add or subtract an even number of odd lengths—for example, $1+3-5+7=6$. By the same token, an odd number of odd lengths always gives an odd sum. Therefore, the distance walked north is an even number of blocks if and only if we have walked along an even number of north and south sides.

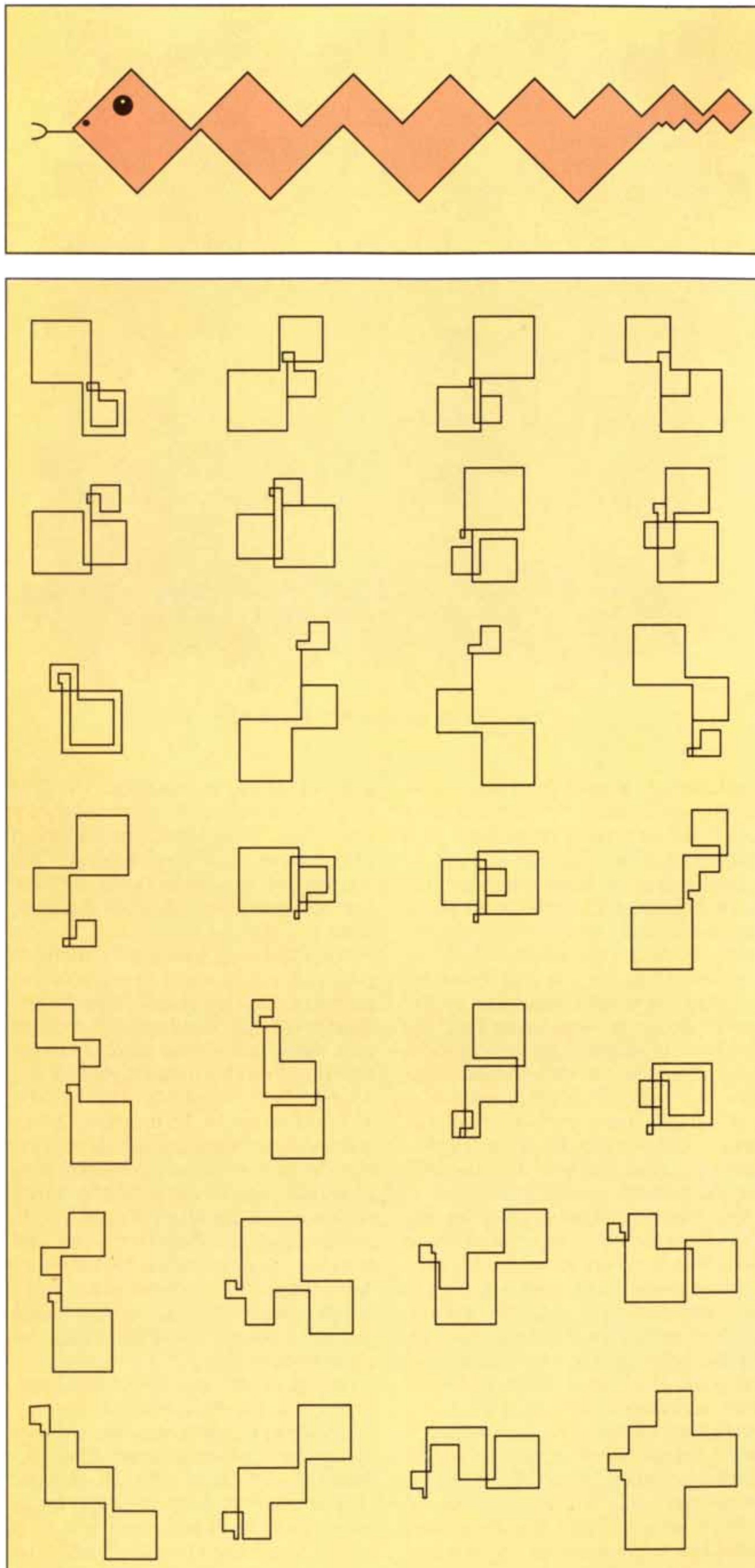
Now if we walk from the starting point around Golygon City and return, the distance to the north of the starting point equals zero. Because zero is an even number, the golygon must have an even number of north and south sides. The total number of sides is twice the number of north and south sides, because for every north or south side there must be an east or west side. Therefore, the number of sides in a golygon is a multiple of four.

How on earth did Gardner show that the number of sides must be a multiple of eight? Let us accompany the master on his journey. We already know that the number of north and south sides in the golygon is even. We also know that the distance traveled to the north must equal the distance traveled to the south. Therefore, the distance north must equal half the total distance that we have walked in both directions, north and south. Hence, the total distance is equal to the sum of an even number of consecutive odd numbers. For instance,

