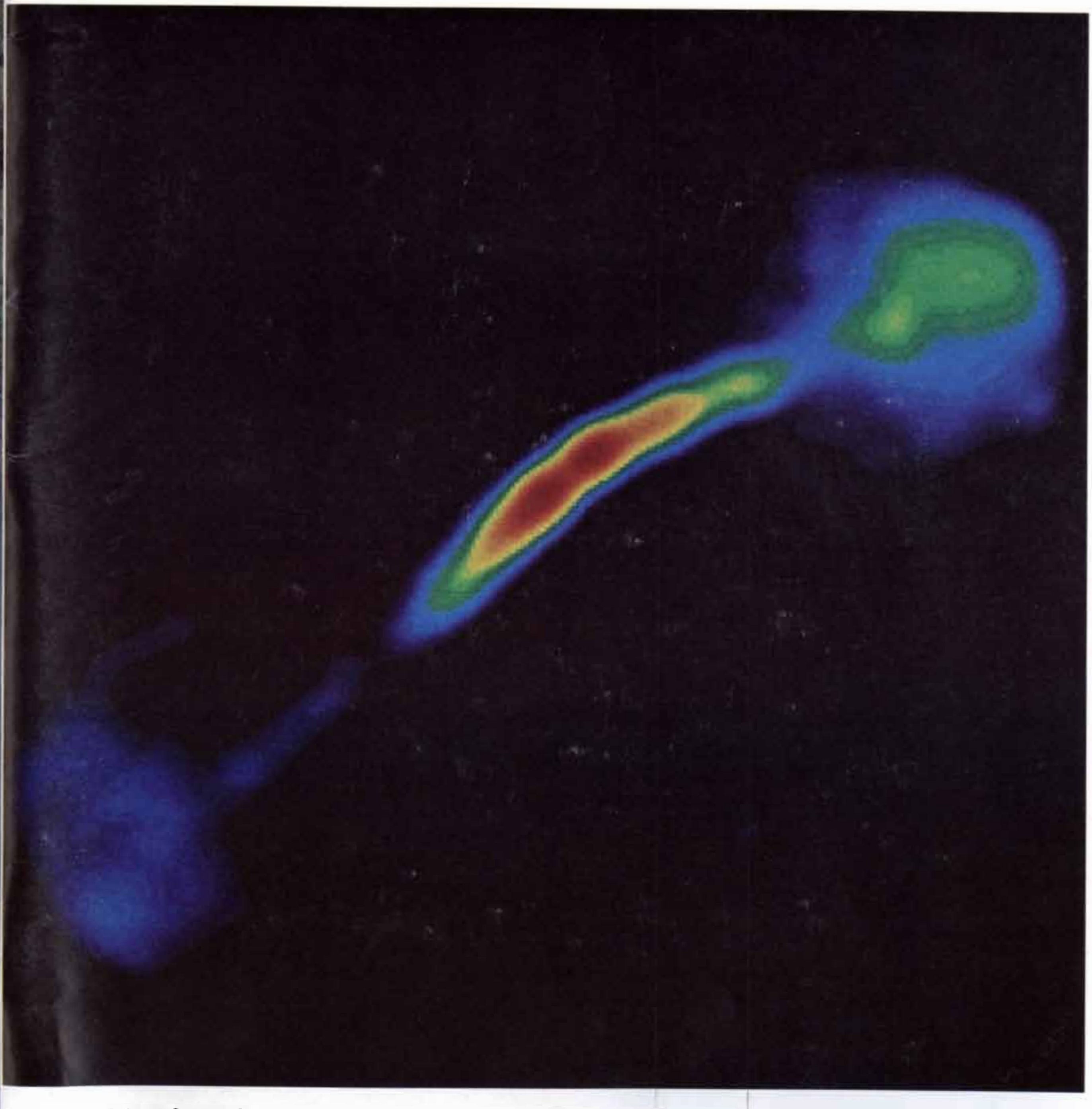
# SCIENTIFIC AMERICAN



COSMIC JETS

$2.00

May 1982

# Why There's America's

## BECAUSE WE'RE LEADING THE WAY WITH EXPERIMENTATION.

As you read this, Ford Motor Company experimental vehicles are criss-crossing the country using seven different alternative fuels. Last year in Brazil 37% of all Fords sold used ethanol.

Ford Motor Company leads the automobile industry in alternative fuel technology. Right now there are propane powered Ford Granadas and Mercury Cougars available to fleet buyers. Propane is not only cheaper than gas, it's cheaper than diesel.

Probe III.

Seat belts save lives–buckle up.

## BECAUSE WE'RE LEADING THE WAY WITH RESOURCES.

The Ford family of affiliated people is 400,000 strong, and lives and works in thirty-one countries of the world.

This extended family supplied the worldwide design and engineering cooperation that developed the successful world cars, Ford Escort and Mercury Lynx.

Overseas we've outsold General Motors for sixteen years in a row.

Ford Aerospace & Communications Corporation is another major resource, which, if shown separately, would rank in the Fortune 500.

Forty-seven satellites that have been successfully launched have been built by Ford Aerospace. Soon two-thirds of all overseas communications will travel by Ford Aerospace built satellites.

So will almost all intercontinental television programming.

# A Ford In Future.

## BECAUSE WE'RE A LEADER IN AERODYNAMICS.

The Mercury LN7 is one of the most aerodynamically efficient standard production cars made. The Ford Escort and Mercury Lynx have less wind resistance than the VW Rabbit, Plymouth Reliant, Chevrolet Citation and Chevette.

In Germany, the most astounding car in the 1981 Frankfurt Auto Show was an experimental Ford, the Probe III. It got the best press, being dubbed–"slippery as a futuristic study in aerodynamics."

Compare the Probe III to most '82 European models, and you come up with a very different shape for the future. You get a 50% reduction in drag coefficient, for example, which conceivably could result in a 27% increase in fuel efficiency at higher speeds.

## BECAUSE WE'RE LEADING THE WAY WITH OUR FORD ESCORT AND OUR F SERIES TRUCKS.

Ford's F series of light, conventional trucks outsells every other car line or truck series sold in America. And they have done so for five straight years. In fact, based on manufacturer's retail deliveries, we had total truck line leadership in the 1981 calendar year, outselling foreign or domestic truck lines for the fifth consecutive year.

Our Escort was the best selling compact car for 1981, outselling every Japanese, every German—every imported car line in America, based on manufacturer's retail deliveries for calendar year 1981.

Is there a Ford in America's Future?

What do you think?

Ford

There's A Ford In America's Future.


·AMARETTO DI SARONNO·ORIGINALE·1525·

FOR FREE FOOD AND DRINK RECIPE BOOKS, WRITE BOOKLETS: P.O. BOX 5145, F.D.R. POST OFFICE, N.Y., N.Y. 10150.

Established 1845 **SCIENTIFIC AMERICAN** May 1982 Volume 246 Number 5

ARTICLES

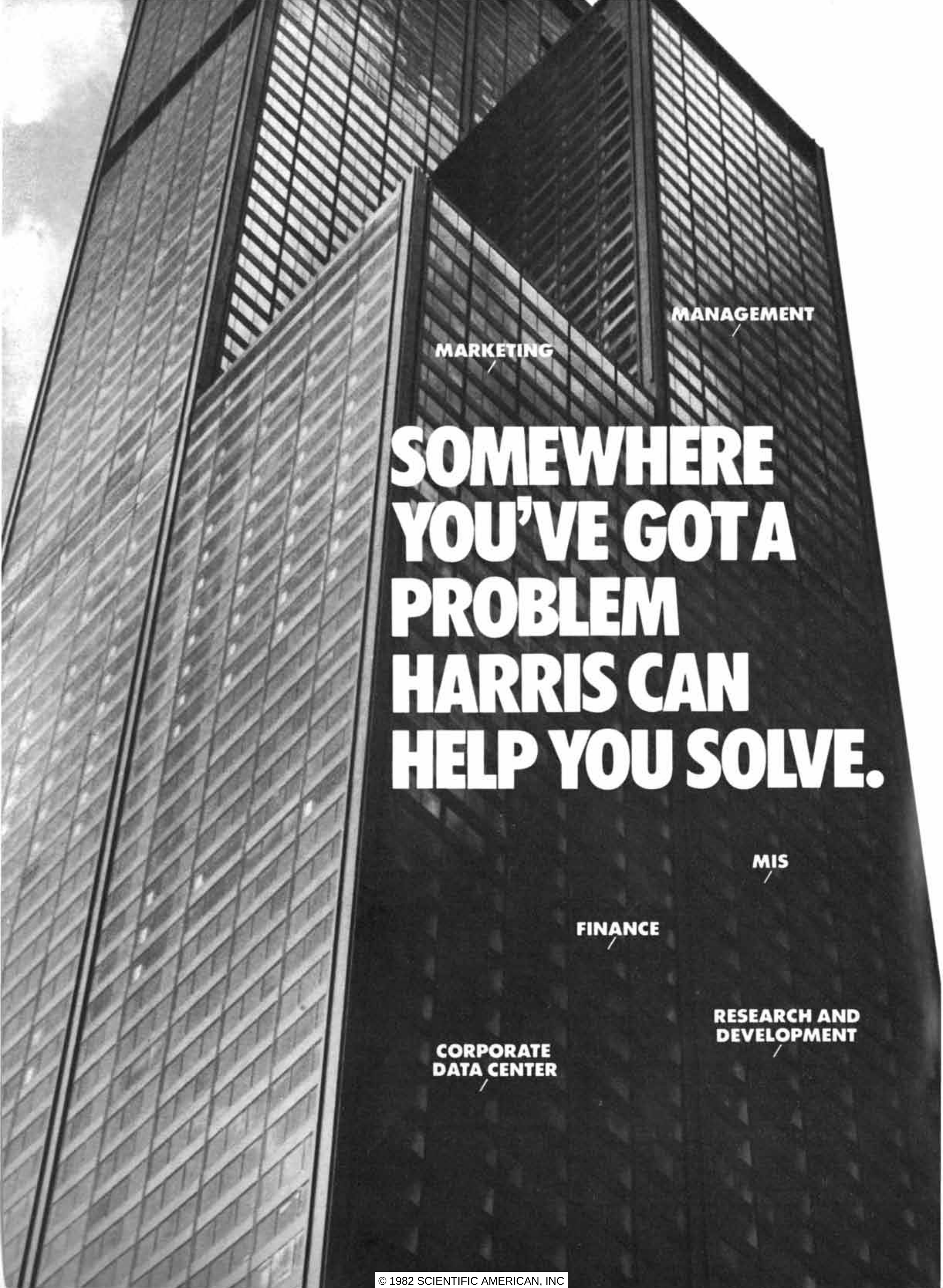


DEPARTMENTS





SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017.  SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $21 PER YEAR IN THE U.S. AND ITS POSSESSIONS; $27 PER YEAR IN ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.


MANAGEMENT
MARKETING
SOMEWHERE YOU'VE GOT A PROBLEM HARRIS CAN HELP YOU SOLVE.
MIS
FINANCE
RESEARCH AND DEVELOPMENT
CORPORATE DATA CENTER

It's the problem of communicating and using information to make your company more productive.

The problem may be in your marketing or R&D group. It could be hurting your financial or engineering department.

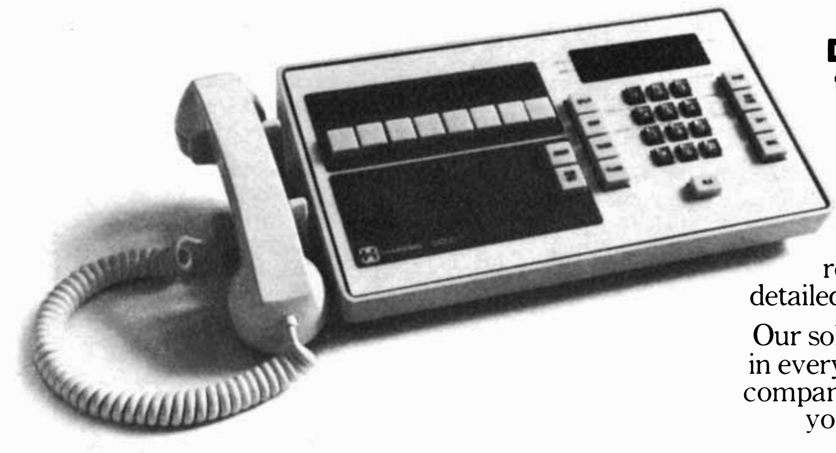
Wherever the problem, it hurts your entire company. But Harris has solutions. We offer the broadest range of products to expand your current systems—*without* costly interruptions.

**Super-minicomputers for your engineering and R&D departments**

Our super-minicomputers are specifically designed for engineering, scientific and general business data processing. They offer outstanding power, performance and reliability. With over 300 software applications, including computer-aided design and simulation.

**Distributed data processing for all of your departments**

Whether you want to support a centralized data base—or a departmental system—we've got solutions. Our remote batch and interactive terminals grow easily into distributed data processing systems. Powerful features include: Support of multiple host computers. An exclusive fiber optic data link for high-speed data transmission. Word processing matched precisely to your needs.



**Digital PBX's to increase productivity**

Harris digital PBX telephone systems help reduce costs and increase productivity–by switching both voice and data. Among the benefits: Least-cost routing. Automatic call back. And detailed station message accounting.

Our solutions help increase productivity in every department. To help your whole company run more efficiently. Wherever your problem, Harris has a solution.

*See us in Houston, June 7-10 at NCC Booth 4000.*

**HARRIS**

*For a copy of the Harris Integrated Information Processing Systems brochure, send your business card or write on your company's letterhead to Jack C. Davis, Group Vice President, Department 3, Information Systems, Harris Corporation, 1025 West Nasa Boulevard, Melbourne, FL 32919. Or phone 305-727-9100.*

## THE COVER

The illustration on the cover shows the pattern of radio emission from the giant elliptical galaxy IC 4296, as mapped with the aid of the Very Large Array (VLA), 27 interconnected radio telescopes at a site west of Socorro, N.M. The galaxy, probably a member of the Centaurus cluster of galaxies, is 120 million light-years away. It is nonetheless the closest of the "classical" radio sources: the sources consisting of two symmetrical lobes of radio emission. The earliest radio maps of IC 4296 showed a patch of radio emission between the lobes. Then three years ago work at the VLA showed that the patch is really two jets of ionized gas emerging in opposite directions from the center of the galaxy. The map on the cover results from observations made last November. The jets appear in red and yellow, the lobes in green and blue. The distance between the lobes is almost a million light-years. Detailed radio observations of IC 4296 reveal a pointlike central source of radiation not visible in this picture. It is presumably the engine that is emitting the jets (see "Cosmic Jets," by Roger D. Blandford, Mitchell C. Begelman and Martin J. Rees, page 124).

## THE ILLUSTRATIONS

Cover courtesy of the National Radio Astronomy Observatory

| Page | Source |
|---|---|
| 24 | Ray Atkeson |
| 58–65 | Albert E. Miller |
| 67–77 | Gabor Kiss |
| 79 | Technicare Corporation |
| 80–85 | Walken Graphics |
| 86 | Technicare Corporation (*top and middle*), *Journal of Computerized Axial Tomography* (*bottom*) |
| 87 | James M. S. Hutchison, Francis Smith and John R. Mallard (*left and middle*), Ilil Arbel (*right*) |
| 88 | Walken Graphics |
| 103 | Richard J. Feldman, National Institutes of Health (*top*); Bunji Tagawa (*bottom*) |
| 105–109 | Bunji Tagawa |
| 110 | Ilil Arbel (*top and right side*); Philip Leder, Harvard Medical School (*left side*) |
| 111–115 | Bunji Tagawa |
| 117 | Albert M. Plaza |
| 118–122 | Tom Prentiss |

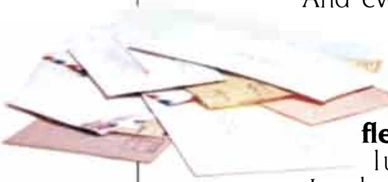
| Page | Source |
|---|---|
| 125–128 | National Radio Astronomy Observatory |
| 129–138 | Ian Worpole |
| 145 | Archives of the City of Siena |
| 146 | Gordon Moyer (*top*), Library of Congress (*bottom*) |
| 147 | Edmondo Lamalle, S.J. (*left*), Cornell University Library (*right*) |
| 148 | National Library of Naples, photograph by Giovanna De Vita (*left*); National Central Library of Florence, photograph by Guido Sansoni (*right*) |
| 150–151 | Ilil Arbel |
| 155 | James F. Shepard, Kansas State University |
| 156 | Patricia J. Wynne |
| 157–158 | James F. Shepard, Kansas State University |
| 160 | Bikrum S. Gill and Lauren Kam, Kansas State University |
| 162–164 | James F. Shepard, Kansas State University |
| 170–177 | Michael Goodman |

# TIME WAS WHEN YOU COULD HARDLY MANAGE YOUR INVESTMENTS WITHOUT YOUR ACCOUNTANT, YOUR BROKER AND YOUR POSTMAN ALL GETTING INVOLVED.

# THEN MERRILL LYNCH INVENTED CMA™.

**It used to be** a hassle to switch investments around—and keep track of them. Your broker handled some...but not if you invested in late 19th-century European oils or a lakefront property.

And even after talking to your broker, there was that wait for the mail.



**Now, instant flexibility.** The revolutionary Merrill Lynch Cash Management Account® financial service eliminates all that.

At the heart of CMA™ are a Merrill Lynch brokerage account and an experienced Merrill Lynch Account Executive.

**Don't lose interest.** The CMA service also makes sure your money keeps earning money virtually all the time, at current money market rates. Your cash balance is invested in your choice of three professionally managed money market funds, with dividends reinvested daily.

If you need extra cash, just write a check or use your special VISA® card—both provided by Bank One of Columbus, N.A. You may automatically borrow up to the *full margin value* of your securities in the account.

**An accountant's delight.** At the end of each month, you receive a single, comprehensive statement that keeps track of everything you've done through your CMA. *Everything.*

**How to get started.** To join the 600,000 investors who already have put more than $30 billion to work through CMA, all you need is a minimum of $20,000 in securities and/or cash.

For more complete information, including a prospectus containing all sales charges and expenses, call 1-800-526-3030 Ext. 559 (in New Jersey, call 1-800-742-2900 Ext. 559). Read the prospectus carefully before you invest or send money.

**THE MORE YOU DEMAND OF YOUR MONEY, THE MORE YOU NEED CMA™.**

**Merrill Lynch**

Merrill Lynch, Pierce, Fenner & Smith Inc.

**A breed apart.**

CMA™ is not available in all states. The Cash Management Account program is proprietary to Merrill Lynch. U.S. Patents Pending. © 1982 Merrill Lynch Pierce Fenner & Smith Inc. Member SIPC.


# Can you spot the five cars headed for Wyoming?


BURLINGTON NORTHERN
L&N
MD&W
BURLINGTON NORTHERN
BURLINGTON NORTHERN
BURLINGTON NORTHERN
CONRAIL
BURLINGTON NORTHERN

# LETTERS

Sirs:

I enjoyed the highly articulate article by Gunther S. Stent and David A. Weisblat, "The Development of a Simple Nervous System" [SCIENTIFIC AMERICAN, January]. Beyond the substantive content of the paper were the superb photographs. The authors refer to the aversion to leeches that "the mere mention" of them can evoke. Since they quote from the script of *The African Queen,* I thought they and the readers of *Scientific American* might be interested in Chapter 64 of Nietzsche's *Also Sprach Zarathustra.* This chapter, "The Leech," includes the following exchange:

"'Then thou art perhaps an expert on the leech?' asked Zarathustra; 'and thou investigatest the leech to its ultimate basis, thou conscientious one?'

"'O Zarathustra,' answered the trodden one, 'that would be something immense; how could I presume to do so!

"'That, however, of which I am master and knower, is the *brain* of the leech:—that is *my* world!

"'And it is also a world! Forgive it, however, that my pride here findeth expression, for here I have not mine equal. Therefore said I: "here am I at home."

"'How long have I investigated this one thing, the brain of the leech, so that here the slippery truth might no longer slip from me! Here is *my* domain!'"

Thank you for excellent articles such as this one.

JOEL BRUMLIK, M.D., PH.D.

Professor and Chairman
Department of Neurology
Stritch School of Medicine
Loyola University
Maywood, Ill.

Sirs:

We appreciate Professor Brumlik's kind remarks about our article and his calling your readers' attention to Nietzsche's insightful recognition that a conscientious man can have no higher goal than to become an expert on the brain of the leech. No other proof seems to be needed to substantiate Nietzsche's claim that *Thus Spake Zarathustra* (written in the very years in which Charles O. Whitman carried out his pioneering work on the developmental neurobiology of the leech) is the most profound book of world literature. The passage referred to by Professor Brumlik is not unknown to modern students of the leech brain. In their introduction to the monograph *Neurobiology of the Leech* (Cold Spring Harbor, N.Y., 1981) the editors quote it to answer their own rhetorical question "Why, then, the leech?"

GUNTHER S. STENT

DAVID A. WEISBLAT

Department of Molecular Biology
University of California
Berkeley

Sirs:

Douglas R. Hofstadter ["Metamagical Themas," November, 1981] paints an intriguing and delightful picture of the world of the strange attractor. However, he omits one very important part of the picture, namely the involvement of topological dynamics. To some extent the "flurry of theoretical work" he hopes may materialize has already done so, in the 1960's. On the other hand, a renewed flurry is under way, thanks to the recently developed links between the mathematics and the physics. Certainly the role of pure mathematicians has not been confined to computer experiments, as Hofstadter seems to suggest.

The iteration of a function is (almost!) what topologists call a "discrete dynamical system." In the 1960's Stephen Smale was trying to classify the "typical" behavior of such systems, and the hope at the time was that all attractors other than pathological accidents would turn out to be nicely behaved and of known type. Smale dashed this hope by finding an attractor (actually this is technically untrue; it is a saddle, but it can easily be modified to give an attractor) called the horseshoe, which is both typical and "strange"—*not* of a known type! It is from this era, and the work of topologists, that the term strange attractor comes.

Smale's example was based purely on the geometry of the horseshoe, that is, he did not exhibit specific dynamical equations leading to it. Although the omission is easily remedied, it was in the spirit of the topological approach not to demand it. Continued analysis from this geometric viewpoint led to a fairly complete understanding of "Axiom A" systems.

Meanwhile, in about 1963, E. N. Lorenz invented his now famous system of differential equations with "chaotic" dynamics—a capital discovery that was ignored almost completely. As interest in chaos grew, topologists learned of Lorenz' work and tried to tie it in with their previous analysis. An immediate problem is that the Lorenz system is *not* Axiom A, so that the existing tools did not work straightaway, but the challenge of the Lorenz attractor is slowly yielding to refined techniques and indeed is suggesting new topological ideas.

(The existence of strange attractors in the equations of celestial mechanics was in fact known to Henri Poincaré in about 1899, but he lacked the necessary topological tools to understand it and did not pursue the matter.)

The connection of the Lorenz equations with turbulence is even more curious. They are *approximations* to certain equations of fluid flow. If one takes *better* approximations, the chaotic behavior disappears! Nevertheless, as was suggested by David Ruelle and Floris Takens in 1971, it does appear that strange attractors are somehow involved in the transition to turbulence in Navier-Stokes flow (although their original model is no longer considered likely in its details). Laser-based measurements by H. L. Swinney, P. R. Fenstermacher and J. P. Gollub show that the evolution of the power spectrum of fluid velocity is consistent with a strange-attractor model in several standard experimental systems.

Currently there is considerable interaction between the various approaches to all these phenomena, and an integrated picture is slowly emerging. It is important to note, however, that pure mathematicians had discovered and analyzed many facets of the problem some 10 years before anyone realized there were specific applications to physics of the ideas concerned. What we are witnessing, in fact, is the full flowering of Poincaré's program to apply topology to dynamics.