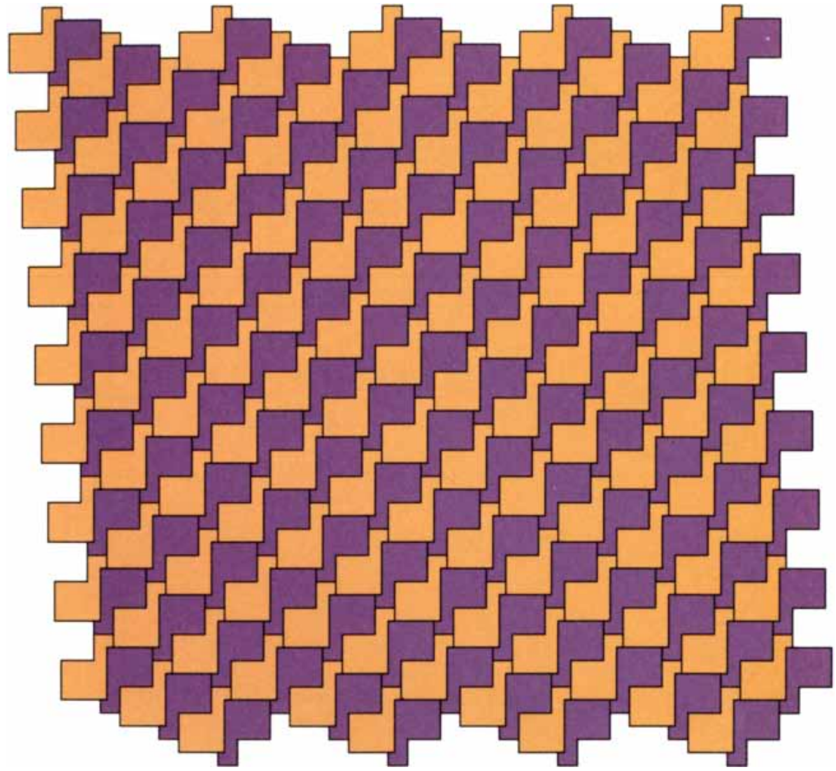
$$1+3=4$$
$$1+3+5+7=16$$
$$1+3+5+7+9+11=36$$
$$1+3+5+7+9+11+13+15=64$$

The result always seems to be a multiple of four.

If the total distance north and south is equal to a multiple of four, then half that distance must be a multiple of two. Therefore, the distance we traveled north must be a multiple of two, and the number of north sides must also be a multiple of two because all north sides are an odd number of blocks long. It follows that the number of south sides is a multiple of two and that the total number of north and



One golygon snake that has 32 sides and all 28 golygons that have 16 sides



The smallest golygon tiles the plane

south sides is a multiple of four. Finally, after we include the east and west sides, the total number of sides in a golygon must be a multiple of eight.

Gardner's proof may seem rather peculiar because odd numbers played a major role and even numbers hardly entered into the discussion. The same conclusion can be reached, however, by making a similar argument on behalf of the even lengths. In fact, the two kinds of lengths play independent roles in the construction of any golygon. One may obtain, for example, a new golygon from two others of the same size by using the even lengths from the first golygon and the odd lengths from the second.

Our recreational path has so far taken us from the question of existence to the question of conditions on that existence; the number of sides in a golygon must definitely be a multiple of eight. For convenience I will call this theorem the $8k$ condition for two-dimensional golygons. It is what mathematicians call a necessary condition. If a golygon exists, the number of its sides is necessarily a multiple of eight. But is this condition sufficient? In other words, if we merely ask that the number of sides, n, be a multiple of eight, is that sufficient to guarantee the existence of a golygon?

We can find a golygon that has eight times any number of sides by applying a set of simple instructions. The first step is to decide how large a value of n one wants. To illustrate the method, I will choose a 16-sided golygon. The second step is to write down consecutive numbers from 1 to n, in this case, from 1 to 16.

The third step would have us place plus and minus signs in front of the numbers. The first quarter and the last quarter of the numbers will receive plus signs, and all the numbers in between will receive minus signs: +1, +2, +3, +4, -5, -6, -7, -8, -9, -10, -11, -12, +13, +14, +15, +16. To translate this sequence of numbers into a tour of Golygon City, you need only know that positive odd numbers run to the north, negative odd numbers to the south, positive even numbers to the east and negative even numbers to the west. The golygon that corresponds to these numbers looks like a snake. A 32-sided "snake" is shown in the illustration on the preceding page.

Having settled the question of sufficiency of the $8k$ condition, Gardner and Sallows began to wonder just how many golygons should exist. They had found one of order eight, 28 of order 16 and 2,108 of order 24. How many would there be of higher orders? Such questions are most frequently asked by mathematicians who specialize in combinatorics, the study of discrete mathematical objects such as sets and graphs [see "Ramsey Theory," by Ronald L. Graham and Joel H. Spencer, page 112]. Gardner decided to check with some experts. He wrote letters to computer scientist Donald E. Knuth of Stanford University and to mathematician Richard K. Guy of the University of Calgary in Alberta. Could either of them assist in the matter of enumerating golygons?

Before long, Knuth had written a computer program that counted all the golygons up to 64 sides. The total number of 64-sided golygons, for instance, is 127,674,038,970,623. Unfortunately, Knuth's program became unwieldy at larger sizes.

Meanwhile Guy was able to develop a formula that expressed the approximate number of golygons for each possible value of k, where k equals the number of sides divided by eight. The formula consists of a fraction, the principal parts of which are 2 raised to the power $8k-6$ in the numerator and k cubed in the denominator. Because, as k increases, the exponential term increases much more rapidly than the cubed term, the growth in the number of golygons is basically exponential. Guy's formula has the further important property of being asymptotic, that is, it gives a value that, expressed as a percentage of the true value, comes increasingly close to 100 percent as k gets larger and larger.

Sallows, Gardner, Knuth and Guy decided to report their various results in a paper. The word "golygon" does not appear in the title. Instead it is called "Serial Isogons of 90 Degrees," a name that rings more solemnly than "golygons." The reference to 90 degrees implies that other kinds of golygons might be assembled. Indeed, some of the authors have looked at golygons in which every angle is not 90 degrees but 60 or 120 degrees. To construct such golygons, readers may need hexagonal-grid paper, which can be found in most graphics-supply stores.

Would this have been a good place to stop? Research (like recreation) never stops. Often, as research in one question progresses, other questions suggest themselves. Sometimes it is very hard to decide just what question to work on.

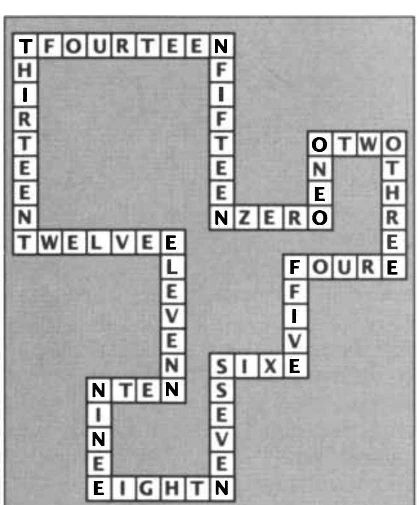
Research problems are no sooner solved than the recreationist in us thinks of new avenues of inquiry: Can we find prime-sided golygons? The lengths of consecutive sides here increase by the sequence of odd primes: 1, 3, 5, 7, 11, 13 and so on.

Another offshoot of golygons deserves mention. Readers are urged to look at the eight-sided golygon once again. It is unique, the only member of its class. Eight-sided golygons will fit together very nicely to make tiling patterns. They will, as the professionals say, "tile the plane." The illustration on the opposite page shows an attractive tiling.

Properly speaking, the region bounded by an eight-sided golygon should be called a polyomino. Not all golygons produce polyominoes because some golygons intersect themselves in their meandering march around the plane. In fact, polyominid golygons (a formidable expression by which readers can lose friends and influence people) are probably increasingly rare in relative terms as *n* gets large. Only four of the 29 golygons shown on page 119 are the boundaries of polyominoes. None of these appears to tile the plane.

But in the matter of tiling, Sallows has a challenge for readers: suppose the eight-sided golygon represents the boundary of one's kitchen floor. Given 13 L-shaped tiles, can readers cover the kitchen floor exactly with these eccentric tiles? For the purpose of this question, the floor may be thought of as divided into 52 squares, like the squares of the special paper on which this enquiry started. Each square is one step on a side. The L-shaped tiles consist of three squares in a row and an additional square comprising the foot of the L. Anyone who finds the task impossible must, of course, prove it to be so.

I close with some final words from Sallows, the engineer who started all this. The "words" are ZERO, ONE, TWO and so on, up to FIFTEEN. They can be arranged in a special golygon all their own in which the letters of these words determine the sides of what can only be called a logolygon [*see illustration below*].



A logolygon created by Sallow

Reader Response

The Emperor's New Mind *and wireworld*

The Pandora's box of minds, machines and metaphysics, which I opened last December, gave me a ringside seat to a battle between titans. In one corner stood Roger Penrose, whose recent book, *The Emperor's New Mind,* has given new heart to those who feel that the human brain is more than a computer. In the opposing corner were members of the artificial-intelligence and computing community who find nothing particularly new or incisive in the book.

The philosopher John R. Searle was called to defend Penrose [see "Is the Brain's Mind a Computer Program?" SCIENTIFIC AMERICAN, January]. Searle put forward the "Chinese room" argument: a computer program that can read and write Chinese does not understand Chinese anymore than a person who imitates the action of the program.

Among the champions of the artificial-intelligence community was Hans P. Moravec, a robotics expert at Carnegie-Mellon University, who sent me an open letter that he had mailed to Penrose. Moravec, whose gung ho attitude is abundantly evident in his book, *Mind Children,* calls Penrose's argument "wildly wrongheaded."

Would Penrose grant the same intelligence to a certain robot as he would to the dog that looks up at him with soft brown eyes and whines? The robot Moravec has in mind is at present more imaginary than real. It says, "*Please,* Roger, it bothers me that you don't think of me as a real person. What can I do to convince you? I am aware of you, and I am aware of myself. And I tell you, your reaction is almost unbearable." The robot's reaction is the result of a special program designed to convince the master that some caring is in order—perhaps like the whining of a dog.

I can only begin to sample the many readers who responded to this column. Jef Raskin of Pacifica, Calif., states that Searle's Chinese room argument is "tantamount to defining intelligence by stating, 'Intelligence is a nonalgorithmic process that...'" Paul Preuss, a computer aficionado and noted science-fiction writer from San Francisco, claims that the Chinese room is not a true thought experiment. It should be performed, Preuss says, "by Dr. Searle himself. If he succeeds in convincing a Chinese speaker that he, inside the room, really can read and write Chinese, he should then report on his subjective feelings as to whether or not he has learned Chinese in the process."

The famous incompleteness theorem of German mathematician Kurt Gödel is often cited as evidence that there are thoughts beyond the reach of algorithms. My query as to whether the theorem enjoys extra-algorithmic status brought a hail of "nays" from a variety of seeming experts, including Greg Ray, who studies the logic and methodology of science at the University of California at Berkeley, and from Robert Wilensky, director of the Berkeley Artificial Intelligence Project. Peter Kugel of the department of computer science at Boston College opined that Gödel's proof is itself an algorithm.