IAN STEWART

Mathematics Institute
University of Warwick
Coventry, England

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

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

**Offprint correspondence** and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 75 cents.

**Subscription correspondence** should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 5969, New York, N.Y. 10017. For change of address, notify us at least four weeks in advance. Send both old and new addresses and enclose an address imprint from a recent issue. (Date of last issue on your subscription is shown at upper right-hand corner of each month's mailing label.)

Name ______________________

New Address ______________________

______________________

______________________

Old Address ______________________

______________________

______________________

# 50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

MAY, 1932: "The most significant scientific discovery of recent years was announced recently in Cambridge, England, where Dr. James Chadwick established the existence of the subdivision of matter known as the neutron. To understand this conception it is necessary to recall that the atom, once regarded as the smallest possible subdivision of matter, is now regarded as being made up of electrons and protons held together by the attraction between negative and positive charges. The neutron is a combination of a single electron and a single proton, the respective electric charges having been neutralized by their union."

"There is no longer any room for reasonable doubt that the Great Nebula in Andromeda is formed of a system of stars that is similar in its essential nature to our own system. It is not the only such system in the sky; hundreds of thousands of others can be observed. Although they are of varied shapes and constitutions, it is found that the greater number of them can be arranged in a sequence. At one end of the sequence are nebulae consisting solely of round, fuzzy masses. Halfway along the sequence are nebulae such as the Great Nebula in Andromeda, which consists of a central fuzzy mass surrounded by stars. Like our own system of stars, these nebulae are generally flat in shape. It is easy to obtain a theoretical interpretation of this sequence. We know how an increase in the speed of rotation of a body is accompanied by a flattening of its shape. It is natural, then, to interpret the sequence of nebulae as one of bodies that are rotating at different speeds."

"'Television is here,' says the technician, and the public looks around for a television 'machine' with which it can see stage presentations, boxing matches, horse and automobile races and all the other interesting events that go to make up our complicated national life. But the public looks in vain: here is a televisor that shows a picture the size of a postage stamp, here is one that gives an image, somewhat distorted by a huge magnifying glass, apparently about six inches square. Where, then, is the kind of television that will give us in our own home the equivalent of a motion picture with sound accompaniment? Quite frankly it is not available. We can go today to an up-to-date radio shop and buy a televisor, but it will be erratic in operation, it will require constant attention on the part of the operator and it will bring to us only mediocre pictures. Thus if you want something that will compare favorably with your home movie equipment in quality of image, you cannot have television today."

"Leaving a note saying, 'My work is done. Why wait?' George Eastman, founder and chairman of the board of the Eastman Kodak Company, shot himself at his home in Rochester, N.Y., on March 14. He had been in ill health for several years and at the time of his death was in bed under the care of a physician. His age was 77. Mr. Eastman was best known for his development of the complicated early camera into a low-priced commercial product, for his transformation of the highly difficult profession of photography into a popular pastime. His further improvements of cameras and films made possible the development of 'still' and motion-picture cameras to their present high state of perfection. He was looked upon as the first American manufacturer to use large-scale production methods and therefore contributed largely to the modern industrial era's efficiency. There can be no nobler tribute than that paid to him by an executive of his company: 'George Eastman played the game to the last. By his own hand he lived his life, and by his own hand he ended it.'"

SCIENTIFIC AMERICAN

MAY, 1882: "The enormous quantities of petroleum that are now produced have led many to consider seriously what is to be done with that magnificent combustible. It is perhaps natural that this production, with the consequent low price per barrel for a substance that contains no ash worth mentioning, should lead to many attempts to utilize it as a means of heat production. Hence we have had prospectus after prospectus of companies that will revolutionize the heat production of the world and give us light, heat and power at almost no expense. Petroleum-burning furnaces for the manufacture of iron and petroleum-burning locomotives have claimed the attention of the world for a brief space and then passed into forgetfulness. But why cannot petroleum successfully be used as fuel in competition with coal? Why have so many attempts to utilize petroleum as a means of producing heat resulted in failure? The answer, in brief, is that at present prices the heat-producing power of a dollar's worth of coal is so much in excess of the heat-producing power of a dollar's worth of petroleum. It is evident that at present prices the excessive cost of the heat-producing power of petroleum is so great that all efforts to substitute the liquid for the solid fuel must almost inevitably result in financial disaster."

"Professor Tyndall has communicated to the London *Times* an account of results obtained by Dr. Koch of Berlin in the investigation of the etiology of tubercular disease, as set forth by him in an address delivered on March 24 before the Physiological Society of Berlin. In pursuing these investigations Dr. Koch subjected the diseased organs of a great number of men and animals to microscopic examination, and he found in all cases the tubercles were infested with a minute, rod-shaped parasite, which, by means of a special dye, he differentiated from the surrounding tissue. Transferring directly, by inoculation, the tuberculous matter from diseased animals to healthy ones, he in every instance reproduced the disease. Dr. Koch has examined the matter expectorated from the lungs affected with phthisis and found in it swarms of bacilli, whereas in matter expectorated from the lungs of persons not thus afflicted he has never found the organism. Guinea pigs infected with expectorated matter that had been kept dry for two, four or eight weeks were smitten with tubercular disease quite as virulent as that produced by fresh expectoration."

"A curious discovery made by Signor Schiaparelli, director of the Royal Observatory at Milan, seems to start again that old and unanswerable question, 'Are the planets inhabited?' This Italian astronomer is one of the most assiduous watchers of the planet Mars. It was he who in 1877–78 first detected the many dusky bands that traverse and subdivide the ruddy portions of the martial orb. During last January and February he was able to observe and map out in more than 20 instances duplications of the dark streaks 'covering the equatorial region of Mars with a mysterious network, to which there is nothing remotely analogous on the earth.' The Italian astronomer has styled them 'canals.'"

"The funeral of Charles Darwin took place in Westminster Abbey on April 26. The pallbearers were United States Minister Lowell, the Duke of Argyll, Lord Derby, Professor Huxley, Sir Joseph Hooker, Sir John Lubbock, Alfred Russel Wallace and William Spottiswoode. Who would have dared to predict 20 years ago that the authorities of conservative England would so soon and so conspicuously recognize the merit of the author of *On the Origin of Species by Means of Natural Selection?* Or in what other age of the world could so radical a revolution in men's interpretation of the facts of life and nature have been wrought during the lifetime of one man?"

# Control Data's PLATO is changing how the world learns.

**Business and industry are turning to PLATO® computer-based training to meet the most massive training and retraining challenge in U.S. history. Technology is changing too fast, competition is too keen, and time is too short for anything less than the state-of-the-art answer to training's toughest problems.**

Developed by Control Data, PLATO computer-based training is a major advance in self-paced individualized learning.

Because of its proven ability to produce better-trained, more proficient employees, it is the choice of companies respected for their training excellence.

- Leading airlines use PLATO for the quality of flight and maintenance crew training that is absolutely essential in the airline industry.
- Major utilities rely on PLATO for the critical job of training nuclear plant operators, dispatchers and maintenance personnel in safety procedures.
- Auto makers are turning to PLATO for the tough challenge of retraining people for the advancing technologies needed to build better cars.

# PLATO
## COMPUTER BASED TRAINING

# THE AUTHORS

DAVIDSON R. GWATKIN and SARAH K. BRANDEL ("Life Expectancy and Population Growth in the Third World") are respectively Senior Fellow and Associate Fellow at the Overseas Development Council in Washington. They are collaborating in a study of the death rate in the developing countries and the effect of international assistance on population growth. Gwatkin received his bachelor's degree in 1962 at Haverford College and his master's degree in 1964 at Princeton University. From 1964 to 1977 he worked for the Ford Foundation. His last two assignments for the foundation were in Nigeria and India, where he served as adviser to the administrators of population programs. In 1977 he moved to the Overseas Development Council. Brandel was graduated from Denison University in 1963. She went on to get two master's degrees: one in South Asian studies from the University of California at Berkeley (1965) and one in demography and sociology from Georgetown University (1977). In 1978 she worked at the World Bank as a research demographer, moving to the Overseas Development Council in the same year. Gwatkin and Brandel would like to acknowledge the assistance of the U.S. Agency for International Development and the United Nations Fund for Population Activities, which have supported the work on which their article is based.

ELLIOTT D. BLOOM and GARY J. FELDMAN ("Quarkonium") are respectively professor of physics and associate professor of physics at Stanford University. Both have been working on states of matter formed by electron-positron annihilations. Bloom's B.A., received in 1962, is from Pomona College. His Ph.D., received in 1967, is from the California Institute of Technology. In 1967 he went to Stanford as research associate, achieving the rank of assistant professor in 1970, associate professor in 1974 and professor in 1980. He has been collaborating with a large group of physicists in Europe and the U.S. in experiments on electron-positron interactions. He plans to spend part of this year continuing the work at the Deutsches Elektronen-Synchrotron (DESY) in Hamburg. Bloom writes that "one of my favorite secondary interests is seismology. I have published a few papers in the seismological literature and I do some consulting on the subject." Feldman's B.S. (1964) is from the University of Chicago; his M.A. (1965) and Ph.D. (1971) are from Harvard University. He went to Stanford in 1971 as a research associate, becoming associate professor in 1979. He plans to spend the next academic year on sabbatical leave at the European Organization for Nuclear Research (CERN) in Geneva.

IAN L. PYKETT ("NMR Imaging in Medicine") is research fellow in physics at the Massachusetts General Hospital. He obtained his B.S. from the City of London Polytechnic in 1974. His Ph.D. in physics was granted by the University of Nottingham in 1978. From 1978 to 1980 he was a research fellow at Nottingham, working on imaging techniques. In 1980 he came to the U.S. as a physicist at the Technicare Corporation, for which he still works. Soon afterward Pykett joined the research team on NMR imaging at Massachusetts General.

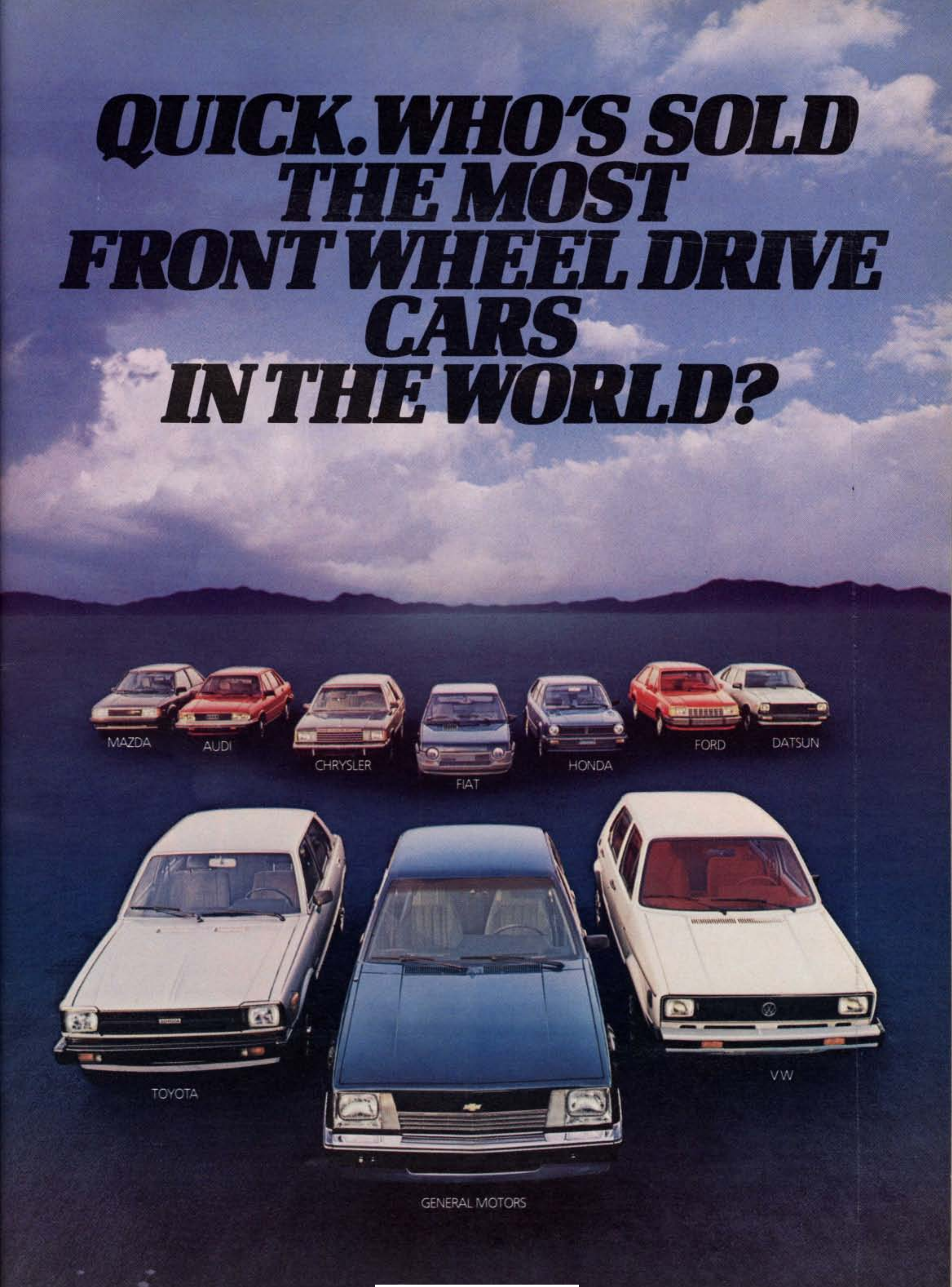
PHILIP LEDER ("The Genetics of Antibody Diversity") is John Emory Andrus Professor of Genetics and chairman of the department of genetics at the Harvard Medical School. He received his bachelor's degree in 1956 from Harvard College and his M.D. in 1960 from the Harvard Medical School. In 1962 he went to the National Heart Institute as a research associate. Later he served as a research medical officer of the National Cancer Institute, head of the section on molecular genetics of the National Institute of Child Health and Human Development (NICHD) and chief of the laboratory of molecular genetics of the NICHD. He moved to Harvard in 1980.

DAVID POLICANSKY ("The Asymmetry of Flounders") is a Cabot postdoctoral fellow at the Gray Herbarium of Harvard University. He was born in South Africa, and he received his secondary education there and in England. He came to the U.S. to study at Stanford University. His Ph.D. (1972) is from the University of Oregon. In 1972 and 1973 he was a postdoctoral student at the University of Chicago. He taught at the University of Massachusetts at Boston before moving to Harvard. His research interests are the evolution of sexual reproduction and how the timing of events in the life cycle, such as reproduction and metamorphosis, is controlled. For their help in obtaining adult flounders or in rearing flounder larvae Policansky would like to thank the staff of the New England Aquarium in Boston, particularly Louis Garibaldi, Michael Kelleher, Bruce Hecker and Paul Sieswerda; the staff of the Steinhart Aquarium in San Francisco, particularly John McCosker; Kunio Amaoka of Hokkaido University; Stanley Fullerton of Santa Cruz, and Jeffery Meyer of San Francisco.

ROGER D. BLANDFORD, MITCHELL C. BEGELMAN and MARTIN J. REES ("Cosmic Jets") are astrophysicists who share an interest in radio-wave sources that lie outside our galaxy. Blandford, born in England, obtained his B.A. (1970), his M.A. and his Ph.D. (both in 1974) from the University of Cambridge. He was Bye Fellow at Magdalene College in Cambridge and then a research fellow at St. John's College. In 1974 and 1975 he was a member of the Institute for Advanced Study in Princeton. In 1976 Blandford moved to the California Institute of Technology, where he is assistant professor of Theoretical Physics. Begelman is a postdoctoral fellow at the University of California at Berkeley. He received both his B.A. and his M.A. at Harvard University in 1974. His Ph.D., granted in 1978, is from the University of Cambridge. He went to Berkeley in 1979, but he has returned to Cambridge for several periods since then. Rees, like Blandford a native of England, is Plumian Professor of Astronomy at Cambridge and director of the Institute of Astronomy there. He got his B.A. (1963) and Ph.D. (1967) from Cambridge. He has held visiting appointments at Cal Tech, the Institute for Advanced Study and Harvard, and he was a member of the faculty of Sussex University before taking up his present post in 1973.

GORDON MOYER ("The Gregorian Calendar") is a free-lance writer on the history of science who has a particular interest in the principles according to which calendars are constructed. He got his bachelor's degree in English from the University of Maryland. Until 1980 he served as consultant to the Davis Planetarium of the Maryland Academy of Sciences on the calendar and the chronology of astronomical events. In that year he moved to Frankfurt am Main in Germany, where he has lectured on the Gregorian calendar at the Institut für Geschichte der Naturwissenschaften of the Goethe-Universität at the invitation of the late Professor Willy Hartner. Moyer has also contributed poetry to magazines and anthologies.

JAMES F. SHEPARD ("The Regeneration of Potato Plants from Leaf-Cell Protoplasts") is professor of plant pathology at Kansas State University. He was graduated from Cornell University with a B.S. in 1963, and he went on to receive his M.S. and Ph.D. (1967) from the University of California at Davis. In 1972 and 1973 he was a visiting associate professor at the University of Wisconsin; there he began his work with plant-cell protoplasts. In 1973 he moved to Montana State University, becoming professor in 1975. In the following year he was made professor and chairman of the department of plant pathology at Kansas State; he was chairman for three years.


QUICK. WHO'S SOLD THE MOST FRONT WHEEL DRIVE CARS IN THE WORLD?
MAZDA
AUDI
CHRYSLER
FIAT
HONDA
FORD
DATSUN
TOYOTA
GENERAL MOTORS
VW


VOILÀ,
RENAULT 5 TURBO
RENAULT ALPINE 310
RENAULT FUEGO
RENAULT 9
792 KD 92

# METAMAGICAL THEMAS

## *Number numbness, or why innumeracy may be just as dangerous as illiteracy*

by Douglas R. Hofstadter

The renowned cosmogonist Professor Bignumska, lecturing on the future of the universe, had just stated that in about a billion years, according to her calculations, the earth would fall into the sun to a fiery death. In the back of the auditorium a tremulous voice piped up, "Excuse me, professor, but h-h-how long did you say it would be?" Professor Bignumska calmly replied, "About a billion years." A sigh of relief was heard. "Whew! For a minute there I thought you'd said a *million* years."

John F. Kennedy told a story about a famous French soldier, Marshal Lyautey. One day the marshal asked his gardener to plant a row of trees of a certain rare variety in his garden the next morning. The gardener said he would gladly do so, but he cautioned the marshal that trees of this kind take a century to grow to full size. "In that case," replied Lyautey, "plant them this afternoon."

In both of these stories a time a long way in the future is related to a time closer at hand in a startling way. In the second story we think to ourselves: Over a century what possible difference could a day make? And yet we are charmed by the marshal's sense of urgency. Every day counts, he seems to be saying, particularly when there are thousands of them. I have always loved this story, but the other one strikes me as much funnier. The idea that one could take such large numbers so personally, that one could sense doomsday so much more clearly if it were a mere *million* years away rather than a far-off *billion* years—hilarious! Who could possibly have such a gut-level reaction to the difference between two huge numbers?

Recently, however, there have been some even funnier big-number "jokes" in newspaper headlines. Jokes such as "Defense spending over the next four years will be $1 trillion" or "Defense Department overrun over the next four years estimated at $750 billion." The only thing that worries me about these jokes is that their humor is probably unnoticed by the average citizen. It would be a pity to allow such mirth-provoking notions to be appreciated only by a select few, and so I have decided it would be a good idea to devote some space to the necessary background knowledge, which also happens to be one of my favorite topics: the lore of very large (and very small) numbers.

I have always suspected that few people really know the difference between a million and a billion. To be sure, they know it well enough to sense the humor in the joke about when the earth will fall into the sun, but what the difference is *precisely*—that is something else. I heard a radio news announcer say, "The drought has cost California agriculture somewhere between 900,000 and a billion dollars." This kind of thing worries me. In a society where big numbers are commonplace we cannot afford to have such appalling number ignorance as we do. Or do we actually suffer from number numbness? Are we growing ever number to ever bigger numbers?

What do people think when they read headlines like the ones above? What do they think when they read about nuclear weapons with 20-kiloton yields? Or 60-megaton yields? Does the number really register or is it just another cause for a yawn? "Ho hum, I always knew the Russians could kill us all 20 times over. Now it's 200 times over, eh? Well, we can be thankful it's not 2,000."

What do people think about the fact that in some heavily populated areas of the U.S. it is typical for the price of a house to be a quarter of a million dollars? What do people think when they hear radio commercials for savings institutions telling them that if they invest now, they could have a million dollars on retirement? Can *everyone* be a millionaire? Do we now *expect* houses to take a fourth of a millionaire's fortune? What has become of the glittering connotations of the word millionaire?

I once taught a small beginning physics class on the 13th floor of Hunter College in New York City. From the window there was a magnificent view of the skyscrapers of midtown Manhattan. In one of the opening sessions I wanted to teach my students about estimates and significant figures, and so I asked them to estimate the height of the Empire State Building. In a class of 10 students not one student came within a factor of two of the correct answer (1,472 feet with the television antenna). Most of the estimates were between 300 and 500 feet. One person said 50 feet—a truly amazing underestimate; another said a mile. It turned out that this last person had actually calculated the answer, guessing 50 feet per story and 100 or so stories, thus getting about 5,000 feet. Where one person thought each story was 50 feet high, another thought the entire 102-story building was that high. This startling episode had a deep effect on me.

It is fashionable for people to write of the appalling illiteracy of this generation, particularly its supposed inability to write grammatical English. But what of the appalling "innumeracy" of most people, old and young, when it comes to making sense of the numbers that run their lives? As Senator Everett Dirksen once said, "A billion here, a billion there; soon you're talking real money."

The world is a big place, no doubt about it. There are a lot of people, a lot of needs, and it all adds up to a certain degree of incomprehensibility. That, however, is no excuse for not being able to understand, or even relate to, numbers whose purpose is to summarize in a few symbols some salient aspects of those huge realities. Most likely the readers of this article are not the ones I am worried about. It is nonetheless certain that every reader of this article knows many people who are ill at ease with large numbers of the kind that appear in the G.N.P., Government budgets, corporation budgets and so on. For people whose minds go blank when they hear something ending in "illion" all big numbers are the same, so that exponential explosions make no difference. Such an inability to relate to large numbers is clearly bad for society. It leads people to ignore big issues on the grounds that they are incomprehensible. Hence anything that can be done to correct rampant innumeracy is well worth doing. As I said above, I do not expect this article to reveal profound new insights to its readers (although I hope it will intrigue them); rather, I hope it will give them the materials and the impetus to convey a vivid sense of numbers to their friends and students.

As an aid to numerical horse sense I thought I would indulge in a small orgy of questions and answers. Let's go. How many letters are there in this magazine? Don't calculate—just guess. Did you say about a million? That has six zeros (1,000,000). If you did, that is a pretty sensible estimate. If you did not, were you too high or too low? In ret-

# "YOU CAN ALWAYS TRUST THE TASTE OF A BOLLA"

—Franco Bolla


Keeping up
with modern times.

Roller skates can gain you a little speed. But they can't help with the unprecedented pressures, problems and paperwork of this information age.

That kind of challenge calls for a special solution—a tool for modern times: the IBM Personal Computer.

## Getting things accomplished.

The most-often-asked questions about this tool are: how easy is it to use, and just what do I use it for?

The first question is simple to answer, because the IBM Personal Computer is surprisingly simple to learn.

The second question has no single answer, however. Because the uses for the IBM Personal Computer are as varied as your needs—as limitless as your imagination.

Storing form letters at the office, for example. Attendance records at school. Or tax records at home.

Keeping track of accounts receivable. Or when you last paid the orthodontist.

Completing sales reports as easily as your kids can revise and print book reports.

Or teaching students "dry" subjects in game form (while they learn there's more to computers than intergalactic battles).

## Plan on it.

Keeping up with modern times is tough enough. But intelligent planning is key. What if interest rates plummet? What if property taxes soar?

What's the latest stock report? What's the Aspen ski report?

With an IBM Personal Computer, and an expanding library of software, you can forecast. Plan ahead. Test contingencies. Tap information networks that bring the world into your lab. Your office. Your classroom and your living room.