The final "Computer Recreations" column, appearing this past January, featured two cellular-automaton software packages created by Brian Silverman of Montreal and by Rudy Rucker of Los Gatos, Calif.

Wireworld, a two-dimensional cellular space in which home experimenters can set up simple computing circuits, inspired a number of readers to experiment on their own. Nyles Heise of Houston, Tex., shared Silverman's enthusiasm for simulated circuits to the point of writing a program that duplicated many features of wireworld. Written in the compact language APL in a single evening, the program enabled Heise to construct memory elements that surpassed my own in speed. Using wireworld, I constructed a circular path that would hold a single live pulse that could be sampled every 13 ticks of the cellular clock. Heise's most efficient memory element has a cycle time of only four ticks. To build such a speedy device, however, he had to invent a new way to transmit pulses along wires.

Other creative souls included Oliver Springauf of Würzburg, West Germany, and Davide P. Cervone of Providence, R.I. Following my more conservative technology, Springauf succeeded in improving my memory loop from 13 to 12 ticks. Cervone built an amazing variety of devices, including a four-bit adder. The layouts sent in by all readers looked like nothing on earth, so alien is the notion of cellular computing to the continuous physics employed by actual computing microcomponents.

FURTHER READING

SERIAL ISOGONS OF 90 DEGREES. Lee Sallows, Martin Gardner, Richard Guy and Donald Knuth in *Mathematics Magazine*. Edited by G. L. Alexanderson (in press).

BOOKS

A chemical classic, creative minds, sexual liaisons, an extraordinary design team

by Philip Morrison

THE MERCK INDEX: AN ENCYCLOPEDIA OF CHEMICALS, DRUGS, AND BIOLOGICALS, edited by Susan Budavari, Maryadele J. O'Neil, Ann Smith and Patricia E. Heckelman. Eleventh edition. Merck & Co., Inc., 1989 ($35).

In 1889 a chemical manufacturing firm published an annotated catalogue of drugs and chemicals available for sale. Over the decades their catalogue, now a book, has grown edition after edition; this 300-page centennial version is fifteenfold longer than the first and physically resembles the bulky dictionary that is found on so many readers' shelves. Today the *Index* is viewed as a standard reference work, a concise encyclopedia of significant substances—animal, vegetable and mineral—each of which is described succinctly in terms of its nature, history and uses. Although the book covers much more than just medicine, it retains "a strong medical character."

As is the case for a big dictionary, it is the sheer number of words that has the most immediate impact on a general reader. The grand cross-index lists about 80,000 names, including all the synonyms in use for the 10,100 alphabetized entries that occupy the bulk of the volume. A million or more compounds are now known to the chemical world; selection of one among 100 is a process too complex to be neatly algorithmic. Prepared foodstuffs and metallic alloys cannot be found, for example, nor can one expect to find listings for Boston cream pie, Alka-Seltzer or bronze. Some 6,000 entries are accompanied by spatial structural formulae (the rings and chains of organic substances); another 2,000 bear symbolic formulae for the inorganics and simpler organics of the laboratory world.

Chemical elements are few in number and superbly defined, and so the *Index* generously includes all of them, even unstable and unused astatine, which has little commercial value. Only from 100 to 200 minerals are listed (out of the thousands that are formally recognized), but the list includes the more familiar ones: galena, realgar, diamond and lapis lazuli. Spinel and pyrite have been dropped in the present edition.

The reader can also probe a layer laid down long ago: the apothecary's shop with its ranked bottles and dusty drawers, their contents prescribed according to law. Remedies include the balm of Gilead (not to be confused with balsam Canada), the flowers of antimony, mandrake root and yogurt.

Not so different in age and ambience are the stocks of the perfumer and the many flavors of the spice dealer. Eight pages are devoted to 50 distinctive oils, including the oils of bergamot, cajaput, hyssop, wormwood and ylang-ylang. One can also read about ambergris, beeswax and myrrh; simpler wintergreen and vanilla no longer reside in this section; they now appear as carvone and vanillin (mainly) in the world of synthetic formulae.

The more or less modern chemicals of factory and laboratory, which include products as well as reagents, are not very numerous. Listed are alizarin dyes (red, blue, green and yellow), ethylene, gallium arsenide, the ferrites found in portable transistors, blue ozone, Teflon, fishy and ammoniacal trimethylamine and vitamins, established and tentative ones side by side. The material is, for the most part, up-to-date.

The heart of the book is its pharmaceutical coverage. Structures bedeck most of the pages in this category—half a dozen to a spread—from aspirin to zidovudine (the latter, which is a reverse transcriptase inhibitor known as AZT, is a recent weapon in the fight against AIDS). The jagged names of geometric precision are also present: the alphanumeric strings, like 2-(4-isobutylphenyl)-propionic acid, that snake so mysteriously across the fine print. They are matched with trade names familiar to most Americans through television commercials. Ibuprofen (the common name of the anti-inflammatory compound spelled out above) is sold under no less than 59 brand names; if not quite overkill, surely this is overhelp! The newest drugs, which include interferons, hormones, monoclonal antibodies and other products of biotechnology, are as yet skimpily present; they surely will flower in later editions.

Controlled substances, both natural and synthetic, are gloomily numerous; there is little hope reading down the list that the problem of American drug addiction, whose roots can be found high in the green fields of the South American altiplano, will be readily solved. Many auxiliary tables are provided, including one that is devoted to the various drugs used in today's cancer chemotherapy; 20 pages of company codes, names and addresses are also provided.

The volume is good for browsing; it is also more or less a necessity for those seriously concerned about the contents of labeled bottles that are neither food nor drink; indeed, attentive people have welcomed its forebears for a long time. Any reader will admire the current *Index*, an impressive testimonial to the growth of chemical sciences, a bargain book for writers, teachers and libraries and a ready tool that can be consulted by modem; an on-line service is now provided by four computer data bases!

CREATIVE PEOPLE AT WORK: TWELVE COGNITIVE CASE STUDIES, edited by Doris B. Wallace and Howard E. Gruber. Oxford University Press, 1989 ($29.95).

This attempt to grasp "how creative people do what they do" provides a detailed and fascinating look at the lives of individuals of unusual creative distinction. Each of the dozen authors has selected one of eight scientists, from Albert Einstein to Jean Piaget, or one of four artists, from William Wordsworth to the contemporary sculptor Melissa Zink, in order to find lifelong themes within their well-documented careers.

The editors' preface sharpens the point of the authors' collective effort. A person who is the subject of too much psychological research, they hold, becomes little more than an object, an unperson called *S*. Instead they define their subject as "a person to be reckoned with, a full-blooded force in the real world and a long-term resident in the investigator's mind."

Creativity is more likely to be unique

than reproducible, yet the task set forth in these pages is not mere description but scientific understanding. No grand epiphany is expected to unite the creative prowess of Albert Einstein, William James and Anaïs Nin. Instead most of the volume's contributors have sought to establish an enduring "network of enterprise," a set of evolving personal projects that together entwine a lifetime of internal and contextual change yet remain recognizably directed toward a goal.

The lives of these individuals might be cut apart, to be patched into coherent, perhaps surprisingly parallel, accounts of, say, their social influences, early childhood experiences or even the moment when the cry "Eureka!" sounded in the bathtub or omnibus. Or they can be examined as they are here, each life woven into one textured narrative that describes the ebb and flow of failure and fruitfulness. There is no way to summarize the dozen accounts presented, most of them resonant with events that many readers will recognize, each one tentative, yet in its own way bold. I shall display a few samples in lieu of the indispensable whole.

One essay is explicitly comparative: Frederic L. Holmes of Yale University examines two chemists, Antoine Lavoisier and Hans Krebs, and makes one unexpected factual finding. Krebs was able to spend five times as many hours per week on research in the 1930's as Lavoisier could in the 1780's under the reign of Louis XVI, and the results Krebs obtained with the help of modern techniques were individually far more reliable than those obtained by Lavoisier. All the same, Holmes concludes that both men followed paths that were equally demanding and sequential.

ALBERT EINSTEIN was employed as a patent clerk in Bern, Switzerland, from 1902 until 1907; the period is said to have been the most creative of his life.

Lavoisier pondered respiration for 17 years before finally concluding that oxygen is taken in and carbon dioxide is given off during the overall process. At first he thought the redness of oxygenated blood was analogous to the red oxidized calx of metallic mercury and simply ignored the presence of carbon dioxide in exhaled breath. By the end of his career he was able to draw parallels between respiration and the combustion of charcoal in air. Although his experiments along the way were few, they were at once difficult and deeply insightful.

In contrast, Krebs's experiments were many, deft and varied. He improvised along several lines of biochemical inquiry, although in retrospect his career path looks strikingly continuous. Unlike Lavoisier, Krebs spent only a few weeks in a quandary over his first celebrated result. At that time he wondered how ornithine, an invariable input for the production of urea from ammonia, could also be a product of the same reaction. He was puzzled by the contradiction, and by his own account the explanation did not come to him in a flash. Instead the realization that the product gives rise to its own source material through a second reaction came to him via a series of small insights, not one of them brilliant enough to have left a lasting trace in his memory. The supposed "single flash of insight" that led to the four-spoke urea cycle and then beyond it to the famous Krebs cycle of oxidative reactions was in fact a protracted and blustery summer storm.

Another remarkable history is that of the organic chemist Robert Burns Woodward, who is sensitively presented by his daughter, Crystal. Before he was 10 years old, Woodward (like many another kid in 1927) had built a chemistry laboratory in the basement of his home in Quincy, Mass. Legend holds that by the age of 12 he had performed all the experiments in a classic German organic-chemistry manual; his daughter confirms that the legend is "substantially true."

Woodward's extraordinary success included the ability to deliver electrifying public explanations of science to his peers and his possession of a deep and refined sense of aesthetics, not only in the laboratory but out of it. Ultimately, however, his success rested on his ability to journey past the long exacting syntheses of quinine and penicillin to formulate a powerful extension of quantum orbital theory, which arose from his careful orchestration of the molecular dance, ring by ring, in a series of linked experiments that so delighted him.

His career plainly evolved as it unfolded, arising from and then in turn forming a unique personality. The riddle of Woodward's precocious start owes less to any fanciful gene than it does to the good fortune of natural selection. He not only was adept at three-dimensional visualization, but he also had a mind that was at once imaginative and tirelessly logical and nourished a devoted enthusiasm.

It is arbitrary that my comments should be directed to three chemists; the reader need not fear monotony among a dozen fascinating subjects, each one visible face to face, diverse witnesses to the protean development of human potential.

Sex and Death in Protozoa: The History of an Obsession, by Graham Bell. Cambridge University Press, 1988 ($44.50).

Witty, literate and penetrating, this brief book by a McGill biologist is not a history of protozoology at all but an intriguing general theory of sexuality, which looks along the way at multicellular organisms, cancers, aging and death. (On the subject of taxes, the biologist is silent.)