## What's in store?

So to keep up with modern times, visit a ComputerLand® store, Sears Business Systems Center or IBM Product Center.

Ask your dealer what programs are available now. Get a demonstration.

### IBM PERSONAL COMPUTER SPECIFICATIONS

**User Memory**
16K-256K bytes*
**Permanent Memory**
(ROM) 40K bytes*
**Microprocessor**
16-bit, 8088*
**Auxiliary Memory**
2 optional internal diskette drives, 5¼", 160K bytes per diskette
**Keyboard**
83 keys, 6 ft. cord attaches to system unit*
10 function keys*
10-key numeric pad
Tactile feedback*

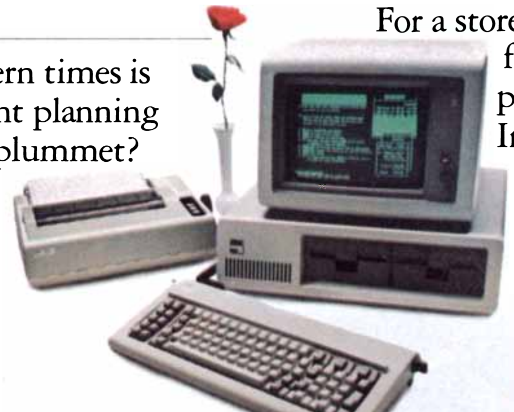
**Display Screen**
High-resolution*
80 characters x 25 lines
Upper and lower case
Green phosphor screen*
**Diagnostics**
Power-on self testing*
Parity checking*
**Languages**
BASIC, Pascal, FORTRAN, MACRO Assembler
**Printer**
Bidirectional*
80 characters/second
12 character styles, up to 132 characters/line*
9 x 9 character matrix*

**Color/Graphics**
*Text mode:*
16 colors*
256 characters and symbols in ROM*
*Graphics mode:*
4-color resolution: 320h x 200v*
Black & white resolution: 640h x 200v*
Simultaneous graphics & text capability*
**Communications**
RS-232-C interface
Asynchronous (start/stop) protocol
Up to 9600 bits per second

*ADVANCED FEATURES FOR PERSONAL COMPUTERS

And be sure to check the hardware features that set this personal computer apart. (Bring the box above to simplify comparison shopping.) You'll find that the quality, power and performance are what you'd expect from IBM. The price isn't.

For a store near you (or for information from IBM about quantity purchases) call (800) 447-4700. In Illinois, (800) 322-4400. In Alaska or Hawaii, (800) 447-0890. IBM®



# The IBM Personal Computer
# A tool for modern times

rospect, does your estimate seem farfetched? What intuitive cues suggest that a million is appropriate, rather than, say, a few thousand or a billion? Well, let's calculate it. Say there are about 100 pages filled with text (most likely something of an overestimate). How many words per page—500, 5,000? Certainly somewhere in between, it would seem. Let's just say 2,000. And how many letters per word? Oh, about six. So we have 100 × 2,000 × 6, which comes to a bit more than a million. Probably that is too high, but certainly it is close. But how could we have sensed this in advance?

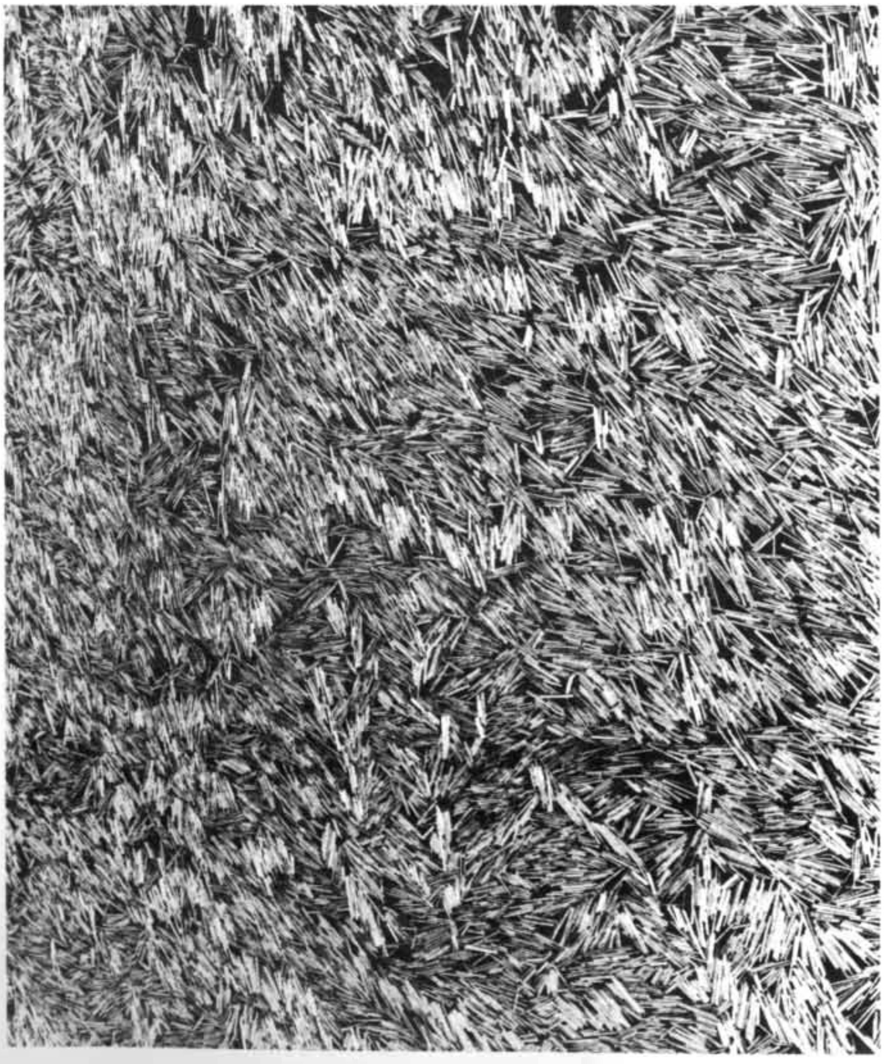
We were faced with a choice. Which is the most likely of the following 12 possibilities: (*a*) 10; (*b*) 100; (*c*) 1,000; (*d*) 10,000; (*e*) 100,000; (*f*) 1,000,000; (*g*) 10,000,000; (*h*) 100,000,000; (*i*) 1,000,000,000; (*j*) 10,000,000,000; (*k*) 100,000,000,000; (*l*) 1,000,000,000,000? (In the U.S. this last number with its 12 zeros is called a trillion; in most other countries it is called a billion. People in those countries reserve trillion for the truly enormous number 1,000,000,000,000,000,000, to us a quintillion.) Making such a guess is not much different from looking at the chairs in a room and guessing quickly if there are two or seven or 15. It is just that here what we are guessing at is the number of zeros in a numeral, that is, the logarithm (to the base 10) of the number. If we can develop a sense for the number of chairs in a room, why not as good a sense for the number of zeros in a numeral?

There is, of course, a difference. It is one thing to look at a numeral such as 100000000000000 and have an intuitive feeling, without counting, that it has somewhere around 12 zeros—certainly more than 10 and fewer than 15. It is quite another thing to look at an aerial photograph of a logjam and to be able to sense that there must be between three and five zeros in the decimal representation of the number of logs in the jam, in other words that 10,000 is the closest power of 10 and that 1,000 would definitely be too low and 100,000 too high. Such an ability is simply a form of perception that is one level of abstraction higher than the usual number perception. One level of abstraction should not, however, be too hard to handle.

The trick, of course, is practice. You have to get used to the idea that 10 is a very big number of zeros for a numeral to have, that five is pretty big and that three is almost graspable. Three zeros would take care of the number of students in your high school: 1,000, give or take a factor of three. (In numbers having just a few zeros we are always willing to forgive a factor of three or so in either direction, as long as we are merely estimating and not going for exactness.) Five zeros is the size of a typical county seat: 100,000 souls or so. Six zeros, that is, a million, is getting to be a large city—Minneapolis or San Diego or Brasília or Marseilles or Dar es Salaam. Seven zeros is getting huge: Shanghai, Mexico City, Seoul, Paris, New York. Just how many cities do you think there are in the world with a population of a million or more? Of them, how many do you think you have never heard of? What if you lowered the threshold to 100,000? How many towns are there in the U.S. with a population of 1,000 or less? Here is where practice helps.

Consider the famous claim made by the McDonald hamburger chain: "Over 25 billion served" (or whatever the number is these days). Is that credible? Well, if it were 250 billion, we could divide by the U.S. population more easily. (This is apparent provided you know the U.S. population is about 230 million. For the purposes of this discussion let us call it 250 million, or 2.5 times $10^8$—a common number everyone should know.) Let us say, then, that the claim is 250 billion. Then we would compute that 1,000 burgers have been cooked for every person in the U.S. But since we increased it by a factor of 10, let us now undo that—let us divide the answer by 10. Is it plausible that McDonald's has prepared 100 burgers for every person in the U.S.? It sounds reasonable to me; after all, they have been around for many years, and some families go there many times a year. I myself have had only a few Big Macs, but I know that this is unusual. Therefore the claim *is* plausible, and the fact that it is plausible makes it probable that it is quite accurate. Presumably McDonald's would not go to the trouble of updating their sign every so often if they were not trying to be accurate. I must say that if this earnest effort helps to reduce innumeracy, I approve highly of it.

Where do all those burgers come from? A staggering figure is the number of cattle slaughtered every day in the U.S. It comes to about 90,000. When I first heard it, that sounded amazingly high, but think about it. Maybe half a pound of meat per person per day. Once again the U.S. population—250 million—comes in handy. With half a pound of meat per person per day that comes to 100 million pounds of meat



*How many logs are there in this aerial photograph of a logjam in Oregon?*

WHOEVER YOU ARE, WHEREVER YOU GO,

# YOU HAVE A WORLD OF CHOICES WITH VISA.


## THE VISA EMBLEM— SYMBOL OF WORLDWIDE ACCEPTANCE.

**Introducing a new Visa card, with preferred banking and travel services.** Financial institutions around the world are introducing a new card, to give you preferred banking and travel services. Each card will feature the unique design of the financial institution and will display the Visa emblem in Gold, White and Gold—your assurance of recognition, acceptance and convenience. Worldwide.

**Visa Travelers Cheques: now with an expanded refund system that reports the loss of your cards, too.** The worldwide refund service protects your Visa Travelers Cheques if they are lost. If your payment cards are lost too, Visa will notify the proper institutions, at no charge, and arrange a cablegram or mailgram to any person you need to contact in the emergency.

**The Blue, White and Gold card.** The Blue, White, and Gold Visa is the most widely used payment card in the world. It is carried by 100 million people and accepted by over 3 million stores, hotels, airlines, car rentals and restaurants in 155 countries. It's good for cash at over 100,000 financial offices, too.

## YOU CAN HAVE IT THE WAY YOU WANT IT WITH VISA.

Each Visa service available at the option of participating financial institutions.


Photographed in Burghausen, Bavaria.

per day. Something like that, anyway. We certainly shall not worry about factors of two. How many tons is that? Divide by 2,000 to get 50,000 tons. An individual animal, however, does not yield a ton of meat. Maybe 1,000 pounds or so—half a ton. For each ton of meat that would mean two animals were killed. Hence we would get about 100,000 animals biting the dust every day to satisfy our appetite. Of course, we do not eat only beef, so that the true figure should be a bit lower. And that brings us back to where we started.

How many trees are cut down each week to produce the Sunday edition of *The New York Times?* Say a couple of million copies are printed, each one weighing four pounds. That comes to about eight million pounds of paper—4,000 tons. If a tree yielded a ton of paper, that would be 4,000 trees. I don't know much about logging, but we cannot be too far off in assuming a ton per tree. At worst it would be 200 pounds of paper per tree, and that would mean 40,000 small trees. The photograph on page 24 shows somewhere between 7,500 and 15,000 logs, as nearly as I can estimate. Therefore if we do assume 200 pounds of paper per tree, the logs in the photograph represent considerably less than half of one Sunday *Time's* worth of trees. We could go on to estimate the number of trees cut down every month to provide for all the magazines, books and newspapers every month, but I shall leave that one to you.

How many cigarettes are smoked in the U.S. every year? How many zeros? This is a classic "twelver"—on the order of a trillion. It is easy to calculate. Say that half of the people in the country are cigarette smokers: 100 million of them. (I know this is an overestimate; we shall compensate by reducing something else somewhere along the way.) Each smoker smokes—what? A pack per day? All right. That makes 20 cigarettes times 100 million: two billion cigarettes per day. There are 365 days per year. Let us say 250, since I promised to reduce something somewhere; 250 times two billion gives about 500 billion—half a trillion. It is just about on the nose, as it turns out; the last I looked (a few years ago) it was some 545 billion.

It is numbers of this size that we are dealing with when we talk about a Defense Department overrun of $750 billion over the next four years. A really fancy single-user computer (the kind I should like to have) costs approximately $75,000. With $750 billion to throw around we could give one to every person in New York City, which is to say we could buy about 10 million of them. Or we could give $1 million to every person in San Francisco and still have enough left over to buy a bicycle for everyone in China. There is no telling to what good uses we could put $750 billion. But instead it will go into ammunition, tanks, fighters, war games, missile systems, jet fuel and so on. An interesting way to spend $750 billion, but I can think of better ways.

Let us think of some other kinds of big numbers. Did you know your retina has about 60 million cells in it, each of which responds to some particular kind of stimulus? And they feed their signals back into your brain, which is now thought to consist of somewhere around 100 billion neurons, or nerve cells. The number of glia, smaller supporting cells in the brain, is about 10 times as large. That may sound big, but in your body there are estimated to be about 60 or 70 trillion cells. Each one of them contains millions of components working together. Take the protein hemoglobin, which transports oxygen in the bloodstream. We each have about six billion trillion (that is, six thousand million million million) copies of the hemoglobin molecule inside us, with something like 400 trillion (400 million million) being destroyed every second and another 400 trillion being made!

The number of hemoglobin molecules in the body is $6 \times 10^{21}$. It is a curious fact that over the past year or two nearly everyone has become familiar, implicitly or explicitly, with a number nearly as big, namely the number of different possible configurations of Rubik's cube. This number—let us call it Rubik's constant—is about $4.3 \times 10^{19}$. The Ideal Toy Corporation was far less daring than McDonald's. They softened the blow, calling it merely "more than three billion," a pathetic and euphemistic underestimate. Moreover, if you are working in the Rubik's cube "supergroup," where the orientations of face centers matter, then Rubik's "superconstant" is 2,048 times bigger, or about $9 \times 10^{22}$!

The hemoglobin number and Rubik's superconstant are *really* big. How about some smaller big ones, to come back to earth for a moment? All right—how many people would you say are falling to earth by parachute at this moment (a perfectly typical moment, presumably)? How many English words do you know? How many murders are there in Los Angeles County every year? In Japan? These last two give quite a shock when put side by side: Los Angeles County, about 2,000; Japan, about 900.

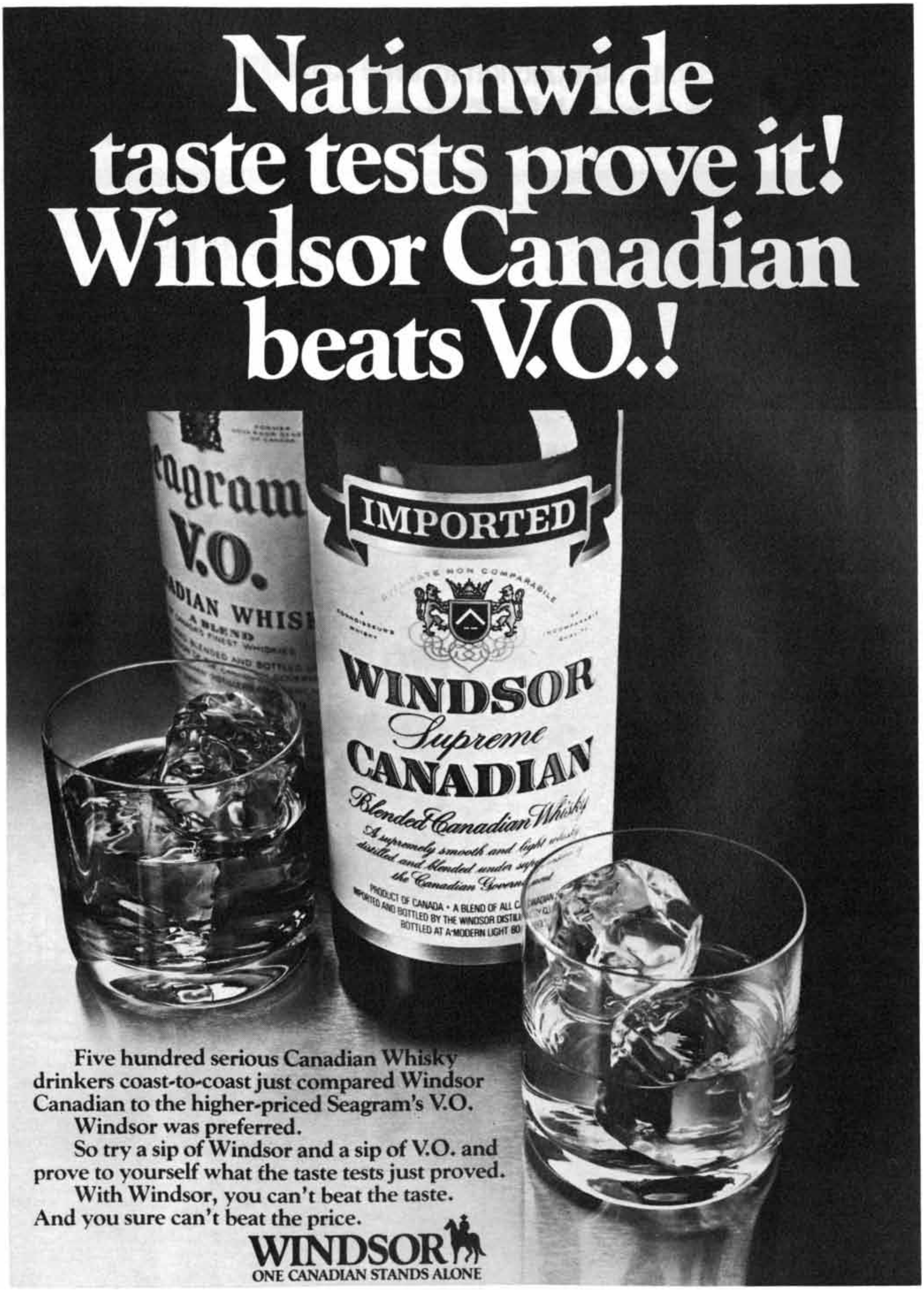
Speaking of yearly deaths, here is one we all sweep under the rug, it seems: 50,000 dead per year (in this country alone) in car accidents. Can you imagine how we would react if someone said to us today: "Hey, I've come up with a really nifty invention. Unfortunately it has a minor defect—every 12 years or so it will wipe out about as many people as live in San Francisco. But wait a minute! Don't go away! The rest of you will love it, I promise!" Now these statistics are accurate for cars. And yet we seldom hear people saying, "No cars is good cars." Somehow collectively we are willing to accept the loss of 50,000 lives per year without any serious worry.

I said I would be a little lighter. All right. Light consists of photons. How many photons per second does a 100-watt bulb put out? About $10^{20}$—another biggie. Is it bigger or smaller than the number of grains of sand on a beach? What beach? Say a stretch of beach a mile long, 100 feet wide and six feet deep. What would you estimate? Now calculate it. How about trying the number of drops in the Atlantic Ocean? Then try the number of fish in the ocean. Which are there more of, fish in the sea or ants on the surface of the earth? Atoms in a blade of grass or blades of grass on the earth? Blades of grass or insects? Leaves on a typical oak tree or hairs on a human head? How many raindrops fall on your town in one second during a downpour?

How many copies of the Mona Lisa have ever been printed? Let us try this one together. Probably it is printed in magazines in the U.S. a few dozen times per year. Say each of the magazines prints 100,000 copies. That makes a few million copies per year in American magazines, but then there are books and other publications. Maybe we should double or triple our figure for the U.S. To take into account other countries we can multiply it again by three or four. Now we have hit about 100 million copies per year. Let us assume this held true for each year of this century. That would make nearly 10 billion copies of the Mona Lisa! Probably we have made some mistakes along the way, but give or take a factor of 10 that is about what the number is.

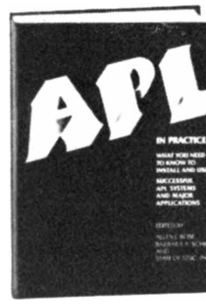
Give or take a factor of 10? A moment ago I was saying that a factor of three was forgivable, but here I am forgiving myself *two* factors of three, that is, an entire order of magnitude (a factor of 10). Well, we are now dealing with larger numbers ($10^{10}$ instead of $10^5$), and so it is permissible. This brings up a good rule of thumb. Say a factor of three is permissible for each factor of 100,000. That means we are allowed to be off by a factor of 10—one order of magnitude—when we get up to sizes such as 10 billion, or by a factor of 100 or so (two orders of magnitude) when we get up to the square of that, which is $10^{20}$, about 2.5 times the size of Rubik's constant. This means it would have been forgivable if Ideal had said "More than a billion billion combinations," since they would have been off by a factor of only 40—about 1.5 orders of magnitude—which is within our limits.