The first half is a critical survey of century-old literature on the life cycle of protozoa, which describes a protracted watch over the long run of a laboratory clone. ("Eleven thousand generations of *Paramecium*" runs one memorable title of 1926.) The rest of the book provides a sophisticated journey toward a more theoretical goal: the evolution of sex, which is provided within the context of general evolutionary theory and explicated through the ideas, mathematics and experiments of population genetics.

The goal can be put in the form of a fine old metaphor. A self-replicating machine is a contrivance that can assemble any finite arrangement it is instructed to build, regardless of the number of individual parts. Detailed instructions for assembly are provided in the form of a long tape, which must be copied and reloaded periodically into the machine. Yet every copying process introduces errors, rare though they may be, which accumulate even when error-correction routines are applied. Indeed, most damaging are errors introduced by the copying and error-correcting routines themselves. Even if a second error-correction system is installed, it must be replicated and passed on, and thus, it too cannot be error-free. It follows that a finite, self-replicating machine cannot be both unchanging and immortal.

Life has resources far more intricate than our machine model, and so the two are not exactly comparable. H. J. Muller drew an analogy between natural selection and a great Ratchet in life that turns inexorably—always in the same direction—as a population persists through time. Mutation pressure and natural selection act slowly, bringing the population to genetic equilibrium over many generations. Although every organism has more than one gene that may be adaptive in a given environment, not every gene can be considered optimal.

As a consequence of random chance, the best lines inevitably vanish. Over time the Ratchet continues to turn slowly, bringing the gamble of life to ruin and causing the population to die out slowly. Suppose that a perfect (error-free) population exists and each member in the population is allowed to produce just one offspring before death. By the author's calculations, the population will not last more than 1,000 generations against the Ratchet (assuming standard values for mutation rates and selective advantage).

In real life, however, it is possible to evade the clicking Ratchet. The easiest way is to increase in number; a population that multiplies, prospers. We are talking large numbers here; cell populations that contain more than 10-to-the-12 cells can beat the Ratchet by fruitfulness. Surely this is one explanation for the dominance of the tiny, self-replicating biochemical wonders, the single-celled organisms that were the first organisms to appear on the earth and that have flourished for four fifths of the geologic past.

The machine model demonstrates that vulnerability prevails during two stages of replication: when the instruction tapes are copied and corrected and when the individual parts are assembled. Replication in multicellular organisms is similarly divided: the germ-cell and the somatic-cell lines are sharply distinguished—in little nematodes as well as in human beings. Bell emphasizes the difference between the two cell lines throughout his book, suggesting that correction of somatic-cell lines differs from that of the germ line. The latter, he thinks, ages as the Ratchet turns, slowly accumulating mildly harmful errors.

In contrast, somatic senescence presents a conflict of interest for the organism. Developmental patterns that lead to early reproduction (although related to functional loss later in life) are favored by natural selection. Multiple organs fail as we age, not by a chance accumulation of somatic mutations here and there but more probably by a slow decline in the efficiency of our overall repair mechanisms.

The longevity of tissue cultures from multicellular organisms is an old topic. Indeed, early investigators thought that isolated cells might be immortal, but they were proved wrong in the 1960's when it was demonstrated that human tissue cells only replicate from 40 to 60 times during development. (The adult human being undergoes a total of about 37 cell divisions from the time of fertilization.) The striking exception is the human tumor cell, many forms of which seem to divide indefinitely. These cells almost always have an unusual karyotype (chromosome pattern), and the author writes, "It is abundantly clear that there is some link between karyotype and clonal lifespan, but the mechanism involved is entirely obscure."

Life's answer to clonal aging is sex, which need have no connection with gender or even with reproduction. In ciliates, "two cells enter sexual conjugation, and two cells emerge from it." The goal is the creation of informational diversity by mingling two separate lines of descent. At the most fundamental level, sex permits a damaged genetic tape to be repaired, which can be accomplished because the elaborate machinery for shifting tape segments among strands of DNA exists. At the next level, sex outpaces the Ratchet by adding an independent source of genetic variation, which can erase deleterious mutations.

Yet it is puzzling that sex persists (given its high cost) among organisms that live in old, complex and stable environments and already have abundant genetic diversity. Bell suspects that sexuality is adaptive in such environments, where the degree of competition between organisms is high. Consider, for example, the arms race that has been under way for so long between parasite and host and between predator and prey.

This original, provocative and wide-ranging book is bound to stimulate a general reader, although it must be said that economic metaphors are the responsibility of the reader alone.

Eames Design: The Work of the Office of Charles and Ray Eames, by John and Marilyn Neuhart with Ray Eames. Harry N. Abrams, Inc., 1989 ($95.50).