Why should we be content either with an estimate that is only 1 percent of the actual number or with an estimate that is 100 times too big? Well, because if you consider the base-10 logarithm


Nationwide taste tests prove it! Windsor Canadian beats V.O.!
IMPORTED
WINDSOR
Supreme
CANADIAN
Blended Canadian Whisky
V.O.
Five hundred serious Canadian Whisky drinkers coast-to-coast just compared Windsor Canadian to the higher-priced Seagram's V.O.
Windsor was preferred.
So try a sip of Windsor and a sip of V.O. and prove to yourself what the taste tests just proved.
With Windsor, you can't beat the taste.
And you sure can't beat the price.
WINDSOR
ONE CANADIAN STANDS ALONE

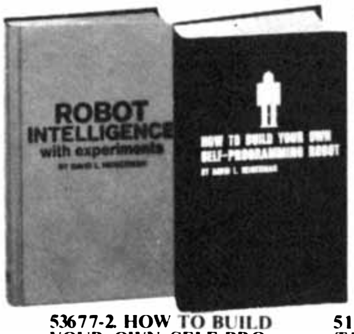

**39960. COMPUTER-ASSISTED DATA BASE DESIGN.** *George V. Hubbard.* A complete look at data models, editing, design tactics, root anchor points and much more. **$24.95**


**39977-2. COMPUTER CONTRACT NEGOTIATIONS.** *Auer and Harris.* Shows the computer professional how to handle every phase of successful negotiation, from preparation through signing the final agreement. *Counts as 2 of your 3 books.* **$34.50**

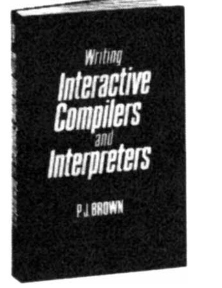



**34115. APL IN PRACTICE.** *Edited by Allen J. Rose et al.* Covers such topics as writing maintainable APL codes, evaluating telecommunications networks, using APL for construction accounting and more. **$25.00**

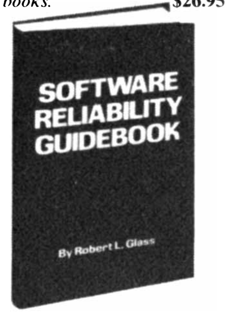

**40045-2. CONVERSATIONAL LANGUAGES.** *Kupka and Wilsing.* Covers interactive program development, languages based on JOSS, syntax and design of conversational languages and more. *Counts as 2 of your 3 books.* **$27.75**



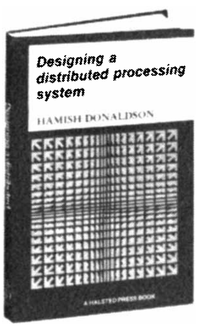

**53677-2. HOW TO BUILD YOUR OWN SELF-PROGRAMMING ROBOT/ROBOT INTELLIGENCE . . . WITH EXPERIMENTS.** *David L. Heiserman.* Two books. *Counts as 2 of your 3 books.* **$29.90**




**87970-2. WRITING INTERACTIVE COMPILERS AND INTERPRETERS.** *P.J. Brown.* Includes how to design an internal language, error detection and diagnosis in the run-time system, and much more. *Counts as 2 of your 3 books.* **$26.95**

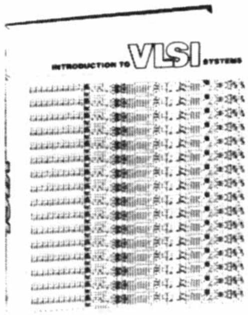



**79167. SOFTWARE RELIABILITY GUIDEBOOK.** *Robert L. Glass.* Analysis of technical tools available for software design and implementation. Includes applications for small operations to mammoth projects. **$18.95**



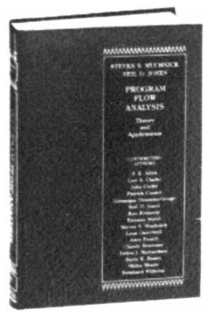

**42053-2. DESIGNING A DISTRIBUTED PROCESSING SYSTEM.** *Hamish Donaldson.* An important casebook on how to manage projects successfully and a technical reference for design problems. *Counts as 2 of your 3 books.* **$34.95**


**51955-2. A GUIDE TO THE SUCCESSFUL MANAGEMENT OF COMPUTER PROJECTS.** *Hamish Donaldson.* Provides invaluable advice on project management and documentation, systems design strategy, and programming. *Counts as 2 of your 3 books.* **$34.95**



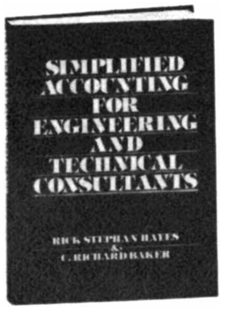

**56505-2. INTRODUCTION TO VLSI SYSTEMS.** *Mead and Company.* Offers computer professionals a practical view of technical problems at the leading edge of integrated circuit technology. *Counts as 2 of your 3 books.* **$27.95**


**34440. THE ART OF ELECTRONICS.** *Horowitz and Hill.* Emphasizes circuit intuition and non-mathematical design techniques. Packed with diagrams, schematics, and illustrations. **$24.95**




**70691. PROGRAM FLOW ANALYSIS.** *Edited by Muchnick and Jones.* A comprehensive look at a statistical method which can yield more efficient and reliable software. **$23.50**


**44160. ELECTRONIKIT: MULTI-EFFECT SOUND SYNTHESIZER. Vol. I.** *James Barbarello.* Explains how to produce a sound synthesizer and how to play it through your own stereo or hi-fi unit. Softbound. **$10.95**




**78333-2. SIMPLIFIED ACCOUNTING FOR ENGINEERING AND TECHNICAL CONSULTANTS.** *Hayes and Baker.* Covers partnership share accounting, compensation plans, time accounting, systems of billing and more. *Counts as 2 of your 3 books.* **$27.95**

## Plus more books to choose from

**48494. FILE TECHNIQUES FOR DATA BASE ORGANIZATION IN COBOL.** *Johnson and Cooper.* Includes technical considerations in choosing a file structure, data structures and more. **$20.95**

**56996. JEWELS OF FORMAL LANGUAGE THEORY.** *Arto Salomaa.* A compact presentation of language theory covering such topics as repetitions, DOL systems and w-words, codes and equality sets and more. **$24.95**

**63060. MORE CHESS AND COMPUTERS: The Micr computer Revolution, The Challenge Match.** *Levy and Newborn.* Softbound. **$12.95**

**42275. DIGITAL SYSTEMS: Hardware Organizations and Design.** *Hill and Peterson.* A classic reference on hardware organization and system architecture. **$23.95**

**79155. SOFTWARE INTERPRETERS FOR MICROCOMPUTERS.** *Thomas C. McIntire.* How to design a reliable software interpreter for any microcomputer system. **$23.95**

**39890-2. COMPUTER DATA-BASE ORGANIZATION.** *James Martin.* An invaluable planning tool by the world's foremost computer author. Contains over 200 diagrams. *Counts as 2 of your 3 books.* **$27.95**

**54455. IMS PROGRAMMING TECHNIQUES: A Guide to Using DL/I.** *Kapp and Leben.* Spells out specific techniques to write application programs in ANS COBOL, PL/1, or Assembler Language in an IMS DL/1 data base environment. **$17.95**

**85847. VIDEOTEXT: The Coming Revolution in Home/Office Information Retrieval.** *Edited by Efrem Sigel.* Examines the explosive growth of television-based videotext systems. **$24.95**

**70230. PRINCIPLES OF SOFTWARE ENGINEERING AND DESIGN.** *Marvin V. Zelkowitz et. al.* A practical guide which shows you how to handle everything from small programs all the way through the design of large, complex software systems. **$24.50**

**60385. MANAGING THE SYSTEMS DEVELOPMENT PROCESS.** *Charles L. Biggs et al.* Provides every computer professional with all the management tools and procedures necessary to keep the systems development process under control, on budget, and on time. **$24.95**

**42303. A DISCIPLINE OF PROGRAMMING.** *Edsger W. Dijkstra.* A programming classic. **$22.95**

**59940. MANAGEMENT INFORMATION SYSTEMS.** *Paul Sigel.* Planning, forecasting, modeling, decision-making, establishing MIS roles, and implementation of MIS—are all thoroughly covered. **$16.95**

**81857. SYSTEMATIC PROGRAMMING.** *Niklaus Wirth.* A classic work on programming as an exercise in constructing algorithms by the noted author. **$23.95**

**70745. A PROGRAMMER'S GUIDE TO COBOL.** *William J. Harrison.* **$18.95**

**70725-2. THE PROGRAMMER'S ANSI COBOL REFERENCE MANUAL.** *Donald A. Sordillo.* Alphabetized and heavily cross-referenced, provides a complete reference to COBOL verbs, syntax concepts and uses. *Counts as 2 of your 3 books.* **$27.95**

**32485. ALGORITHMS + DATA STRUCTURES = PROGRAMS.** *Niklaus Wirth.* "A valuable, novel contribution to the computing science literature."—Edsger Dijkstra. **$20.95**

**62651. MICROPROCESSORS AND MICROCOMPUTER SYSTEMS.** *Guthikonda V. Rao.* Coverage of selection, design and testing. **$24.50**

**42420. THE DISTRIBUTED SYSTEM ENVIRONMENT: Some Practical Approaches.** *Grayce M. Booth.* Covers horizontal and hierarchical distributed structures, networking and much more. **$21.95**

**42213. DICTIONARY OF BUSINESS AND MANAGEMENT.** *Jerry M. Rosenberg.* Over 8000 entries. **$24.95**

**79240. SOFTWARE PSYCHOLOGY: Human Factors in Computer and Information Systems.** *Ben Shneiderman.* **$24.95**

If the reply card has been removed, please write to
**The Library of Computer and Information Sciences**
**Dept. 7-BB4, Riverside, N.J. 08075**
to obtain membership information and an application.

of the number—the number of zeros—when we say 18 and the real answer is 20, we are off by only 10 percent!

What entitles us to cavalierly dismiss the magnitude itself and to switch our focus to its logarithm (its *order* of magnitude)? Well, when numbers get this big, we have no choice. Our perceptual reality begins to shift. We simply *cannot* visualize the actual quantity. The numeral—that string of digits—takes over: our perceptual reality becomes one of numbers of zeros. When does this shift take place? It begins when we can no longer see, in our mind's eye, a collection of the right order of magnitude. For me this perceptual logjam begins at about $10^4$—the size of the logjam I remember in the photograph.

There are other ways to grasp $10^4$, such as the number of soup cans that would fill a 50-foot shelf in a supermarket. Numbers much bigger than that I simply cannot visualize. The number of tiles that line the Lincoln Tunnel between Manhattan and New Jersey is so enormous that I cannot easily picture it. (It is on the order of a million.) In any case somewhere around $10^4$ or $10^5$ my ability to visualize begins to fade and to be replaced with that second-order reality of the number of digits (or to some extent with number names such as million, billion and trillion). Why it happens at this size and not, say, at 10 million or 200 must have to do with evolution and the role that the perception of vast arrays plays in survival. It is a fascinating philosophical question but one we cannot hope to answer here.

In any case a good rule of thumb is this: Your estimate should be within 10 percent of the correct answer—but this need apply only *at the level of your perceptual reality*. Therefore you are excused if you guessed that Rubik's cube has $10^{18}$ positions, since 18 is pretty close to $19\frac{1}{2}$, which is about what the number of digits is. (The factor of 4.3 counts for a bit more than half a digit, since each factor of 10 contributes a full digit, whereas a factor of 3.16, the square root of 10, contributes half a digit.)

If, perhaps, you were to start dealing with numbers with millions or billions of digits, the numerals themselves (the endless strings of digits) would cease to be visualizable and your perceptual reality would take another leap upward in abstraction—to the number that counts the digits in the number that counts the digits in the number that counts the objects concerned. Needless to say, such third-order perceptual reality is highly abstract. Moreover, it occurs very seldom, even in mathematics. Still, you can imagine going far beyond it. Fourth- and fifth-order perceptual realities would quickly yield, in our purely abstract imagination, to tenth-, hundredth- and millionth-order perceptual realities.

By this time, of course, we would have lost track of the exact number of levels we had shifted, and we would be content with a mere *estimate* of that number (accurate to within 10 percent, of course). "Oh, I'd say about two million levels of perceptual shift were involved here, give or take a couple of hundred thousand" would be a typical comment for someone dealing with such unimaginably unimaginable quantities. You can see where this is leading: to multiple levels of abstraction in talking about multiple levels of abstraction. If we were to continue our discussion just one zillisecond longer, we would find ourselves smack-dab in the middle of the theory of recursive functions and algorithmic complexity, and that would be too abstract. So let's drop the topic right here.

Related to this idea of huge numbers of digits, but more tangible, is the computation of the famous constant pi. How many digits have so far been calculated by machine? The answer (as far as I know) is one million. It was done in France a few years ago, and the million digits fill an entire book. Of these million how many have been committed to human memory? The answer strains credulity: 20,000, according to the latest *Guinness Book of World Records*. I myself once learned 380 digits of pi, when I was a crazy high school kid. Later I met several other people who had outdone me. All of us had forgotten most of the digits, but we all remembered the first 100 solidly, and so we would occasionally recite them in unison—a rather esoteric pleasure.

What would you think if someone said that the entire book of a million digits of pi had been memorized by someone? I know I would dismiss the claim out of hand. I once heard that Jerry Lucas, the memory and basketball whiz, had said he knew the entire Manhattan telephone directory. Here we have a good example of how innumeracy can breed gullibility. Can you imagine what memorizing the Manhattan telephone directory would involve? To me it seems about two orders of magnitude beyond credibility. To memorize one page seems fabulously difficult. To memorize 10 pages seems at about the limit of credibility. Incidentally, memorizing the entire Bible (which I have occasionally heard claimed) seems to me about equivalent to memorizing 10 pages of the phone book. To have memorized 1,500 dense pages of telephone numbers, addresses and names is literally beyond belief.

Let us consider the scale on which one measures a given phenomenon. Take pitch in music. If you look at a piano keyboard, you will see a linear scale along which pitch can be measured. The natural thing to say is, "This A is nine semitones higher than that C, and the C is seven semitones higher than that F, hence the A is 16 semitones higher than the F." It is an additive scale. This means you could assign each note a number, and then the distance from one note to another would be the difference between their note numbers.

On the other hand, acoustically speaking, each pitch is better described in terms of its frequency than it is in terms of its position on a keyboard. Frequency is another number associated with each key on the keyboard. At the bottom of the keyboard that low A vibrates about 27 times per second, whereas the C three semitones above it vibrates about 32 times per second. You might be inclined to guess, then, that in order to jump up three semitones one should always add five cycles per second. Not so. You should always *multiply* by about 32/27 instead.

Thus when you have gone up one octave, which is 12 semitones, your pitch has been multiplied by 32/27 four times in a row, which is 2. Actually the fourth power of 32/27 is not quite 2, and since an octave represents a ratio of exactly 2, 32/27 must be a slight underestimate. That, however, is beside the point. The point is that the natural operations for comparing frequencies are multiplication and division, whereas the natural operations for note numbers on a keyboard are addition and subtraction. What this means is that the note numbers are logarithms of the frequencies. Here is a case where we think naturally in logarithms!

Logarithmic thinking happens when you perceive only a linear increase even if the thing itself doubles in size. For instance, have you ever marveled at the fact that dialing a mere seven digits can connect any telephone to any other in the New York metropolitan area, where some 10 million people live? Suppose New York were to double in population. Would you then have to add seven more digits to each phone number, making 14-digit numbers, in order to reach those 20 million people? Of course not. Adding seven more digits would *multiply* the number of possibilities by 10 million. In fact, adding merely three digits (the area code in front) enables you to reach any phone number in North America. This is simply because each new digit creates a tenfold increase in the number of phones reachable. Three more digits will always multiply your network by a factor of 1,000: three orders of magnitude. Hence the length of a phone number—the quantity directly perceived by you when you are annoyed at how long it takes to dial a long-distance number—is a logarithmic measure of the size of the network you are embedded in. That is why it is preposterous to see huge long numbers of 20 or 30 digits used as codes for people or products when without any doubt a few digits would suffice.

I was once sent a bill asking that I transfer a fee to account No. 60802-

620-1-1-721000-421-01062 in a bank in Yugoslavia. For a while this held my personal record for absurdity of numbers encountered in business transactions. Recently, however, I was sent my car-registration form, at the bottom of which I found this enlightening constant: 01010136121820030107001-4263117241512003603600030002. For good measure it was followed a few blank spaces later by 19283.

One place where we think logarithmically is in number names. Up to a certain point we in America have a new name every three zeros. From thousand to million to billion to trillion. Each jump is "the same," in a sense. That is, a billion is exactly as much bigger than a million as a million is bigger than a thousand. Or a trillion is to a billion exactly as a billion is to a million. (On the other hand, is $10^{103}$ to $10^{100}$ exactly as a million is to a thousand? I would be inclined to say no. It is a little tricky because of the shifts in perceptual reality. We are thinking less here of pure mathematics than we are of the psychology of perception of abstractions.)

In any case, we seem to run out of number names at about a trillion. To be sure, there are some official names for bigger numbers, but they are about as familiar as the names of extinct dinosaurs: quadrillion, octillion, vigintillion, brontosillion, triceratillion and so on. We are simply not familiar with them, since they died off a dinosillion years ago. Even "billion" presents cross-cultural problems, as I mentioned above. Can you imagine what it would be like if in Britain "hundred" meant 1,000? The fact is that when numbers get too large, people's imaginations balk. It is too bad, though, that a trillion is the largest number with a common name. What is going to happen when the defense budget gets even bigger? Will we just get number?

The speed of automatic computation is something whose progress is best charted logarithmically. Over the past several decades the number of primitive operations (such as addition or multiplication) that a computer can carry out per second has multiplied about tenfold about every seven years. Nowadays it is some 100 million operations per second or, on the fanciest machines, a little more. Around 1975 it was about 10 million operations per second. In the later 1960's one million operations per second was extremely fast. In the early 1960's it was 100,000 operations per second, in the mid-1950's 10,000, in the late 1940's 1,000 and in the early 1940's 100.

In fact, in the early 1940's Nicholas Fattu was the leader of a team at the University of Minnesota that was working for the Army Air Force on some statistical calculations involving large matrixes (about 60 × 60). He brought about 10 people together in a room, each of whom was given a "Monroe-matic" desk calculator. These people worked full time for 10 months in a coordinated way, carrying out the computations and cross-checking each other's results as they went along. About 20 years later, out of curiosity, Professor Fattu redid the calculations on an IBM 704 in 20 minutes. He found that the original team had made two inconsequential errors. Nowadays, of course, the entire thing could be done on a big computer in a second or two.

Still, modern computers are not infinitely fast. The notorious computer proof of the four-color theorem, done at the University of Illinois a few years ago, took 1,200 hours of computer time. When you convert that into days, it sounds more impressive: 50 full 24-hour days. If the computer was carrying out 20 million operations per second, that would come to $10^{14}$, or 100 trillion, primitive operations—a couple of hundred for every cigarette smoked that year in the U.S. Whew!

A computer doing a billion operations per second would really be moving along. Imagine breaking up one second into as many tiny fragments as there are seconds in 30 years. That is how tiny a nanosecond—a billionth of a second—is. To a computer a second is a lifetime! Of course, the computer is dawdling compared with the events inside the atoms that compose it. Take one atom. A typical electron circling a typical nucleus makes about $10^{15}$ orbits per second, which is to say a million orbits per nanosecond. From an electron's-eye point of view the computer is as slow as molasses in January.

Actually the electron has two eyes with which to view the situation. It has both an orbital cycle time and a rotational cycle time, since it is spinning on its own axis. Now, strictly speaking, "spin" is just a metaphor at the quantum level, and so you should take the following with a big grain of salt. Nevertheless, if you imagine an electron to be a classically (non-quantum-mechanically) spinning sphere, you can calculate its rotation time from its known spin angular momentum (which is about Planck's constant, or $10^{-34}$ joule-second) and its radius (which we can equate with its Compton wavelength, which is about $10^{-10}$ centimeter). The spin time turns out to be about $10^{-20}$ second. In other words, every time the superfast computer adds two numbers, every electron inside it has pirouetted on its own axis about 100 billion times. (If we took the classical radius of the electron instead, we would have the electron spinning at about $10^{24}$ times per second—enough to make one dizzy! Since this figure violates both relativity and quantum mechanics, however, let us be content with the first figure.)

At the other end of the scale there is the slow, stately twirling of our galaxy, which makes a leisurely complete turn every 200 million years or so. And within the solar system the planet Pluto takes about 250 years to complete an orbit of the sun. Speaking of the sun, it is about a million miles across and has a mass on the order of $10^{30}$ kilograms. The earth is a featherweight in comparison, a mere $10^{24}$ kilograms. We should not forget that there are some stars—red giants—whose diameter is so great that they would engulf the orbit of Jupiter. Of course, such stars are very tenuous, something like cotton candy on a cosmic scale. In contrast, some stars—neutron stars—are so tightly packed that if one could remove from one of them a cube a millimeter on an edge, its mass would be about half a million tons, equal to the mass of the heaviest oil tanker ever built, fully loaded!

These large and small numbers are so far beyond our ordinary comprehension that it is virtually impossible to keep on being more amazed. The numbers are genuinely beyond understanding, unless one has developed a vivid feeling for exponents. And even with such a feeling it is hard to give the universe its awesome due for being so extraordinarily huge and at the same time so extraordinarily fine-grained. Number numbness sets in early these days. Most people seem entirely unfazed by words such as billion and trillion; they become simply synonyms for the meaningless zillion.

That hit me particularly hard a few minutes after I had finished a draft of this column. I was reading the paper, and I came across an article on the subject of nerve gas. It stated that President Reagan expected the expenditures for nerve gas to come to about $800 million in 1983 and $1.4 billion in 1984. I was upset, but I caught myself being thankful that it was not $10 billion or $100 billion. Then, all at once, I really felt ashamed of myself. How could I have been relieved by the figure of a "mere" $1.4 billion? How could my thoughts have become so dissociated from the underlying reality? One billion for nerve gas is not merely lamentable; it is odious. We cannot afford to become number-number than we are. We need to be willing to be jerked out of our apathy, because this kind of joke is in very poor taste.

Combating number numbness is basically not so hard. It simply involves getting used to a second set of meanings for *small* numbers, namely the meanings of numbers between, say, 5 and 20, when they are used as exponents. It would seem revolutionary for newspapers to adopt the convention of expressing large numbers as powers of 10, yet to know that a number has 12 zeros is *more* concrete than to know that it is called a trillion.

I wonder what percentage of our population, if people were shown the numerals 314,159,265,358,979 and 271,828,182,845, would recognize that the former magnitude is about 1,000 times greater than the latter. I am afraid that the vast majority would not be able to see it and would not be able to read these numbers out loud. If this is the case it is something to be worried about.

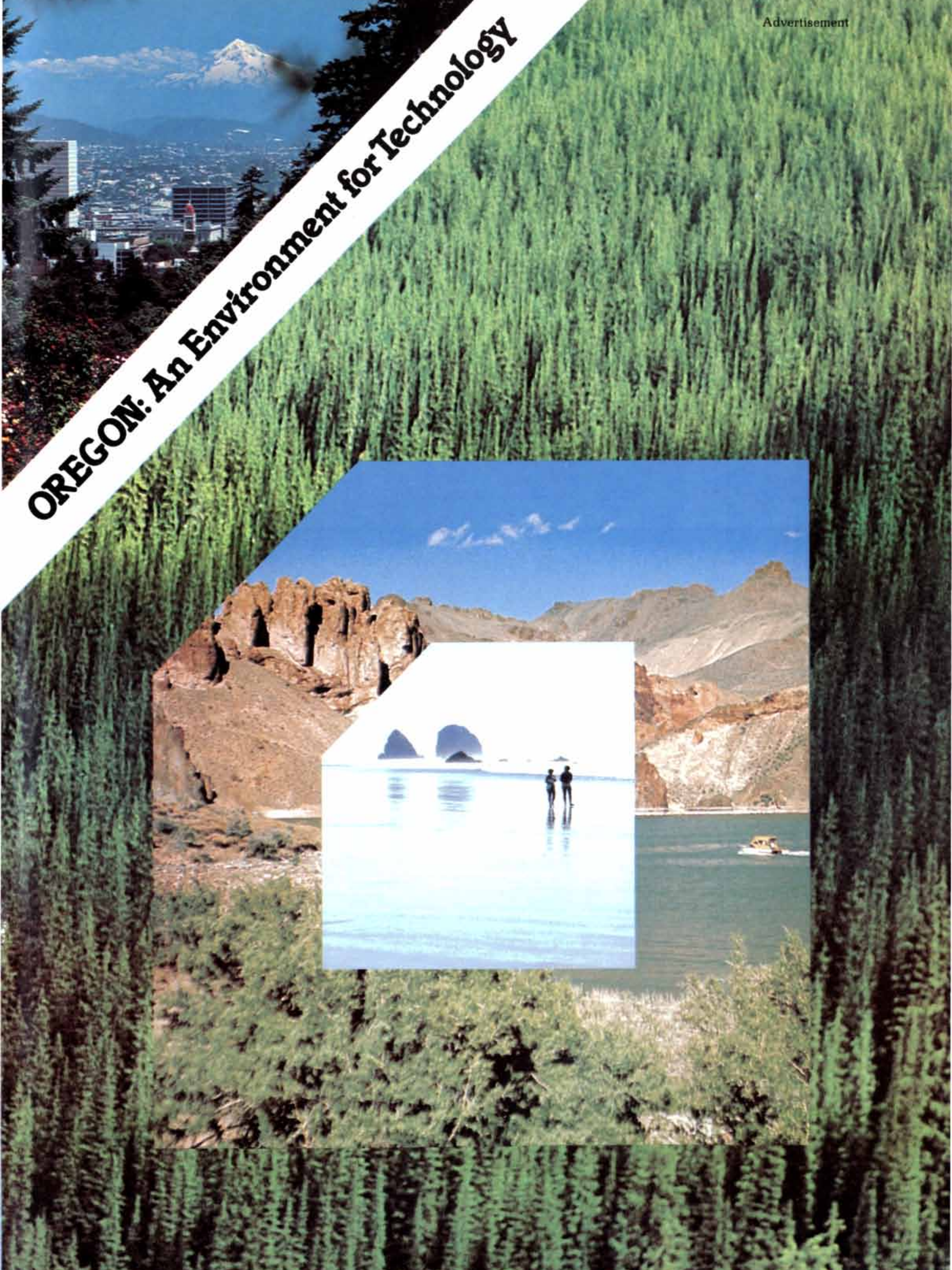
One book that attempts to combat such numbness, a book filled with humility before some of the astounding magnitudes we have been discussing, is called *Cosmic View: The Universe in Forty Jumps*, by a Dutch teacher, the late Kees Boeke. In his book Boeke takes us on an imaginary voyage in pictures, in which each step is an exponential one, involving a factor of 10 in linear size. From our own size there are 26 upward steps and 13 downward steps. It is probably not coincidental that the book was written by a Dutchman, since the Dutch have long been internationally minded, living as they do in a small and vulnerable country among many languages and cultures. Boeke closes in what therefore seems to me a characteristically Dutch way, by pleading that his book's journey will help to make people better realize their place in the cosmic scheme of things and in this way contribute to drawing the world closer together. Since I find his conclusion eloquent, I should like to close by quoting from it:

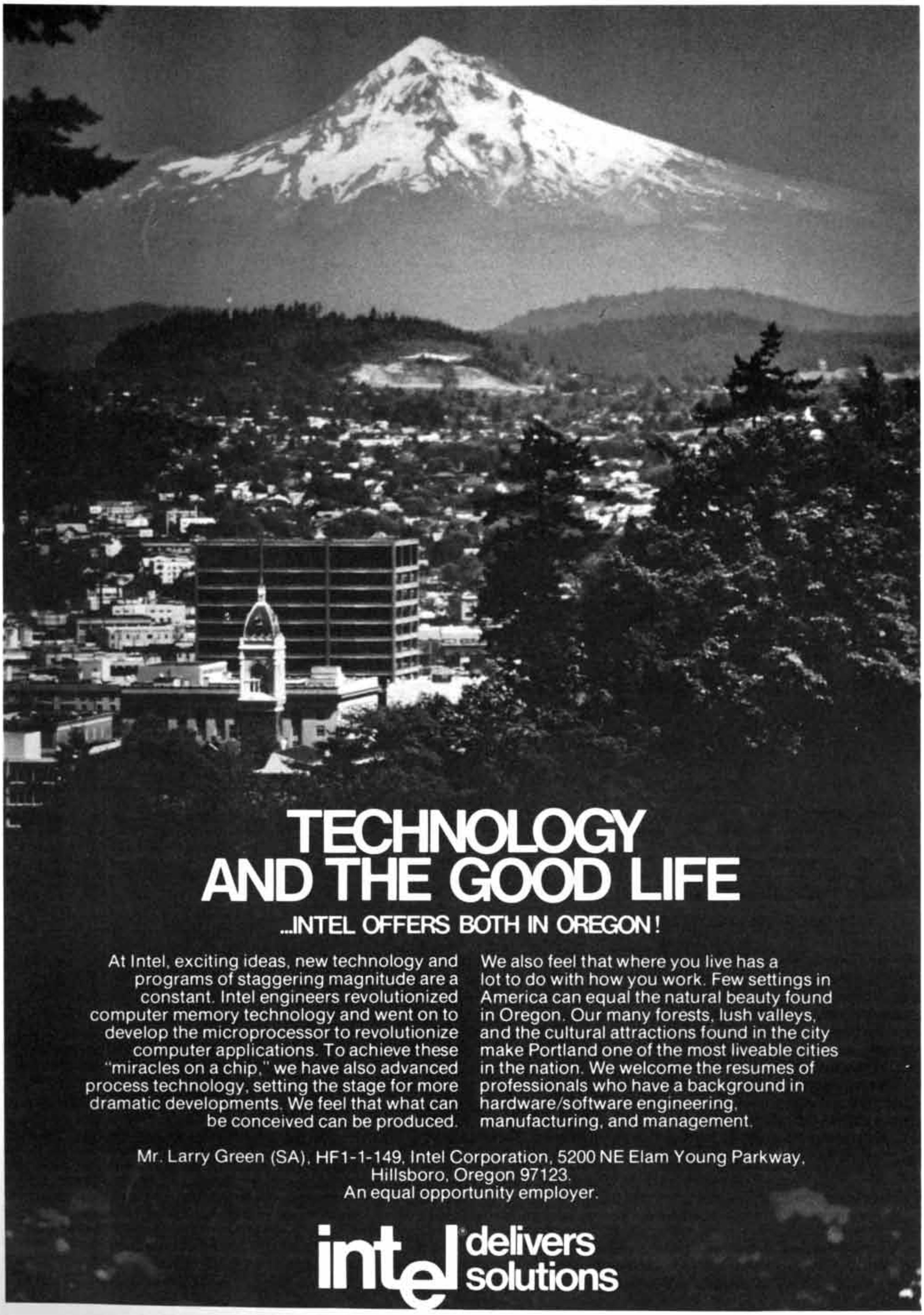
"When we thus think in cosmic terms, we realize that man, if he is to become really human, must combine in his being the greatest humility with the most careful and considerate use of the cosmic powers that are at his disposal.

"The problem, however, is that primitive man at first tends to use the power put in his hands for himself, instead of spending his energy and life for the good of the whole growing human family, which has to live together in the limited space of our planet. It therefore is a matter of life and death for the whole of mankind that we learn to live together, caring for one another regardless of birth or upbringing. No difference of nationality, of race, creed, or conviction, age or sex may weaken our effort as human beings to live and work for the good of all.

"It is therefore an urgent need that we all, children and grown-ups alike, be educated in this spirit and toward this goal. Learning to live together in mutual respect and with the definite aim to further the happiness of all, without privilege for any, is a clear duty for mankind, and it is imperative that education be brought onto this plane.

"In this education the development of a cosmic view is an important and necessary element; and to develop such a wide, all-embracing view, the expedition we have made in these 'forty jumps through the universe' may help just a little. If so, let us hope that many will make it!"


OREGON: An Environment for Technology


TECHNOLOGY
AND THE GOOD LIFE
...INTEL OFFERS BOTH IN OREGON!
At Intel, exciting ideas, new technology and programs of staggering magnitude are a constant. Intel engineers revolutionized computer memory technology and went on to develop the microprocessor to revolutionize computer applications. To achieve these "miracles on a chip," we have also advanced process technology, setting the stage for more dramatic developments. We feel that what can be conceived can be produced.
We also feel that where you live has a lot to do with how you work. Few settings in America can equal the natural beauty found in Oregon. Our many forests, lush valleys, and the cultural attractions found in the city make Portland one of the most liveable cities in the nation. We welcome the resumes of professionals who have a background in hardware/software engineering, manufacturing, and management.
Mr. Larry Green (SA), HF1-1-149, Intel Corporation, 5200 NE Elam Young Parkway, Hillsboro, Oregon 97123.
An equal opportunity employer.
intel delivers solutions

# Oregon

Almost to its surprise, Oregon, goal line of the great westward migration, is emerging as a high technology province.

Oregonians are still alive who entered the state in canvas-roofed ox wagons. Oregon Trail ruts still furrow some eastern counties; grooves still scar standing Douglas fiq trees, burned when Conestogas were snubbed to them, roping down the Cascades.

A state with a dozen wilderness areas, in the west it has forests to the high-tide mark, and in the east, mustang overpopulation. It is defined by the Pacific Ocean, the volcanic Cascade Range, and the violent Snake and rolling Columbia rivers.

And in the four of its counties, electronics is the business of the biggest employer.

World technology is measured and shaped by the instruments streaming from Tektronix shops. The computer industry watches introductions of Oregon-designed, Oregon-built Intel microcomputers as if they were moves in a world championship chess game—which they may be.

Oil tankers and fishing boats alike navigate by Oregon-built Morrow Loran computers. The western world builds it nuclear reactors with Oregon zirconium, and its high-speed aircraft with Oregon titanium.

Surgeons circumvent cardiac problems with multiple bypass techniques that orginated in Oregon, and many of the pacemakers or pain-blankers they implant are made in Oregon. And the nation's grocers use Oregon-made laser checkstand scanners to read prices and inventory data.

Before the end of the century, high technology will displace forest products as Number One employer, predicts the Bonneville Power Administration, power planner for the region. The forecast is credible, particularly after a drive along Interstate 5, the north-south link with the California and Washington industrial hubs, where the "spec" warehouses are clustered at every population center.

High technology manufacturing for a world market isn't new to Oregon.

What's new is the entry of a number of highly visible California manufacturers, large and small, compounding the growth and multiplication of native companies. The Californians, pinched at home by their own bigness, have been looking increasingly to the Pacific Northwest for expansion and a less jaded workforce, and escape from housing costs that make it difficult to attract young professionals.

The California entries include some of the biggest, but most are small and consider themselves growth companies.

A kind of Mason-Dixon line runs across the middle of the state. Below it, electronics plants spiritually face south, and their products are likely to be subassemblies to be flown or bused to California.

Above it, the output tends to stand alone, competing directly on the world market without further handling. Even Intel and Hewlett-Packard, the biggest California employers in the upper half, operate there as autonomous profit centers, with their own R&D and marketing roles.

## HIGH-TECH PIONEERS

Early federal research played a key part in electronic and metallurgical development—not unlikely, since more than half of Oregon's real estate is in federal hands. Studies at the University of Oregon medical school in Portland also were productive.

The laboratory seeds sprouted in a distinctive population of resourceful native scientists and radio enthusiasts, often ham operators, hungry to make a living at their developing art without migrating to more industrialized states. Oregon was limited by dependence on rural resources, and had only a low "carrying capacity" for service industries, yet her riveting beauty and hardy style of life were hard to leave.

Benignly neglected by the forest-farm-fishery establishment, the high-tech pioneers concentrated on specialty markets and reached international stature with precision products that often were unknown at home.

Radio Specialties Corp. was organized in the 1930's by a man who quit high school after he'd taken all the electrical courses. The first contract was for two-way radios for the U.S. Forest Service, whose laboratory in Portland had been designing them since the early '20's. The company still makes a frequency-modulation deviation meter that is a standby around the world according to Gale Sells, the founder, who is now retired.

Tektronix was conceived in a Portland appliance-repair shop by a Reed College-educated physicist with a visionary partner who had passed up college to get more quickly into radio, which he saw remaking civilization. They put the first synchronized oscilloscope on the market in 1946, and launched a new generation in waveform analysis.

Howard Vollum, the physicist, had worked on oscilloscopes with a University of Oregon medical experimenter, and then for the Army Signal Corps.

Electro Scientific Industries began in Portland during World War II making impedance bridges on defense contract, and was taken into the commercial world in 1950 by Douglas Strain, a Caltech engineer who preferred Portland to Los Angeles for his family. He first saw Portland while assigned to the Forest Service radio labs for the Office of Science Research and Development, designing a warning network to deal with submarine-launched Japanese balloon-borne firebombs.

Exact Electronics, an early Tektronix spinoff, is believed to have been the first maker of the versatile signal source called a function generator. Recently merged into Dynatech Corp., it specializes in analog signal sources for laboratory use. Dynatech is moving it from coastal Tillamook to Carson City, Nev., where it has other plants.

The roots of Oregon's exotic-metals primacy don't go that deep.

The U.S. Bureau of Mines built a regional laboratory in a vacant college in agricultural Albany in 1943 that quickly grew to strategic importance.

A pilot plant for the new Kroll zirconium-manufacturing process was built there in 1950, and produced zirconium tubing for the submarine test reactor at Idaho Falls, Idaho, and then for the pioneer nuclear submarine, the U.S.S. Nautilus. Commercial production was spun off by the Bureau, and the first contract was won by the Wah Chang Corp., an international metals trader and processor.

Wah Chang Albany is now a Teledyne Corp. subsidiary, and believes it is the western world's major producer of zirconium, titanium, columbium and hafnium.

Other specialty metal processors opened at Albany later. The laboratory was closed in a 1971 budget reduction, and some of the companies failed to survive that decade, but Oregon Metallurgical Corp. produces large amounts of titanium, titanium alloy and zirconium castings. Another titanium company, recently formed, is not in production.

## SHAPED BY NATURE

Geography has ruled Oregon's economy since before statehood in 1859.

Tenth largest state in size, it is the 29th in population. Half of the workforce and 40 percent of the 2.6 million Oregonians live in the three Portland metropolitan counties. (The metropolis also spills across the Columbia River into Vancouver, Wash.) Most of the remaining residents either share the north-south Willamette Valley with Portland or live in two other valleys, the Rogue and Umpqua, directly to the south.

As a result, the out-of-doors still is large in Oregon life. While evidence of

# Oregon

mankind is seldom lacking, whether farms, roads or logging traces, solitary landscapes are frequent and wildlife is close to hand.

The western one-third of the state is wetted generously by incoming Pacific storms compressed against the Cascades. Thickly forested, mainly with Douglas firs, that region has formed much of the standard image of the state.

East of the Cascades, Nature's domination is clear. Craggy mountains, jagged canyons, lonely basin-and-range high desert are set off in some areas by vast wheat ranches, or in others by the green circles of irrigated potato or alfalfa fields.

Below the surface, or sometimes armoring the surface, are the Columbia River Basalts, volcanic flows that covered hundreds of thousands of miles up to a mile deep in Miocene times, and according to current theory may be screening large oil and gas deposits.

What does this do for high technology? Not a lot, at first glance. A country made for naturalists and hikers, it lacks the scholarship and professional elbow-rubbing that attract first-rate physical scientists and engineers, say some observers.

"Where is your high-tech bar in Salem, where the computer people all go for a drink after work?" asks former Stanford Research Institute staffer John H. Wensley, founder of August Systems, a microcomputer manufacturing firm 60 miles from Portland at the state capital.

"An M.I.T. person would have culture shock here," says Ken Patton, general manager of Hewlett-Packard's plant in McMinnville, an hour from Portland. His 170 employees make medical resuscitation systems and special high-speed industrial X-ray units used for freeze photography of bullets speeding through gun barrels, shaped-charge explosions and similar short-lived, violent phenomena.

Patton, a Purdue mechanical engineer who was transferred to Oregon from Massachusetts, says he now would be reluctant to take his family away from the state.

Typical of Hewlett-Packard, he has his own research and development and marketing operations, with 40 or more engineers and some physicists.

The plant and process were acquired by H-P from a company headed by its inventor-developer, Dr. Walter P. Dyke, a physics professor from nearby Linfield College.

## CULTURE SHOCK

Patton's qualms about culture shock are shared by many in the business, and in academia. They say quality technical education and research opportunities are necessary if outstanding minds are to be attracted to the state so that its manufacturers can compete in the world marketplace.

"Most graduates will tend to take jobs near their campuses," is a theme heard frequently in discussions of the problem. References to the M.I.T.-Harvard axis, and Prof. Frederick Terman's nucleating role in the Silicon Valley phenomenon are also common.

One national, two state and at least four private educational institutions, all near Portland, are working at separate improvement programs and also talking about coordinated curricula for a broad appeal in electronics, computer and biomedical sciences.

They include the Health Sciences University, Portland State University, Reed College, the University of Portland, Pacific College, the Oregon Regional Primate Research Center and an unusual applied-research institution, the Oregon Graduate Center.

At the Health-Sciences University, which includes the state's medical, dentistry and nursing colleges, a building project for an Institute for Advanced Biomedical Research is underway. A program to line up private funds for endowed clinic and science chairs, general endowments and research fellowships has been organized, with a blue-chip board of overseers headed by Robert B. Wilson, chairman of the board of Weyerhaeuser Co., the timber giant.

Dr. Leonard Laster, president of the university, says the research institute will materially improve a school that has produced a long line of research benefits despite chronic underfinancing and lack of space. He points to world stature for achievements in heart surgery, unclogging arteries, stroke research, arrythmia, sleep disorders, kidney transplants, cornea replacement and many other fields.

At the Oregon Graduate Center, Pres. F. Paul Carlson is carving out an educational institution with an applied-research charter. OGC was organized in 1969 by industrial and civic forces who felt the need for research facilities and leaders close to the users of regional technology.

The center offers graduate degrees in applied physics and electronic science, environmental science, chemistry and biochemical science, materials science, computer science and engineering.

Carlson is basing his school's character on close working ties with industrial research leaders. Intel, Tektronix, Electro Scientific Industries, Weyerhaeuser Co., Southern Pacific, Hughes Research Laboratories, Honeywell Laboratories, Battelle Memorial Institute, Omark Industries and Bethlehem Steel Corp. are among the participants.

His site is near several high-technology office campuses including those of Intel and Tektronix. It is back-to-back with the Oregon Regional Primate Research Center, one of seven national biomedical research institutions supported by National Institutes of Health grants, working with resident monkey populations. Carlson aims at a close, synergistic linkage where workers and academicians share scholarship and enthusiasms, he says.

Portland State University Science Dean William Paudler, a chemist, says he came to the school last year aiming to help turn it into a major urban institution, at the same time coping with an explosive increase in science and math enrollment.

A passionate believer that the rural land-grant college is obsolete, Paudler is pressing for funds to improve what he says is already an impressive research program at PSU, with distinguished work in fish biochemistry, coal chemistry and liquefaction, blood pressure abnormalities, chemistry and earth sciences. He is expanding the mathematics and computer sciences offerings.

Oregon's chancellor of higher education, Dr. Roy E. Lieuallen, agrees that more programs are needed, and at a higher level, at Portland. He blames the state Legislature for withholding funds, and for setting salary levels that cannot compete with private offers to faculty members.

The state's University of Oregon and Oregon State University, both some distance south of Portland at Eugene and Corvallis, repectively, have sufficiently broad programs and simply need more money he said.

But the industry isn't waiting on legislative budgeters. Tektronix's Vollum and John D. Gray, the board chairman of Omark Industries who brought the chainsaw to maturity and revolutionized logging, have set out on an endowment campaign to energize the private campuses.

With sizable personal challenge grants to Reed College, Oregon Graduate Center and the Biomedical Research Institute, they have mounted a regional support drive that has made important differences.

"They helped us make a transition," says Pres. Paul Bragdon of Reed.

"Business needs the colleges more now than when we (he and Gray) started," says Vollum. "The technology wasn't as sophisticated or expensive then."

Far to the south, near the California border at rural Klamath Falls, another technological element in the state system of higher education has a peculiar claim to fame.

The Oregon Institute of Technology, with more than 1,000 students in its electronics and computer technology courses, educates them for two- and four-year science and engineering degrees, ready to step up to test benches.

John Lund, head of engineering technologies, says half the students are local, 90 percent are from somewhere in Oregon, and they don't want to leave the state. "They turn down good starting offers in the Midwest, Texas and California in the $16,000 to $24,000 range, even though we encourage them to get out and get some experience," Lund says.

The unusual aspect of their education, however, is not related to curriculum or job offers.

Their classrooms and labs are completely geothermally heated, and Lund, the associate dean, is an acknowledged expert in the arcane technology of the direct use of moderate-temperature goethermal energy.

The Klamath Basin has well-known hot water aquifers that draw their heat from magma cells 15,000 to 20,000 feet below ground and then are deflected closer to the surface.

As associate director of the Geo-Heat Center of OIT, the national clearinghouse for direct-use technology, Lund has traveled to China, New Zealand, Iceland and France as a consultant to national agencies, and is helping a local group develop a geothermally heated industrial park alongside the OIT campus.

They hope it will be taken by high-technology companies.

## ELECTRONICS BIG FOUR

Oregon's electronics landscape is dominated, but far from controlled, by four big names: Tektronix, Intel, Hewlett-Packard and Floating Point Systems.

All have made waves in science and technology, lifting them to new plateaus and at the same time making their founders wealthy.

Tektronix has been the state's principal seedbed, drawing talented specialists to the Portland region who from time to time would spin off their own ventures, and frequently furnishing their capital.

Dozens of companies, including Floating Point Systems, were bankrolled in the beginning by their founders' Tek profit-sharing accounts. Some have attracted direct investment by Tek executives.

Tektronix was opened in 1946 by Howard Vollum, his longtime partner

# Oregon

the late Jack Murdock, and three others. In five years, when the firm had grown to 100 employees, IBM buyers told Tek management they would soon have 2,000 and drew laughs, but facts overran the prediction. By 1972 there were 8,300 employees, and the number tripled in the last decade to 24,000.

It is the largest employer in two Oregon counties. It presently has three Oregon manufacturing campuses and three more blueprinted, plus one in Vancouver, Wash., and offshore plants in Japan, The Netherlands, the United Kingdom and the Isle of Guernsey. Net 1981 sales were 1.06 billion dollars.

The product range has been aggressively expanded into growth areas during the past decade. The catalog is headed by an impressive series of microprocessor and microcomputer support instruments, data communication testers, semiconductor testers and logic analyzers.

A range of computer display terminals and desktop computers with graphic displays and software is offered, along with display monitors for OEM and end-use customers. Television and cable instruments are listed, and radio-frequency spectrum analyzers.

At the bottom of the catalog is the instruments division, still producing oscilloscopes, calibration standards, signal processors and other laboratory essentials.

## SHADOW OF TEK

If Tektronix was important as a reservoir of trained people and of capital, it also cast a big shadow with its style of business—unostentatious, almost anonymous in the community—its main Beaverton campus, of rolling fields and stately oaks, didn't carry the company label for more than a decade.

Vollum says the omission was calculated. "I think the big advantage that the West had and still has to a degree was a lack of factory heritage," he said recently. Factory heritage contains built-in management-labor conflict, he said, and restrictive views of mutual responsibilities.

Tek's original workforce basically was of farm stock, people who were certain they could accomplish any task they were set to, said Vollum, whose father homesteaded near Portland. He and Murdock set a high value on that attitude.

They never seriously thought of setting up the company outside Oregon, according to Vollum, because there was no place they would rather live. Later, analyzing their position, they saw that customers don't really care where an instrument is made, but simply whether it works.

"Freight wasn't a problem, and you introduce a product at a national show, where the user can tell fairly quickly if it does the things he wants it to."

## INTEL ARRIVES . . .

A newly-arrived Textronix neighbor named Intel has been reshaping the computer industry with microprocessors produced in nearby Hillsboro.

In the mid-1970's the Santa Clara memory giant saw its growth and quality threatened by Silicon Valley crowding and workforce conditions, and decided to spread out to Oregon and Arizona.

The company now occupies three "campuses" in Oregon's Washington County, within minutes of Tektronix, Electro Scientific, Floating Point and a mass of other high-tech groups.

Complete with their own research and development staffs, they are carrying a major part of Intel's race against all rivals, domestic or foreign.

The Oregon plants produce the entire Intel line of singleboard computers, including the workhorse 8086 and 8088 microcomputers, and are churning out a stream of new products. "The new System 86/330, Intel's first fully integrated microsystem, was designed and developed here," said Keith Thompson, director of systems operations for the company.

He also pointed to the iAPX 432 family of 32-bit micromainframe processors, a set of microprocessor chips with the computing ability of a midsized mainframe computer. It is expected to have applications in telecommunications, online office systems, computer-aided design and simulation units, multiuser business systems, and factory automation and control.

Other computer products include the iSBC 88/40 process-control computer, and a line of In Circuit Emulators (ICE) for use by microprocessor designers.

Intel's memory-chip operations are also based in Washington County, churning out about 100 products with sales of "hundreds of millions of dollars a year," according to general manager Scott Gibson. Seven or eight new products are due out this year, including two 64K Random Access Memories. One is a "second generation 64K RAM" with eight separate 8,000-bit memories. "We're also working on our 256K RAM, just like everyone else," he added. "We expect them to consume all the land and facilities we now have," he said.

## SPINOFF COUNTRY

Just a few minutes drive from either Intel or Tektronix in Washington County, Electro Scientific Instruments (ESI) occupies a science park in the community of Cedar Mill, with 600 employees and a 30 percent annual growth rate.

Founder Douglas Strain, who invented a vacuum-tube electrometer while in high school in the 1930's, has brought the company across a mind-boggling span.

The first product was an impedance bridge, a variation of the classical 1833 Wheatstone circuit.

Present-day bestseller is the Microlase wafer processor, a computer-controlled laser trimmer that tunes integrated circuits to design specification by etching their microscopic components. The Microlase laser also can disconnect faulty microcomponents and "rewire" with built-in spares. Circuit modifications are also possible with the Microlase. The work is servopositioned with a tolerance of 2.5 microns.