The Office was three blocks inland from the beautiful and then unfashionable Venice beach in westernmost Los

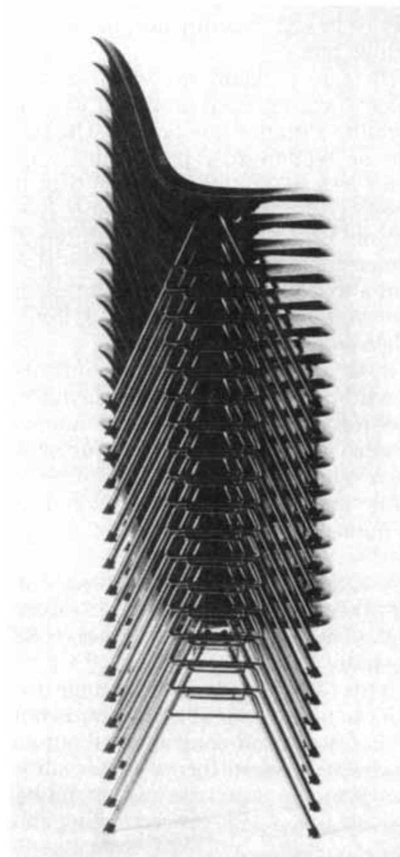
Angeles. For more than 40 years its industrial floor and windowless bays enclosed the tools and everyday work of an untiring pair of American artists. Charles Eames, architect—"for us, the name is synonymous with the word 'designer'" the book begins—died in 1978. Ray Eames, his wife and partner, died 10 years after him.

During the glowing decades of their Office, an atelier that would fit into the pages of Vasari, dozens of the most talented artists worked for and with the Eameses. Among them were John and Marilyn Neuhart, now faculty members at U.C.L.A. Theirs is the devoted authorship of this volume, although Ray's sensibilities can be found throughout the book in every image and in most phrases. It was her determination to turn the resources of the studio into a legacy for those who followed that has so endowed this rich view of the Eames enterprise; it offers by count 3,504 illustrations, 2,107 of them in color!

After all, the Office was once home to some 40,000 35-millimeter slides from around the world, most of them images taken by Charles. The wide, windowless bays were filled with a slowly moving parade of toy ships and trains, an intricate furniture lathe, computer-controlled animation cameras, an octopus holding tank, carved duck decoys, quaint clockwork calculators, dry tumbleweed and a plywood radar dome that had been hot-molded during wartime right in the studio. Everything had its place and purpose, even beyond the aesthetic pleasure provided by its presence. Most of what the Eameses considered to be "good stuff" can be glimpsed somewhere in the images that crowd these 450 big pages, arranged in order of maturation over the passing years.

At first glance, a dozen of these pages look like contact sheets back from the film processor, arrayed as they are in a grid of 18 to 24 images. Yet the prints are no mere run of the spool but rather the select catch of an artist's ubiquitous eye: one grid consists entirely of images of the circus—from carved wagon hubs to painted clown faces—another juxtaposes the patterns within a digital computer, discrete transistors and magnetic cores; another portrays the satisfying textures of decorated papers and master weavings from Florence to Kyoto.

The great themes of this lifework are sounded with enchanting variations. The Eames House is presented, in multiple gatefold, standing in its meadow high above the sea. It is the Case Study



EAMES stacking chair is a design classic.

they built and lived in from 1949 until their deaths. Both intimate and dramatic, its industrial warehouse appearance becomes domestic; its elegant interior is as diverse and personal as was their Office a few miles south. The famous Eames chairs are here, model after model over the decades, of molded form, adorned with novel fastenings so painstakingly developed, in styles from chaises deluxe to frugal shells. So, too, are the multiple tandem sling assemblages, the chairs that first emerged in 1962 and that now decorate the long corridors of so many airports around the world.

The entire collection of Eames films are annotated and sampled by chosen frames. Eighty films make up the list, among them the celebrated 1952 study of white suds washing across a blacktop schoolyard, the nine precisely animated minutes that deal so rigorously with the relative sizes of objects in the universe (the film was produced in two versions a decade apart) and the three animated minutes that magnify the posturing actors of the commedia dell' arte, which Jacques Callot engraved in

17th-century Tuscany. In another film the viewer looks directly into the eyes of famous circus clowns, each one painstakingly applying his distinctive makeup in the dressing tent, while outside the band plays on, cuing each performer to the moment of his appearance in the ring.

The gamut of subjects is broad: one film looks at the Victorian cast-iron wheels and brass gears of Lick Observatory; others act as instructive primers for feedback and communication theory; still others give light-hearted visual treatments of symmetry or map Roman imperial expansion. One film celebrates bread, one the brushwork in the late works of Paul Cézanne. In another film a score of varied tops, among them a thumbtack on the drawing board, spin to inevitable stillness. And there are 50 more films.

Those films, like the exhibitions the Office made for clients both great and small, are marked by plenitude: images, examples, memorabilia, juxtapositions flood past—often on many screens at once—or on long walls set intricately with text, image and object. Yet, however playful the feast of images might seem, nothing passes by without meaning. Ray insisted on mastery of color and form; Charles was stubbornly determined to gain clarity for himself before capturing a subject, exhausting many a consultant's little bag of expository devices. Their corpus is free of any taint of the cynicism that prevails on today's screens.

One 1972 film is about as self-reflective as anything the Eameses made. It was called *Design Q & A;* the exchange owes much to the innocent gravity of the French interviewer: "Q. What is your definition of 'design'? A. A plan for arranging elements in such a way as to best accomplish a particular purpose... Q. Is it a method of expression? A. No—it is a method of action... Q. Does the creation of design admit constraint? A. Design depends largely on constraints. Q. What constraints? A. The sum of all constraints...each problem has its own peculiar list. Q. Does design obey laws? A. Aren't constraints enough?...Q. To whom does design address itself: to the...masses...the specialists...to a privileged social class? A. To the need."

Possibly the most welcome line in this bright treasury is the final note that tells the reader where to go to get the Eames films now available (Pyramid Films in Santa Monica, Calif.) and happily promises us more "Eames films...on video cassette," to address the need.