Down the street in the same science park is the inevitable spinoff, P/M Industries Inc., with the inevitable job shop. There, longtime ESI engineer Paul F. Parks, Sr. makes his own line of less complex laser machine products, offers custom laser-trimming services for makers of microelectronic circuits, and also offers laser-machined ceramic substrates for short-run projects.

## FLOATING POINT SYSTEMS ET AL

Perhaps more than any other industry in history, electronics and computers are a spinoff technology. Scientists, engineers and perceptive workers at several levels see an undeveloped opportunity in their employers' business and they split away to start their own company—often after having offered the idea to the employer and gotten his blessing.

When former Tektronix sales engineer Norm Winningstad needed to expand outside membership on the board of his Floating Point Systems (FPS), he persuaded Tektronix President Earl Wantland to take one of the chairs.

FPS makes array processors—auxiliary computers that increase the number-handling speed of standard computers and minicomputers so they can take on extremely complicated calculations.

Typical users are oilfield geophysical surveyors, meterologists, petroleum recovery engineers, astrophysicists and the operators of CAT-scanners, the X-ray tomography image reconstruction process that creates three-dimensional images of human internal organs.

IBM and Digital Equipment Corp. both market Floating Point equipment, and FPS also has its own international marketing system.

Headquarters is in Beaverton, across the road from the main Tektronix campus.

Another Tektronix spinoff in Beaverton—there are many—is a cable-technology firm called CableBus Systems Inc.

Founded by former Tektronix engineer Cliff Schronk and partly financed by Tek founder Vollum, the company makes digital communications devices that operate on narrow radio-frequency bands over community cable networks.

Their first products are aimed at the residential-security market, but they are also supplying computerized hospital tv systems that can distribute special medical educational programs to individual patients, preempting the entertainment receivers. CableBus also is working on digital voice cable systems for municipal fire departments, and is discussing continuous telemetering of household electricity so power companies can reduce peak loads by giving customers preferential rates for shedding load at critical times.

It is also discussing an arrangement with the city of Portland for smoke detectors in every home, connected to a central fire department dispatcher.

## INFORMATION ON DISPLAY

The information-display division of Tektronix is spawning impressive descendants of its own, capable of attracting the venture capital they need.

Just now coming to market, they may be a new generation.

In Hillsboro, close to Intel, Dr. Gene Chao, former director of applied research at Tektronix, organized a color graphics company called Metheus Corp., that just began delivery of the first Omega 400 color graphics systems.

Chao recruited his key people from Intel and Tekronix both, and got the system to market just seven months after incorporation.

His series of high-resolution advanced computer graphics systems is aimed mainly at original equipment manufacturers and sophisticated end users, with emphasis on price.

The cost will be controlled, Chao says, through "aggressive use of the latest Large Scale and Very Large Scale Integration technology, using 64K RAMs, ECL logic and high-speed bipolar microprocessors."

High resolution, fast displays will have vector drawing rates of one million pixels per second, and sell for less than $15,000.

They are aimed at cartography, graphic arts including newspaper layout, computer-aided design and process control applications.

Mentor Graphics Inc. is a nine month old computer-aided engineering sys-

## HIGH TECHNOLOGY PARKS

Three new Oregon high technology real estate developments are underway, two of them education-linked and one reminiscent of Stanford Research Institute.

The Oregon Graduate Center, near Portland, has assembled a 175-acre project including part of its campus for phased development by research-dependent companies.

The research and real estate plan is part of the strategy of OGC Pres. F. Paul Carlson for an intensely creative campus with his faculty and graduate students, hobnobbing synergistically across the lawn with industry's gurus, or convening symposia with the scientists of nearby Tektronix, Intel or Electro Scientific Industries.

Carlson's design shows two million square feet of clustered but separate buildings on an open campus. He will encourage advanced energy conservation with solar heating, cooling and lighting, waste-heat recovery, geosource heat pumps and computer-controlled energy management.

And a neighbor just north of the Graduate Center has unveiled a 200-acre high tech development, Rock Creek Industrial Park. The venture, headed by Standard Industrial Insurance Company, will have 1-to-100-acre sites and be operated as a planned unit development.

Standard Insurance previously developed Tanasbourne, a nearby "newtown" with associated commercial and light industry zoning.

Three hundred miles southward, just above the California border at rural Klamath Falls, the 125-acre College Industrial Park development has been laid out alongside the campus of Oregon Institute of Technology, overlooking Upper Klamath Lake.

The project will be geothermally heated, as is the OIT plant. The heating venture is managed by Dr. John Lund of OIT's Geo-Heat Utilization Center, a national clearing house for non-steam geothermal energy data.

OIT was sited in the first place to take advantage of a geothermal aquifer that traverses the northern neighborhoods of Klamath Falls, which is in a mountain basin along the eastern flank of the Cascade Range, south of Crater Lake.

Developers expect the location on major highway and rail routes, with commercial air service, to be attractive to high-technology developers drawn by the pool of skilled OIT technician and engineer graduates. Klamath Falls presently has some small computer and electronic companies, but no major ones.

# Oregon

tem-maker with 20 employees; three hold M.B.A. degrees, another is an Arthur D. Little alumnus, and still another was the head of the Tek graphic-display division.

Their simulators will depict a structure without the need for building a prototype—"an absolute necessity with the coming shortage of designers," says Pres. Tom Bruggere. First shipments of the independent desktop workstations are scheduled for fall, even though no product announcement will be issued until June.

"We have every expectation of becoming significant," said Bruggere. "Smallness is just a development state.

"It is possible to create a product in Oregon that will attract high-quality investors," he said. "We've done it.

"We didn't leave Tektronix with a product idea, but with a marketing concept," he said. "Our customers will buy a system—hardware, software, peripherals, consultation and documentation."

Spunoff companies can generate their own spinoffs, in turn, and so it is with Floating Point Systems.

In the same industrial park as Mentor Graphics, a shiny new "spec" warehouse development more than half-filled with computer firms, are two young companies specializing in accessories and services for Floating Point equipment.

APTEC Computer Systems is a year-old creation with a contract from Control Data Corp. to making floating hardware for the Cyber System 1700 minicomputer. They have developed an array-processor timer and a FORTRAN complier that simplifies writing array-processor programs. They also have acquired rights to the original FPS product, the floating-point hardware to increase a computer's computational speed.

Former FPS production-line manager Woodrow Wittmayer is president.

## GYPSY TECHNOLOGY

The almost gypsy-like portability of small high-tech operation sometimes gives it curious, improbable settings in Oregon, whose typical town has less than 25,000 residents. (Only three Oregon cities have more than 50,000 residents; Portland with 363,000, Eugene with 87,500 and Salem with 73,000).

Some examples:

In Southern Oregon near the community of Kerby, in a 12-foot-by-24-foot plywood house on the slopes of Eight Dollar Mountain, with four microcomputers served by 1,350 feet of Romex electrical conduite from the nearest power line, lives the fabled personal-computer programmer Paul Lutus.

Lutus brought in the commercial power after he was unable to develop enough hydrogeneration on his land. He also uses it for a satellite television receiver he seldom watches. He has a collection of 1,000 books, and a Cessna 172 airplane which he bought for cash one day. Lutus says he rode his bicycle to the airport, handed the dealer a personal check for $21,000, and cherished the reaction when a phone call established its validity. Among Lutus' royalty-earning works is AppleWriter, the popular word-processing program for Apple microcomputers.

On the coast west of Kerby, at the fishing harbor of Brookings, Martin Palmer operates TVRO Systems Inc., a computerized satellite antenna system supplier that has never made a sale in Oregon, but placed fifteen systems in Germany last year and does most of its business with East Coast luxury condominiums.

In Portland, in a handsomely-restored commercial building in the shadow of the city's "suicide bridge," Oregon Software Inc. produces a Pascal II language compiler for DEC PDP computers, and is developing programs for the new Motorola 68,000 microprocessor. The firm is a "programmers' programmer" with worldwide trade. DEC and Honeywell have been recent big customers.

And in Klamath Falls, in brick-walled industrial space built when sawmills were the dominant technology, is physicist Art Merkl's Inventron Industries Inc.

Merkl pulled the 23-employee shop out of Los Angeles to find a more stable workforce. "People in LA auction themselves off every six months to the highest bidder." His children like the schools, he sails and waterskiis on Oregon's biggest lake, and manufactures ultrasonic level-control systems using an 8085 microprocessor with noncontact ultrasonic gauging in tanks and other enclosures.

In the rural town of Sweet Home, in the auriferous foothills of the Western Cascades, White Electronics Inc. makes a leading metal detector.

At Salem, the state capital, 60 miles from the nearest harbor, Morrow Electronics Inc. manufactures fathometers, Loran navigation computers and plotters, electronic fish-finders and handheld underwater sonar units for divers. Morrow has just introduced a lightweight Loran navigator for general aircraft that gives instant headings to any of 200 airports specified by the user.

Founder-President Ray Morrow Jr., educated as a game warden, left a Santa Clara Valley engineering job to return to Oregon and set up the company with $3,000. "I went out of here with the first 20 units in my truck and had to get cash for them, and did," says Morrow. He makes a point of using all-domestic components. Morrow learned the business from his father, an amateur radio operator who manufactured Morrow communications receivers in Salem in the 1940's and '50's.

Chairman of the Board at Morrow is former Citibank executive Ted Achilles, who left the Manhattan bank after setting up an Oregon outpost for it.

August Systems Inc., "The Can't Fail Computer Company," is another Salem firm in the bootstrap tradition.

Founder John H. Wensley quit the staff of Stanford Research Institute, and he and four partners sold their price-inflated Palo Alto homes for start-up financing.

Their "fault-tolerant" technique uses inexpensive Intel 8086 microprocessors in triplicate at the heart of process-control systems. All three monitor the process, and if they disagree, a "vote" is taken and the majority prevails.

A typical installation was in a large electrochemical plant where an August Systems unit selects a power source from among several different-priced, interruptible supplies. The power operates a weeks-long process that cannot survive interruption.

A Salem software house, Relational Systems International, recently published a universal personal-computer program by which inexperienced operators can quickly create special custom programs. They call the product PEARL (Producing Error-free Automatic Rapid Logic) and compare it to VisiCalc, a well-known free-form financial program.

Not all smalltown Oregon operators are small. Two of them are the biggest private employers in their counties.

Hewlett-Packard's personal computer and calculator divisions are on a 140-acre campus at Corvallis, home of Oregon State University. The workforce of more than 2,500 is the largest in Benton County. Fred Hansen, manager of integrated-circuit operations, said the company is pleased with the university and with transporation facilities, and enthusiastic about the workforce and community.

Their products include the recently introduced H-P personal computer, the HP87, and the earlier professional personal computers the HP83A and HP85A.

*(Also in Corvallis: Tiny Janel Laboratories, a TV antenna-coupler maker owned by engineer Robert Larkin. He quit Bell Laboratories to avoid a transfer from New Jersey to Chicago, and the family toured the country to choose a new hometown. Janel has nine employees.)*

In Grants Pass, Litton Industries

# U.S. Bank Would Like To Make A Profit Statement.

In any free market the profit motive is the catalyst that fuels growth and achievement.

And if you believe that profitability is a valid measurement of management skill, you'll understand the pride we take in being recognized as one of the most efficient and profitable banks in America.

To a large extent, our success is built on partnership with companies large and small. It comes from the kind of close working relationship we have with some of today's most exciting growth companies. As clients such as Tektronix, Floating Point Systems and Electro Scientific Industries grow and prosper, so do we.

We have grown, for example, to include 11 affiliates in 8 western states. So we can provide the broader vision and support large and emerging growth companies need: from investment services to international banking. And our commercial service teams are constantly developing new financial management options and strategies to meet changing economic realities.

Profitability built on partnership. Through good times and bad, that simple idea has produced positive results for us and our clients. Which is one statement we will be happy to repeat.

To find out how it can work for you, call Robert R. Mitchell, President, at (503) 225-6111. He will be happy to help.

**U.S. Bank**

The U. S. National Bank of Oregon
A member of the F. D. I. C. and the U. S. Bancorp family.

**Managing Today's Smart Money.**

# Oregon

Guidance and Control Systems is Josephine County's largest employer, with 425 people. They build military inertial-navigation subassemblies. Human Resources Manager Robert F. Morrison says the company, low-profile by policy, adopted an active civic role because the community expected it. "We find it enjoyable," says Morrison, who is on the community college and Cancer Society boards. Worker turnover is below three percent a month, and when upper-level jobs are posted in other parts of the Litton organization it creates "mass pandemonium—everybody wants to come to Oregon."

Preceding Litton Industries in Grants Pass was the Oregon Technical Products Division of the Bell Aerospace Division of Tektron Inc. They make military subassemblies with 240 employees, almost all locally hired. The community college helps with training, and "there's a better work ethic," says manager William L. Renton.

*(Also in Grants Pass: Met One Inc., 25 employees, making particle counters and meterological instruments. Owner Louis Petralli, feeling crowded in Sunnyvale, Calif., made the decision after driving through the town once. "This is it," he recalls thinking.)*

Medford, an orchard and sawmill town south of Grants Pass, has two new high-tech operations recently opened by Silicon Valley veterans who came out of retirement for the assignments.

James A. Little, president of Simco/Ramic Inc., makes sawmill-control computer systems using harsh-environment technology developed for satellites and missiles by the parent corporation, Quantic Industries of San Carlos, Calif. The Littles live on their dude ranch near Northern California's Trinity Alps. He commutes 100 miles each way, and says he has the best of two worlds.

DeWayne L. Spickler, general manager of the Oregon Division of International Memories Inc., Cupertino, Calif., had retired to the hamlet of Rogue River and built a 3,800-square-foot house. Then he persuaded IMI to set up a Winchester disk-drive plant at the site of a former Veterans Administration hospital, to make use of the under-employed, high-quality labor pool. He has 89 workers and is building toward 250. "These people come in, sit down and get to work without being harrassed; they thoroughly understand you have to do a job right to be productively employed," he said.

## WINDSHIELD-WIPER KIDNEY

Medical technology hasn't gone much beyond the entrepreneurial stage in Oregon yet, despite significant products developed in local hospitals.

One exception is B-D Drake Willock, a subsidiary of Becton-Dickinson, in the Portland suburb of Milwaukie.

With 350 employees, they ship about 400 kidney dialysis machines to the world market monthly. Besides the basic hemodialysis process, the firm makes a peritoneal dialysis unit for children and frail patients which uses the patient's own abdominal lining as a filter.

The company grew around a machine invented by Oregon City engineer Charles Willock to help a neighbor, Dr. Richard Drake, who had a patient dying for lack of an artificial kidney machine. Willock's invention was the first automated dialysis machine and was later converted into an home unit whose price was reduced by using commonly available parts, including a truck windshield-wiper motor for the pump.

Cardiovascular devices are built in Portland by Instromedix and a spinoff company, Powers Medical System, among others. Instromedix products include external pacemaker monitors, and remote electrocardiogram and blood-pressure monitors. They use microprocessors to hold the data, then replay it through a telephone adapter. The firm has about a dozen specialists in its research department.

## AN ENVIRONMENT FOR TECHNOLOGY

What does it all portend?

Many observers see fulfillment of Computer-Age Prophecy in the Oregon trend.

If high-technology people can perform their work anywhere within limits of communication and transportation, numbers of them will wish to do it in Oregon, they believe.

Numbers are, already. The landscape is changing in suburban Oregon.

One morning in February on the edge of Beaverton, two mallard ducks cupped their wings and dropped into a steep landing glide, slaloming toward a lowland where mallard guidance systems predict sheltering creeks and marshes.

The landing was aborted when they saw a landscaped parking lot covering the lowland; an asphalt sea broken by atolls of new "spec" warehouses.

And 80 percent of the warehouses carried the logos of computer, electronic or biomedical companies.

The economists of the Bonneville Power Administration, the federal regional power-system operator and planner, expect high-tech manufacturing to be Oregon's biggest industry before the year 2000.

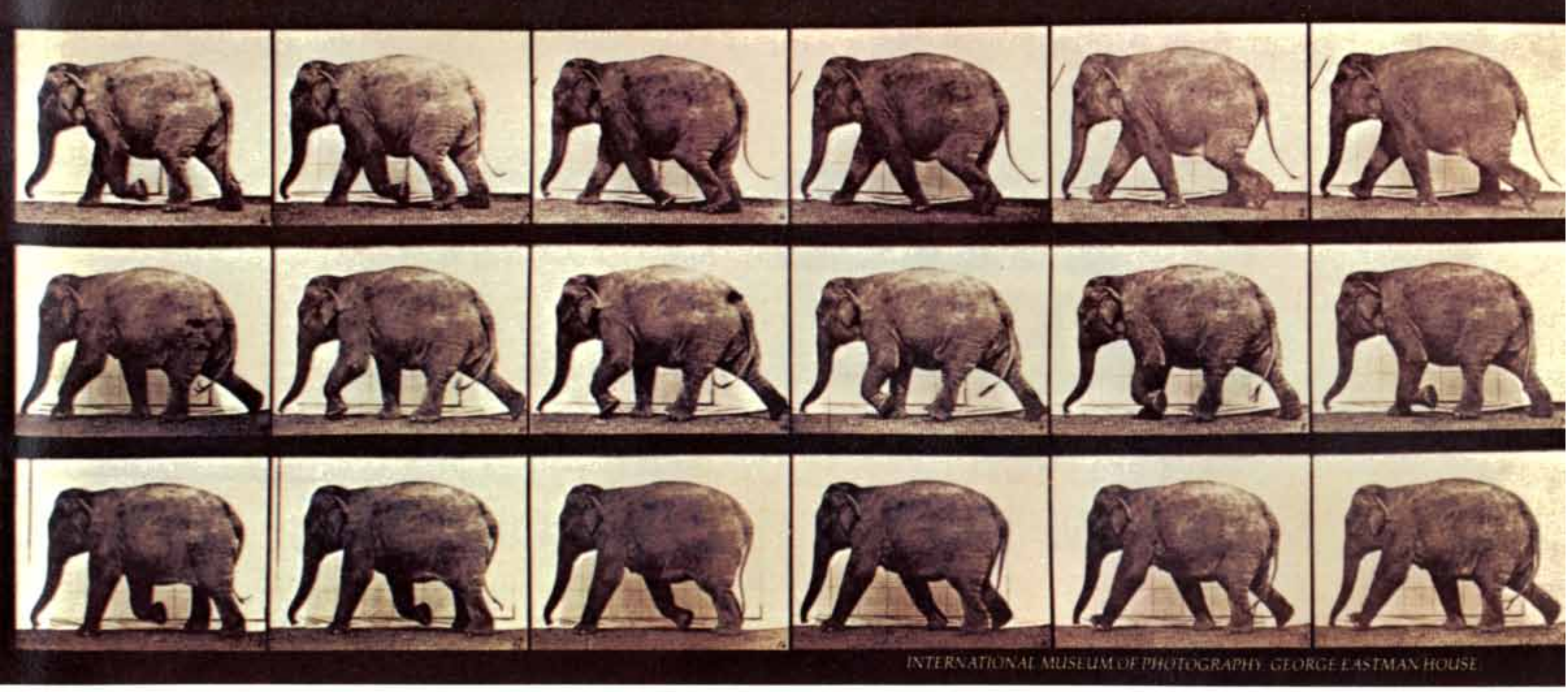
Marple's Newsletter, a respected regional business report from Seattle, noting the pressure for computerized productivity that is expected to maintain high technology as a high-growth industry, seconded the Bonneville Power Administration view.

"The region has only begun to scratch the surface," concluded Marple's.

*The Oregon Travel Information Office offers a free, 80-page color book with photos, maps and details on all of Oregon's many attractions and visitor facilities.*

*The Travel Office will also assist in providing specific information on the state's many meeting and convention facilities including some of the nation's more spectacular seacoast and mountain resorts in addition to convenient urban accommodations. Oregon tour information is also available.*

*Phone the Oregon Travel Office toll free 1-800-547-7842 for the Oregon Book plus any specifics mentioned above.*




# THE BIGGER YOU ARE, THE HARDER THEY FALL.

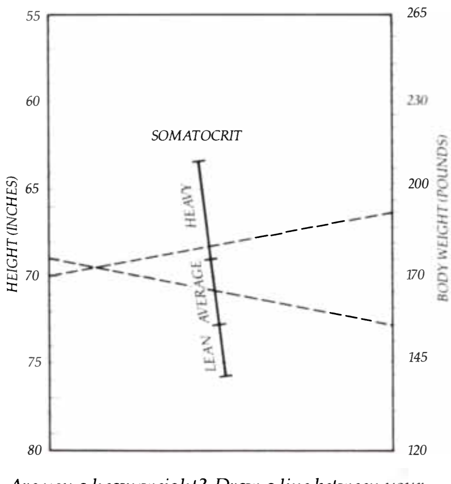
All runners punish their feet. But the heavyweights—they do it with authority.

If you're a runner who's 5′10″ with 30 extra pounds (even if it's all muscle), you are the lucky recipient of about 20 percent more vertical shock.

If that isn't frightening enough, our studies also show the forces side-to-side and fore-to-aft on your foot will also be greater.

While that may be a perfect set-up for injury, frankly, up until now, no one seemed to care. If you were built like an elephant, you just weren't expected to be much of a runner.

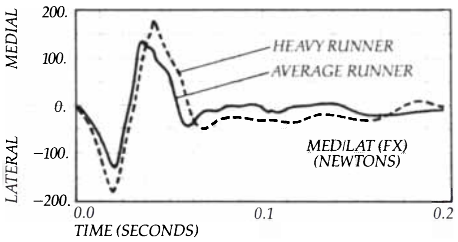
We found that slightly odd. For a couple of reasons.




*Are you a heavyweight? Draw a line between your height and weight and see where it intersects the somatocrit scale.*

First, an aroused pachyderm can cover 100 meters in about 9.1 seconds.

Second, and more important, our surveys now show that at a typical marathon, nearly eight percent of all entrants could be classified as heavyweight runners.




*Typical medial and lateral (side to side) forces are significantly greater for the heavy runner vs. the average runner, even when speed and running style are the same.*

No small problem.

And no shoe to solve it. Anywhere. While our LDV, Columbia, Titan and Intrepid may be fine for borderline cases, when a really big time athlete starts smacking a shoe around, it can easily bottom out.

The challenge was similar to designing a spring that would let the foot sink, but not all the way. Because a midsole that's totally compressed, if only for an instant, isn't doing you any favors.

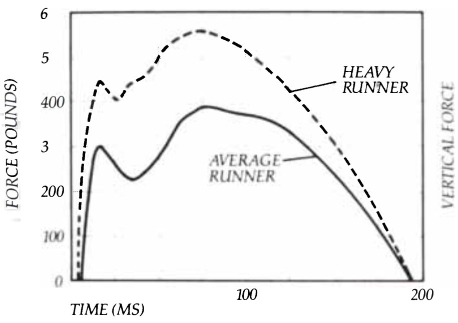
However, our new Nike Centurion certainly will. Thanks to a unique combination of firm wedge and soft midsole, it gives heavyweight runners the kind of comfort only the less developed used to enjoy.

To increase stability, we widened the Centurion's base of support and added leather reinforcements on the upper. Then upped the heel counter, to 60 weight, extended it, and locked it into place with a Stability Saddle.

Hopefully, the Centurion will not only save your feet, but your pocketbook as well. Durability tests indicate stout runners experience no more loss of cushion in the forefoot than normal. Even less in the heel. Ditto for outsole wear.

The last thing we want to do, however, is give athletes an excuse to beef up. Generally speaking, the less you weigh the faster you'll run. And to burn fat, run slow and run long.

But if you never get below 200 pounds and still crack a three hour marathon, we want to hear about it. Because all things considered, that's world-class time.




*Vertical ground reaction forces for a 200 lb. runner and a 150 lb. runner at the same pace. Forces under heel and forefoot are both proportionally larger for the heavy runner.*

And nothing excites us like athletes who reach their potential.