ESSAY

The genome initiative: how to spell "human"

by Norton D. Zinder

The beginning of the human genome initiative harbors a strange irony. In 1984 Robert Sinsheimer, as chancellor of the University of California at Santa Cruz, was thinking about large telescopes and the large amounts of money they cost; he was also thinking of putting Santa Cruz on the map biologically. He came up with the idea of sequencing the nucleotides of all of the human genes. So in 1985 he called together a group of scientists for a workshop. His account of the meeting raises many major points that still occupy us today. This is the same Robert Sinsheimer who in 1977, by suggesting that there should be limits to inquiry, caused the molecular biologists fits over their plans to do recombinant-DNA experiments. Had we listened to him then, the human genome initiative would be impossible now.

Initially almost all the scientists at the 1985 meeting were highly skeptical. Two almost diametrically opposed positions were formulated. One was that we would not learn enough to make the large and expensive effort worthwhile. This feeling was in part based on the fact that about 90 percent of the human genome may have no function (for one, it does not code for protein). Some call it junk (but it is not garbage). A related idea held that a large applied-research program would distort the workings of science.

On the other side were those who felt we would learn far too much. Having the human genome at hand might provide an infinite number of new reasons for genetic discrimination by employers and insurance companies; it might even inspire Nazi-like eugenic measures. At a minimum, the number of genes known would increase manyfold, and there would still be a long lag between the time a disease gene is identified and the time treatment might be available.

Still, as the meeting ended, almost all those attending were if not enthusiastic about the project at least approving of it. This pattern of skepticism evolving into enthusiasm has repeated itself over and over in subsequent meetings and workshops. The consensus has spread outward, concentrically. A rotating core of some dozen well-known molecular biologists attended each meeting. At each new event an equal number of newcomers were added to expand the pool for future meetings.

The next gathering was at Santa Fe, N.M., sponsored by the Department of Energy. The DOE was used to expensive, large projects, and needing some new ones to maintain the national laboratories, it had picked up the genome initiative. The DOE was not without portfolio: it had been doing some radioactivity-related genetic work for a long time. Other meetings were held at Cold Spring Harbor in New York and then came studies by the Office of Technology Assessment and the National Research Council.

The spring of 1988 saw a quantum jump in the program's credibility when James D. Watson, of DNA-structure fame, agreed to be an assistant director at the National Institutes of Health for the human genome program. An NIH conference at Reston, Va., focused in an operational sense on what should be done. Again, the recommendation was to proceed, and following the lead of the NRC report, the conferees set forth programmatic goals. There was to be a five-year effort to map the human genome by both genetic and physical means. So that more detailed information on human evolutionary and developmental questions could be obtained, similar efforts to map the genome of several animal models were to proceed simultaneously. The conferees also saw that the price of sequencing must be reduced to less than 50 cents per base—amounting to a fivefold improvement in the technology.

The project finally got under way on January 3, 1989, when the first meeting of the Program Advisory Committee of the NIH took place. It is a talented committee of biologists chaired by the author. Since then, grants have been given, workshops on chromosomes have been organized and, most important, the Human Genome Office has been staffed and turned into an independent unit of the NIH. This center now controls the distribution of monies. The DOE carries out its program, mostly in-house at the national laboratories.

A joint task force on information technology (JITF) has been established with the DOE. Ultimately the JITF may join with its European and Japanese counterparts to create a single informatics advisory group. A strong program has been established in collaboration with the DOE on ethical and legal matters. It will both provide grants-in-aid for studies and organize, for public information and discussion, town meetings throughout the country.

Last summer at a retreat at Cold Spring Harbor, representatives of the NIH, DOE and other invited experts developed a five-year program. We submitted a report to Congress that will be revised yearly. We have estimated that $200 million per year for about 15 years will be needed; the clock would start in the 1991 fiscal year.

In sum, after about three years of discussions and one year of activity, a strong, broadly popular program has been created that is sensitive to criticism from public and scientific sources. Let us then ask what it is that has caused the initiative to carry the day at each of the meetings at which the plan was discussed. We must also consider that the program has captured the imagination of many all over the world from politicians to laypeople.

The answer is often phrased in terms of expected medical advance. Other arguments rest on important findings to be made in human biology, particularly in evolution, development and neurobiology. Perhaps there is also a subtle sense that as we gain the ability to identify the 100 or so genes that distinguish one individual from another, the desire to discriminate on the basis of a few genes might be mitigated.

No, I think the answer is both more simple and yet more profound. Determining the map and sequence of the human genome is about us. And we are greatly interested in us. We can only consider these efforts as a form of human anatomy: the beginnings of a great adventure in human biology. We have to do the sequence only once, and for all time we shall have for study the human genetic dictionary.

With the map and sequence of our genome, everyone will be able to see and participate. We have brains, we have hearts, we have legs and we have nucleotides. The last is what makes us us, and that is exciting. No wonder, then, that whenever groups of people considered whether to go forward, they came down on the side of "go."

Still, the two types of opposition persist: learning too little and learning too much. They will force us to question continually what we do.

NORTON D. ZINDER is professor of molecular genetics at Rockefeller University. He chairs the Program Advisory Committee for the NIH human genome project.


All of GM, going all out for you.
"Used to be that an American car in Switzerland was something of a rarity.
Not anymore. GM sells a lot of cars over here these days–good cars at good value. Which is nice for me–I'm the local dealer."
LU·136662