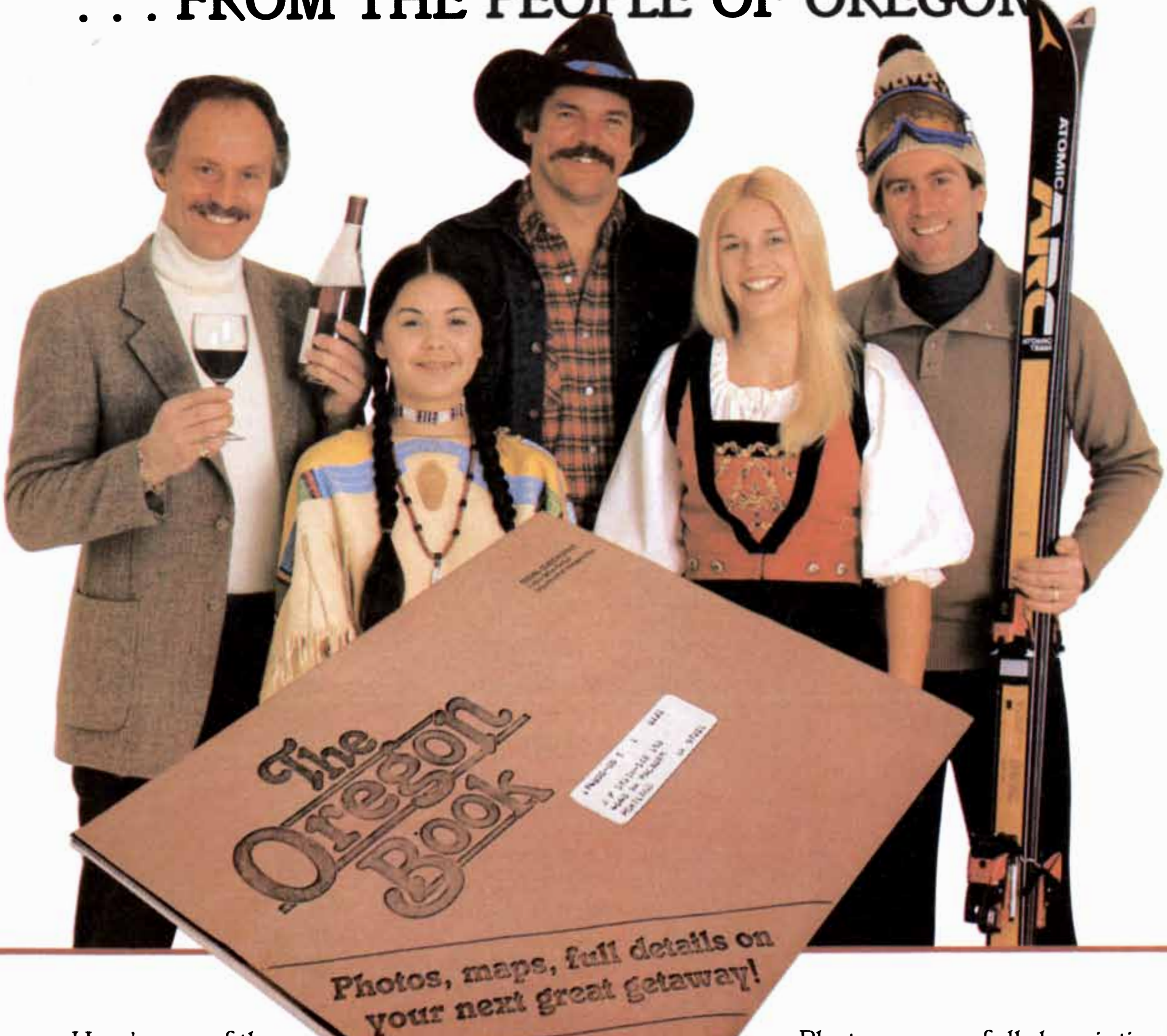
Even if they insist on throwing their weight around.

NIKE

Beaverton, Oregon

# TAKE OUR FREE ADVICE

## ON YOUR NEXT VACATION . . . FROM THE PEOPLE OF OREGON




Here's one of the world's great travel destinations, presented in full 64-page detail:

**Everything from wine-tasting to whitewater . . . seacoast frolics to guest ranching . . . plush resorts to rustic cabins . . . fishing rewards to rockhounding . . . year-round skiing to Shakespearean theater . . .** Oregon has it all, and it's all in the book!

Photos, maps, full descriptions to invite you among the friendly people of Oregon. Yours, free for the asking!

**PHONE TODAY for your FREE copy: 1-800-547-7842**

(outside Oregon) or write to Oregon Travel, Room SC482, Transportation Building, Salem, OR 97310.

# BOOKS

## *The Viking Atlantic, the history of fossils, cosmology, a handbook of physics, orchids*

by Philip Morrison

THE CONQUEST OF THE NORTH ATLANTIC, by G. J. Marcus. Oxford University Press ($25). The Atlantic stretches broad from Gibraltar to the Spanish Main. The long voyages over that passage under sail defined the European Age of Discovery, marked by an indispensable continuity. G. J. Marcus, a British maritime historian who pays as close attention to the experience of deep-water sailors as he does to port records, gives instead a convincing account of medieval pioneers across the same ocean but in those icy latitudes where the arc of the earth itself is only half as long and the passage is twice broken by the big islands of the subarctic.

This very readable story deserves to be called a saga; its judicious and well-documented arguments center on recent scholarship and archaeology on the Vikings, on their poets and their ships, their smiths and their axmen. It was not the long ship, a specialized fighting craft of low freeboard for amphibious expeditions, but its faster, broader merchantman counterpart, the *hafskip,* that carried these voyagers, under one square sail, *vestan um haf,* west from Norway across the sea. The first ventures beyond their coasts must date back to about A.D. 700; by a century later the Norse were settled in the Faeroes and the other sea islands north of Ireland and Scotland proper. Iceland became home to the Norse first along its northern coast, less surf-beaten than the southern one, the stream of immigrants reaching flood height in about 900. There were scores of ship arrivals each year, the 40-ton ships laden with families, livestock and timber for house building.

Just as we learned in school, it was Eirík Raudi, a formidable red-bearded chieftain with a keen sense of public relations, who discovered Greenland and settled there. The old books tell us that he knew of one Úlfsson who had caught distant sight of those shores during a long, involuntary storm-tossed voyage to the west. Banished from Iceland for three years in punishment for fatal quarrels, Eirík sailed angrily away not east to Norway but due west to a landfall fronting the ice cap. He wintered three times on Greenland, mainly around the southernmost cape and along the more hospitable western shores, their fjords so like those of Norway.

From Eirík's time, about 985, to the extinction of the Greenland settlements nearly six centuries later Greenland was a remote but never-forgotten colony of Norway. Its domestic economy centered on pastured livestock; timber was lacking, grain was very scarce ("the majority of the people . . . had neither tasted nor seen bread") and iron too was short. For these necessities the colonies traded in Arctic rarities, mainly walrus and narwhal ivory by the shipload, together with hides and once in a while a white falcon or a live polar bear for royal astonishment. The ruins of churches are still to be seen in Greenland; Eskimo lore recounts the sound of the great cathedral bell, of which fragments have been found. It was the icy cold worsening over the centuries, and perhaps a final coup by English pirates, that eroded and then ended the settlement, whose long succession of worthy bishops is the matter of centuries of explicit records.

The *Groenlendinga saga* tells of young Bjarni Herjólfsson, rash and inexperienced, whose fog-beset voyage from Iceland overshot Greenland and caught distant sight three times of lands with low forested hills before he made enough northerly way to reach the glacier-guarded Greenland harbor he sought. Leif the Lucky, Eirík's son, only followed Bjarni's lead, but he made landings, first on Markland and then on Vinland, the lands he named for wood and for grapes. A series of voyages followed; these places were real enough. The archaeologists know Viking remains of unquestioned authenticity on this continent only at L'Anse aux Meadows in Newfoundland. There is no evidence, however, that this hamlet was Vinland the Good; it may or may not be the place visited repeatedly, always by the Greenlanders, in the 11th century.

How did they navigate? The magnetic compass came into general if tricky use in these polar parts before 1300. For a couple of centuries the Norse had already practiced a regular deep-sea trade without it, if not without risks. One key was latitude measurement by the polestar and the sun, with the angle judged more by sight line and mast shadow than by any instrument. It was Leif himself in 999 who first of all captains sailed straight along the parallel direct from Greenland's tip to Hernar in Norway, about 1,100 nonstop nautical miles. There were finally written directions for such routes, the line plotted north of the Shetlands and south of the Faeroes. Dead reckoning was well developed early, aided by some version of a ship's log and by the sounding lead but not by charts. Viking distances are generally good estimates; their length unit was a day's sail, six nautical miles times 24, not a literal time at sea.

Dead reckoning is scarcely by itself enough; the Vikings had the essential astronomical insights. The sun at upper and lower culmination was their main resource in the busy summer season of the midnight sun, when no star near the pole was visible. Cautious, detailed lengthy attention to sea states, ground swell and the types and flight directions of birds—with the exploitation of distant looming during infrequent but remarkable mirages—all played a part. It should be noted that Viking seafarers were at home in Sicilian waters too; big sky shifts on northing and southing were by no means strange to them. The sunstone and other such suggested instruments are not mentioned by Marcus; his view of navigation sees the growth of craftsmanship, free of surprises. His master mariner, like the Gilbertese on the other side of the globe, "knoweth euery coste in his viage. . . . The depper see, the gladder he. . . . He knoweth euery rok and euery race/, The swolewys & the starrys, sonde & sholde." (A swolewys is a whirlpool, in this recension of a 15th-century English poem, *Knyghthode and Bataile.*)

The Irish were first to the westward, step by step over the centuries. Their eremites, for the sake of true isolation, settled themselves sparsely in the islands all the way to Iceland; they were there when the Vikings first came to those shores. A few monks (*papar,* the Norse called them) dwelt on Iceland's southern coast, but they could not long remain once the Vikings took over. Theirs was "the first notable geographical discovery of the Middle Ages," in the sixth and seventh centuries. But the voyages told of St. Brendan, the sea pilgrim of Clonfert, far beyond the nearer isles are probably not to be taken as historical. The photograph of the hermitage now on the forbidding rock of Skellig Michael, off the coast of county Kerry, is eloquent proof of the audacity of these recluses.

The book ends with an account of the entry of the men of Bristol into wester-


"A telephone system
that'll save Westinghouse
$66,000,000 in ten years?
GEE!"
(No, GTE!)
And what a system it is!
To start out, it will serve 15,000 telephones. At 20 different locations. (That's bigger than most of the country's independent telephone companies.)
The network handles basic telephone service and high-speed data transmission over the same lines. And it allows Westinghouse to use such "office-of-the-future" services as, for example, electronic mail and video conferencing.
And since they own the system, Westinghouse won't have to pay for a single phone call between their 20 locations. They won't have to pay for equipment rental. Nor worry about increased rates in the future.
All of which will add up to that phenomenal $66,000,000 ten-year savings. Or, according to Westinghouse, a 28% return on their investment.
Where else can you put your money today and do that well?
GTE

ing; when they finally engaged John Cabot himself, some of their bolder fishermen already knew the swarming cod grounds off his New Found Land, perhaps as long as a generation before Columbus. It was more the sea than the land that had lured their captains. For every considerable entry in the history of ideas there must be a prehistory; this admirable look backward in the end strengthens the common view. Ours is the first age of a single world, and rightly is it called post-Columbian. It seems a fair bet, although we shall not easily prove it, that some lost hunter in a kayak of skins long ago made the first transatlantic passage of all, blown against his will not westward but eastward from the edge of the ice.

LIFE HISTORY OF A FOSSIL: AN INTRODUCTION TO TAPHONOMY AND PALEOECOLOGY, by Pat Shipman. Harvard University Press ($19.95). To a paleontologist any animal is a fossil in the making. It was the Russian paleontologist J. A. Efremov who proposed in 1940 a new branch of paleontology whose object it should be to study the transition from living form to lithosphere that is the story of every fossil. Taphonomy is his word; it is from the Greek roots for the laws of burial. Always implicit in the study of the fossil past, it has become explicit, often salient, in vertebrate paleontology over the past decade or less, particularly in the tempting and vexing problems of the study of our ancient hominid forebears. We ask of fossils, what was it really like, that life past? The aim of taphonomy is to recover a clear signal from the noisy and eroded message coded in the details of any find.

This crisp survey, the first introductory text in the discipline, opens the subject for students and the general scientific reader with steady hand and light touch. One of the many witty but direct line drawings by David Bichell shows the loss of information stage by stage, from a lively little antelope through the kill and the trampling and the rains to the fault slip and the exposure by partial erosion in a distant epoch. The drawing is less a caricature than it is a summary; indeed, the chapters systematically follow a related sequence, although of course they emphasize genuine examples and the techniques now in use to puzzle out the blurred record in the rocks.

Death begins this story. Its causes vary, and they clearly modify the result. Starvation accompanies drought; many carcasses are found near waterways, to be preserved by rising water, often in a time of surfeit for predators. Only a few bones reach quiet burial for long preservation in the sediments; nearly all are lost, to predators, to scattering by wind and water, to rolling and slumping, to disintegration by weathering and chemical decay. These matters are all systematically examined. Sorting by water currents is semiquantitatively treated; it depends on known if complex properties of a bone, such as its density and effective diameter. (The treatment here is insightful from the user's standpoint, although the author and the editor alike are ill at ease with the hydrodynamic formulas.)

One example is drawn on repeatedly from several different points of view. It is the Fort Ternan hominid site in Kenya, first excavated with care by the Leakeys and their crews between 1959 and 1967. It was painstakingly reworked in the modern manner in the mid-1970's specifically to gather taphonomic data by a team including Dr. Shipman. She tells us of high standards; each bone bigger than a couple of inches was located in three dimensions to that same accuracy. Key maps are shown, and the conclusion appears firm. The bones were concentrated along a net of small channels quite visible in the rock sections. But orientation and dip data on more than 1,000 bones show a random pattern in plan together with a predominant horizontality.

It was not wind or water that distributed these bones, even though they do lie along small, water-worn channels. It appears that the sediments come from a basin where the old rains cut small gullies. The bones entered these crevices with the ancient mud, in a slumping flow that did not move them any significant distance. They were buried almost as they lay, in horizontal positions. In one Uganda site some long, slender bones are instead found nearly vertical. There is no support for the idea that high-speed flows of water swept the bones into such unstable positions. The suggestion is made that animals as they drank trampled the bones at random into the mud, first stepping on one end and thereby raising the other end high, then driving the entire bone vertically downward. The lesson is plain: it is easy to jump to unwarranted conclusions. Noisy data are hard to read.

Not only environment but also nature is marked in a bone. Species assemblies have long been puzzled out as a guide to origins. The trace-mineral content, weathering pattern and microscopic wear are clear invitations to newer study. At Fort Ternan there were no bones so weathered as to have fallen mostly apart in situ. To be sure, such fragments are easily missed. But the site was so well examined—a large volume was sieved for every fragment—that it suggests the assembly was fresh on burial; it had been weathered for a decade or so at most. Species and bone-frequency counts call for much care, even though they are rich in inference. Water currents, weathering, owls, porcupines, hyenas and hominids assemble distinct bone populations. This approach has yielded spectacular results, both false and true, from the bone remains found in the limestone caverns of South Africa, described lately in a review in these columns.

The most exciting recent taphonomic result is the discovery that the diet of an animal in the last months before its death is worn into the tooth surface, to be visible under the scanning electron microscope. An illustration (made by Alan Walker) shows six such surfaces. The orangutan's tooth looks nothing like that of the carnivorous cheetah. We are led to conclude that the robust australopithecine too was a gentle frugivore. Here is a line drawing of five reconstructed versions of that old ancestor of ours, all based on the same skull. They look very different, but not more so than some gifted character actor made up in startlingly different ways. We can foresee a long period of new texts lucidly deciphered from the Book of Sediments.

COSMOLOGY: THE SCIENCE OF THE UNIVERSE, by Edward R. Harrison. Cambridge University Press ($24.95). Unusual, discursive, nonmathematical, full of reflective comments and disturbing questions, packed with unexpected citations, this text is a gentle but far from innocent introduction to cosmology today. The modern science of large-scale models of the universe bears the weight of those ancient treasuries of epic and wisdom that all societies have amassed to sum up the total picture. Professor Harrison, himself an adept among contemporary traffickers in Robertson-Walker metrics and such, offers a comprehensive account of the modern view more or less on the terms of the timeless tales. Naturally his graphs and other illustrations are wiggling waves and diverging world lines rather than gods and demons tugging on the two ends of a serpent. His aim is nonetheless to make a volume intellectually continuous with the traditional past, at an easy level consonant with his view that "cosmology is the 'natural' science for students who are not majoring in a science." It is for such students (at the University of Massachusetts at Amherst) that he has wrought, not avoiding mathematics at all costs but doing his best to convey in the common speech the physical-geometrical arguments about first and last things, about bounds and beginnings.

Professor Harrison offers 20 chapters, each with a coda of questions and comment. The first chapter addresses the cosmologies of the past (rather too briefly), three others summarize the world we empirically find: stars and galaxies and the bare elements of the discussion on the origins of life. Four particularly fresh chapters discuss the issues of space and time. "The big bang did not occur somewhere in space. ... It occupied the whole of space.... If space is infinite, the big bang was also infinite in extent.... Cosmic edges do not exist." That home truth is conveyed without formulas but in depth, not by mere assertion.

It is this explicit facing up to the conspicuous paradoxes of the subject that marks the book. Half a dozen further chapters treat the more familiar basic ideas of general relativity, spatial expansion, black holes and the many possible overall models within current theory, classified as progressions of the scale factor among bang (condensed), whimper (diluted) and static states. Professor Harrison's careful account of the red shift and its direct interpretation in terms of the scale factor is the best popular transcription of the theory, although not all readers will much care for his view that its classical limit, the Doppler shift arising from motion, is not a useful interpretation. Two more topics that are valuable contributions are his treatments of the dark night sky and of the cosmic box, a little cell representative of the entire homogeneous universe, which we can examine without the need for distant travel. "I could be bounded in a nutshell and count myself king of infinite space," said the Prince of Denmark.

There are plenty of dreams, too, such as the first minute of particle origins, white holes (sensibly cut down to size) and the role of the possibly changing great numbers of the physicist. The interesting topic of the horizons of our knowledge is discussed in some detail, as it refers both to events beyond our ken and to enduring structures we cannot observe. Not quite as up to date as the newest proposals for multiple vacuum states in the early quarky world, the text is fully at home with ideas of the transience of matter and considers the model of infinite dilution and incredibly slow decay: "Ultimately, when the cosmic radiation has cooled,... the universe will consist only of photons, neutrinos and gravitational waves, and all else will have vanished. There will exist no matter and antimatter, and the universe will at last be baryon symmetric and in a state of maximum entropy—for eternity. *Sic transit....*" (One ought not to rely on those photons and neutrinos; they have no specific sources left, whereas the colliding gravitons remain sources of themselves.)

Meticulous and acute, Professor Harrison is nonetheless a bit overcome with metaphysical anguish. "The whimper universe is surely the most appalling state of damnation ever conceived by the human mind," he says, as though infinite time were a novel relativistic idea. Many readers will not quite share his metaphysical melancholy about self-reference; how can our cosmology include the mind of the cosmologist, with an idea of the cosmos including the mind within itself? That infinite regress is not so fearful: the map that is knowledge is both social and necessarily incomplete. A reader is quite apt to be content with a cosmology containing, if not his own world view, that of some other mind. Even the retinal image has its blind spot.

Bernard Shaw and Austin Dobson and the length of Brahma's day are all found here, citations witty and profound, among more technical universes from Thomas Aquinas through Willem de Sitter to Tycho Brahe and Thomas Wright. A beginning serious interest in cosmology can find no better satisfaction than in this helpful overview, taken with taste. Crowding hard on the Hausdorf axioms, the question of the heretic Pelagius appears here, directed against the concept of original sin. "If it is necessary, then it is not a sin; if it is optional, then it can be avoided." Let theologians and logicians contend; a cosmology with predictive strength is not yet at hand.

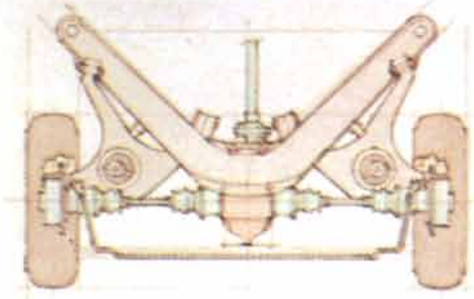
PHYSICS VADE MECUM, Herbert L. Anderson, editor-in-chief. American Institute of Physics, 335 East 45 Street, New York, N.Y. 10017 ($25). Some 60,000 American physicists and students of physics dwell peacefully under the umbrella of the American Institute of Physics, which has for just 50 years served to link and to manage the organizations of that diverse and powerful science, nine national societies with a dozen journals, chiefly archival repositories for the papers that are the symbol and substance of published scientific research. The profession is strongly differentiated: there are specialists in acoustics and in optics, in low-temperature physics and in medical physics, and many more, in addition to the familiar mainstream of the big laboratories of nuclear and particle physics, of solid-state wonders and the physics of the atom.

The AIP serves astronomers as well; in our country many single university departments include both physics and its ancient progenitor. Physics has always been fissiparous too. The dynamics of fluids, for example, preserves the names of physicists such as Helmholtz and Kelvin, Rayleigh and Landau, even though its development has long engaged applied mathematicians, aeronautical and hydraulic engineers, meteorologists, oceanographers and physiologists. Their muster is now much larger than that of the parent specialty, but a powerful minority within the smallish circle of physicists remains caught up in the intricacies of fluid motion; *The Physical Review* has a companion journal, *The Physics of Fluids.*

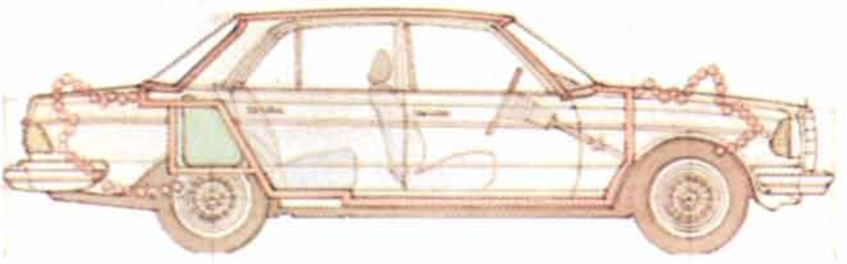
An anniversary gift to the substantive unity of the science, this attractive compendium was prepared to be of use "to the wide spectrum of physicists" affiliated with the AIP. Portable rather than bulky, it presents a dozen pages of the definitions, data, formulas and references most useful for any physicist in each of some 20 fields "physicists are supposed to know something about." A general section, four times longer, presents the units, the mathematical and physical formulas that might be held as a common foundation, plus a choice set of practical laboratory matters and the needed properties of everyday laboratory materials.

There are predecessors for this compendium, of course. They are usually either more specialized (the entire volume compares in size and form with the wonderful *Astrophysical Quantities* of C. W. Allen of London that is now at the elbow of most astrophysicists) or much more comprehensive and perforce much larger. This scheme, enforcing an austere choice on its specialist authors, is fresh. No holds are barred, of course; although the chapters include the fundamentals in field after field, they are in no way introductory. Let him who is without understanding not enter here, but he who comes prepared will find paths to sound conclusions.

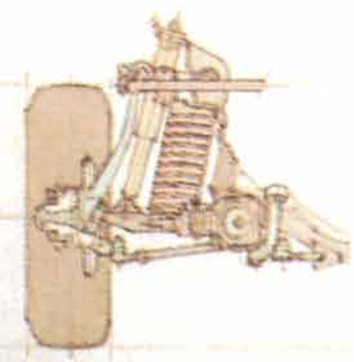
A dozen pages is not much for today's physics. Take the crystallography chapter. It opens with a little history and a citation map to show that the field has most transactions with chemical journals. There follow a table of the point groups, some notation and two tight pages of lattice constants for substances from diamond to Rochelle salt, quartz, penicillin and myoglobin. Then there are five pages on the geometry and physics of diffraction and diffractometers, with plenty of fine points, and a table of mass attenuation values. Next we read of atomic scattering factors, for X rays and for neutrons, with a table of dispersion corrections. Thermal motion is discussed, with references for its treatment



Overhead view of diagonal pivot rear axle, a key element of the 300D's elaborate fully independent suspension system. To Mercedes-Benz, sports car handling and sedan handling should be the same.

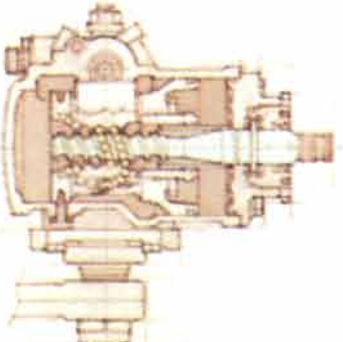


There are 120 safety features incorporated in every 300D–and the car's body itself is one. Front and rear sections are carefully devised to progressively yield on heavy impact (dotted lines), managing kinetic energy before it can be transmitted with full force to the passenger area within.

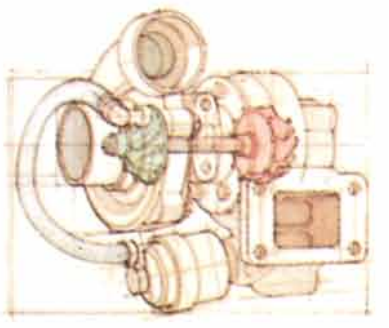


Zero-offset geometry of front suspension mobilizes esoteric physics to guide the 300D along a true forward path.

Floor-mounted shift lever connects to a rugged four-speed automatic gearbox.



Twenty-four tiny ball bearings form the heart of the 300D's exquisitely smooth and precise recirculating-ball steering system.



The 300D's rank as America's most potent diesel automobile owes much to this 17-lb. piece of engineering brilliance: a turbocharger.

Instrument panel of the 300D is not decked out with gadgets but laid out with ergonomic exactitude. Look of rich wood is achieved by using rich wood.

300D's vivid performance is excelled by no other diesel automobile sold in America.

as a disturbance of the simple Bragg intensities, and the chapter ends with brief accounts of the clever ways to infer phases from the intensities that are all experiment yields.

Computation is the key, and the digital computer (the most popular program is MULTAN) has now of course turned crystallography into a routine analytical tool for molecules of a few dozen atoms, not to mention a unique source of structural detail for the giant molecules of life. One small result of the unity sought in this book is the fact that the atomic-substitution techniques that earlier helped to repair the ambiguities of X-ray-diffraction studies have been extended by infrared spectroscopists to locate the vibrating atoms in the molecules. Heavier or lighter substituents shift the infrared resonance so that the molecule can be more accurately mapped.

There is no way to summarize a score of specialties. Some points impress a reader: Cryogenics now deals with superconductors in wild variety, each with its special properties and its claim to technical advantages. Millikelvin temperatures are routine; the temperature scale has been extended down there by use of the Fermi substance helium 3. Physicists now worry about large-scale energy economics; their tables include the end uses of U.S. energy, the thermal resistance of vinyl floor tile and world energy consumption over time by type of source. Stars and quarks and flow through rough pipes and gamma-ray calibration sources (with photon energies given to seven decimal places) are all here, along with ferroelectric and ferromagnetic transition temperatures, the properties of plasmas from those of the interstellar gas to those of tiny laser-made puffs, and a picture of the globular tertiary structure of myoglobin.

Few will be adept in all these arts or will be called on to apply them. Many will want the chance to consult one or another book on library shelves or closer by. The cross sections cut here are wide and cunningly selected, if necessarily thin. The editors and the contributors deserve praise, to be shared with the designers and all others who have participated in the making of the book; it is easy to use. A psychoanalyst might find it tempting to explain why shock waves, explosives, fission products and transient thermonuclear reactions are entirely omitted; not even the energy content of TNT makes the table of fuel values, although it surely is the basis of one of the best-known energy units outside of everyday affairs. The somewhat accidental history of the societies of physicists does limit the technical treatments; ubiquitous electronics hardly enters either. There are plenty of other handbooks; this authoritative and advanced one has a pleasing tension between oneness and variety.

THE ORCHIDS: NATURAL HISTORY AND CLASSIFICATION, by Robert L. Dressler. Harvard University Press ($27.50). On a sunny morning in the cloud forests of Panama the euglossine bees, nonsocial tropical kin to northern bumblebees, are abundant and diverse. Any skillful cast of the net will trap a bee, large or small; there may well be 50 species of the bees in a small area. This bee chaser, however, is a botanist. His attention is in fact drawn by the orchids in flower perched high on the limbs of the big trees. The winged bees can get there; they seek resinous terpenoid droplets from the flower surfaces, which they store in their hollow hind legs to modify somehow for use as pheromones. The bees do not leave the blossoms unscathed; each forager is fixed by one or two tiny golden banderillas, rather as the brave bull is prepared for the matador. These darts, which at millimeter scale are not at all sharp points but are merely glued to the surface of the insect, bear the packaged pollen of the orchid.

An illustration shows the field results of the author, a Smithsonian scientist. He identified half a dozen distinct placements on the bee's body, every location specific to a known form of orchid. It is the flower structure—its various tunnels, traps, obstacles, even trip wires—that manages that precision; one can be pretty sure that the pollen parcel is removed by just the right contact with the right stigma in the next blossom of the species the bee enters. Indeed, the parcels bend and twist for a minute or two as they dry, so that they do not reach the removal position until there has been ample time to enter a second flower.

Such precision pollination allows the sexual isolation that defines each species of orchid. The flush of sweet apple blossoms in the Temperate Zone spring exploits chance; as long as a bee enters one flower to pick up the dusty pollen it is pretty sure to shed some in another blossom nearby. The diverse, sparsely strewn, nutrient-poor orchids need more precise methods. Their numerous seeds are dust-fine, airborne, able to settle on a limb above the forest floor, there to find once in a while the right fungus symbiont to feed them in germination; they develop into saprophytic seedlings. The many-seed flower therefore needs many pollen grains, all sent to the right address. Sweet nectar might bring in any hungry insect; orchids are subtler. Specific attractive fragrances and precise flower forms allow long-distance specific pollination. By way of a single bee species many flower species can send distinct messages in different locations on the body of one small wanderer.

A color photograph shows a departing wasp bearing two neatly affixed bundles of pollen. Another shows the remarkable dark, glossy, well-formed fleshy growth that is to be the object of mistaken copulation by some wasp male, whose ardors will gain him only a pollen package or two. Sugar is easy for rich flowers to offer, but love and its promise carry far more information at low energy cost. Hence the speciation of orchids rests not on a long accumulation of genetic changes but in many instances on a single protopheromone or on some other trait quite unrelated to the actual fusion and development of pollen and ovary. Crossing is thus fully possible between forms that are isolated in nature only because of their unwitting messengers. The orchidist in the hothouse takes full advantage of this breeding flexibility. Even in the wild there are all sorts of orchid hybrids; some big or bumbling pollinators have not yet been excluded from the evolving chain.

This reflective and expert book considers the manifold orchids from the evolutionary standpoint. Since before Darwin this group of flowering plants—so beautiful, so diverse—has held the botanical imagination. But it is the flower forms and their variety, the hybrid richness now open to the florist and the spice grower, that has motivated most studies. Rich volumes of color plates of orchids are numerous, but they tell little of the intricate story. This is the first book in some time to try to piece out the evolution of the group and to essay a new classification in a fully modern vein: chromosome counts, biochemistry. Dressler even provides a final chapter of hints for further study in the laboratory or the field, aimed at both the professional botanist and the amateur orchid grower. The volume, rather technical on the side of botanical structure, is nonetheless full of interest for the general reader. Its color photographs are small but striking, remarkable examples of the stratagems described.

The editors of this book have not served it well. Cross-references between the text and the photograph captions—entirely absent—would have much aided the reader to grasp what he is seeing: one flower pretending to be carrion, another sporting a set of rich pollen-bearing anthers, all false, a moth-aimed blossom, a bird-luring one. Long ago the ur-orchid was a kind of tubular lily; once its radial symmetry was broken because the insects always crawled in along one side, the intricate pattern began. Those flowers then lived on Cretaceous rocky cliffs; the rest arises from the elaborate details. Orchids now are mainly tropical epiphytes; there are some 8,000 species known in tropical America alone. Four dozen species may bloom on a single tree. It is clear that there is work ahead and to spare.

# Placido Domingo, the complete musician, discusses his favorite instrument.

Every half century or so, a leader emerges in his field of such substance and force that he stands out head and shoulders above the rest, and the best.

Even to people who have never graced the great opera houses of the world, the name and the voice of Placido Domingo are justifiably hailed. But for those who will stand in line all night to share the beauty of this man's singing, he is a legend.

A legend which can be heard from Hamburg to Paris, from Milan to New York.

"I can only sing five or six performances a month," he says. "In order to give of my best, I must make sure I do not sing too much. That is my responsibility to the people who try so hard to see me."

Placido is not simply the world's greatest tenor, but rather a complete musician who also possesses a marvelous voice.

At rehearsals, his mastery of the piano enables him to sit and play through the score, thinking of the emotions that words and music are attempting to communicate.

His experience as a conductor gives him objectivity, not only about his own interpretation of the part, but also of the total performance.

"To understand the part," he says, "one must first musically and dramatically understand the whole. I was lucky to have been given the talents to do this."

Placido Domingo also has an extremely good understanding of the watch he chooses to wear.

A Rolex Oyster GMT-Master in 18kt. gold.

"This watch is perfect for me," he says, "because it simultaneously tells me the time in two different countries...which is extremely useful considering the amount of traveling I have to do.

"And opera people all over the world are pleased too, because now I don't get them out of bed when I phone them. And, unlike me, this watch never needs a rest. You could say it's my favorite instrument."

For the complete musician.
The complete watch. By Rolex.

ROLEX


*Pictured: The Rolex GMT-Master. Available in 18kt. gold or stainless steel, with matching bracelet.*

*Write for brochure. Rolex Watch, U.S.A., Inc., Dept. 151, Rolex Building, 665 Fifth Avenue, New York, N.Y. 10022-5383. World headquarters in Geneva. Other offices in Canada and major countries around the world.*

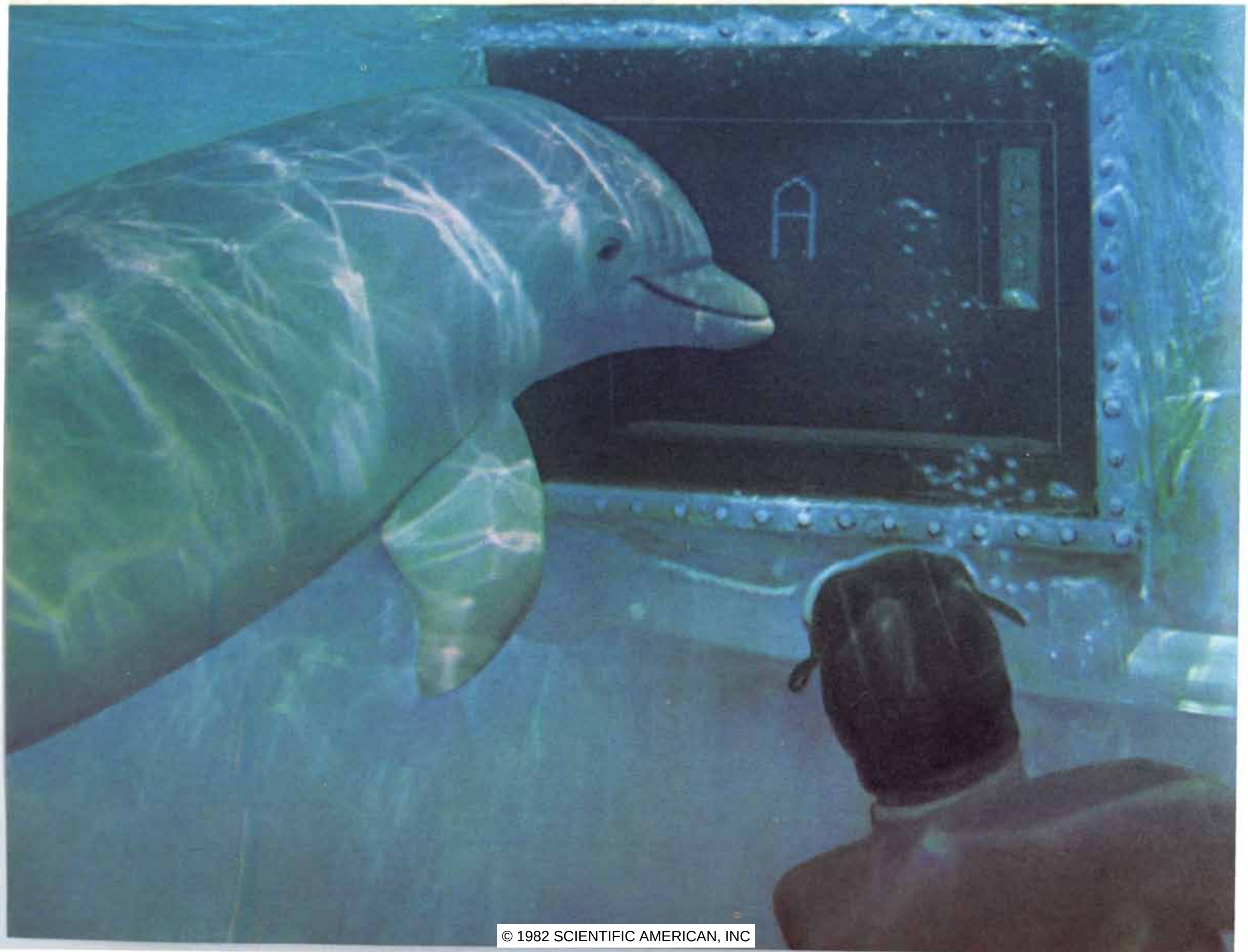
## Make waves with an Apple.

If you'd like to spend more of your research budget on research and less on computer costs, consider the discoveries of Dr. John Lilly and the Human/Dolphin Foundation.

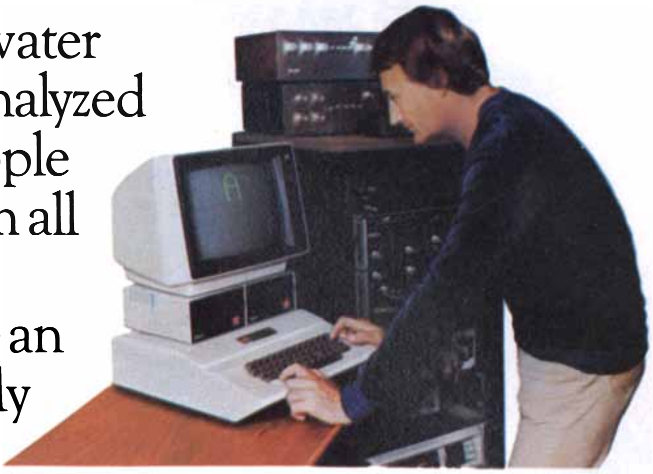
Dolphins vocalize at 2,000-40,000 Hz (compared with 300-3,000 Hz for humans) and "converse" 10-15 times faster than their bipedal brethren.

In 1968, Dr. Lilly's interspecies communication experiments stalled for lack of affordable computer power to bridge this gap. But today, with the help of Apple Personal Computers and a DEC® PDP/11, things are going swimmingly.

A new program called JANUS (Joint Analog Numeric Understanding System) uses a 48K Apple II Plus to generate dolphin-comprehensible wave forms matched to dolphin-



viewable symbols on an underwater screen. Dolphin responses are analyzed through a PDP/11. A second Apple monitors and analyzes data from all phases of the experiment.



The objective is to create an artificial language that is mutually intelligible to both species, with a beginning vocabulary of 48 sound/symbol morphemes associated with objects, locations and actions.

But, with all the micros available, why pick Apples?

Because they're inexpensive enough to allow the use of a stand-alone computing system for dedicated functions.

Because they're portable and rugged enough for field use in a wide range of environments.

And because they can be configured for many scientific applications with no special I/O devices. (Apple's IEEE-488 interface card will program and operate virtually any test, measurement or control instrument.)

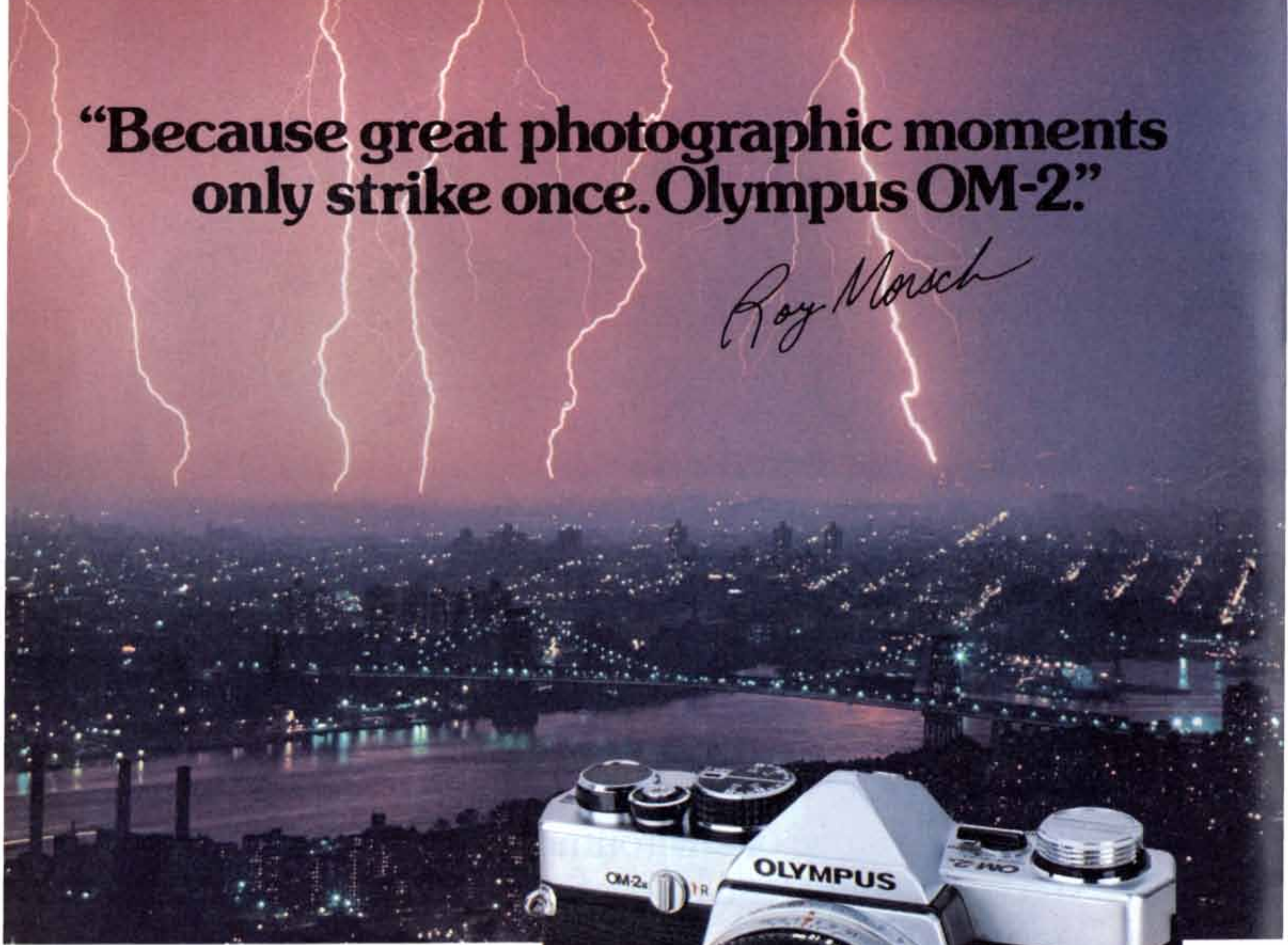
| | APPLE II | APPLE III |
|---|---|---|
| Maximum RAM | 64K | 256K |
| I/O | 8 expansion slots | 4 expansion slots + built-in interfaces for disks and printer. |
| Disk Drives | Up to 6 140K floppies. | Built-in 140K + up to 3 additional floppies. Available with ProFile™ 5MB hard disk. |
| Languages | BASIC, Pascal,™ PILOT, COBOL, Assembly. | Enhanced BASIC, UCSD Pascal,™ Assembly. |
| Keyboard | Fixed 96-character ASC II. | Fully programmable 128-character, includes numeric keypad. |

There's more software available for Apples than for any other microcomputer on the market. And more full service dealers worldwide.

Communicate with one and find out how an Apple can make waves for you, whatever your, ah... porpoise.

The personal computer. apple®

The Human/Dolphin Foundation is a non-profit research organization. For more information, write Dr. John Lilly, P.O. Box 4172, Malibu, CA 90265.
For the authorized Apple dealer nearest you, call (800) 538-9696. In California, call (800) 662-9238. Or write: Apple Computer Inc., 10260 Bandley Dr., Cupertino, CA 95014.







"My OM-2 worked perfectly where most automatic SLR's couldn't work at all," says award-winning commercial and *NY Daily News* professional photographer Roy Morsch. "It automatically made a perfect picture from 5 successive lightning bolts, without guesswork. My OM-2 never misses a shot!"

What makes OM-2 perfect for professionals like Roy Morsch makes it unbeatable for you. Its revolutionary OTF™ system sets exposures *as* you shoot, not before. Off-the-film. Automatically fine-tuning changing light in micro seconds. As no ordinary SLR can do.

OM-2 even makes pro-quality flash shots automatically easy. With its OTF™-controlled Olympus thyristor flash units. So you command more creative options than ever before. The OM-2's LED viewfinder ready light even confirms exposure accuracy. Automatically.

And OM-2 is more than an incomparable camera. It's the heart of the incomparable OM system. With 300 superb lenses and accessories to let you do almost anything, anytime. See OM-2 now at your Olympus dealer. You'll see why it's taking the pros by storm...

For information, write Olympus, Woodbury, NY 11797. In Canada: W. Carsen Co. Ltd., Toronto.

# OLYMPUS® OM-2

## The camera with the press card

Established 1845 SCIENTIFIC AMERICAN May 1982 Volume 246 Number 5

# Life Expectancy and Population Growth in the Third World

*Explosive growth has been caused by a rapid decrease in the death rate since World War II. The decrease has slowed, but even if it resumed its former pace, the ultimate effect on population size would be small*

by Davidson R. Gwatkin and Sarah K. Brandel

Should the industrialized countries make a strenuous effort to help reduce the death rate in the Third World? Life expectancy in the developing countries is now from 15 to 20 years less than it is in Europe and the U.S. Humanitarian considerations suggest that the gap should be narrowed as quickly as possible. It must be considered, however, that the rapid gain in life expectancy that has taken place in the developing countries since World War II has been one of the main causes of population growth. The rapid growth of population seriously aggravates economic and social problems. It is thus reasonable to ask whether a substantial increase in life expectancy in the Third World in a short period might not make matters worse.

The best current estimates suggest that the population of the Third World is likely to triple in the next century and thereafter remain stable. Even if life expectancy were to rise at what appears to be the fastest rate possible, the effect on the ultimate, stable population of the Third World would be small. The reason is that the rate of population growth in the developing countries has become increasingly insensitive to changes in the death rate. The most important influences on growth are future trends in fertility and the large numbers of young people now reaching childbearing age, mainly as a result of high fertility in the recent past. If population growth is to be kept to a minimum, attention to reducing the birth rate will be most important. No substantial demographic consideration need stand in the way of the industrialized countries' carrying out their responsibility to help increase life expectancy in the developing countries.

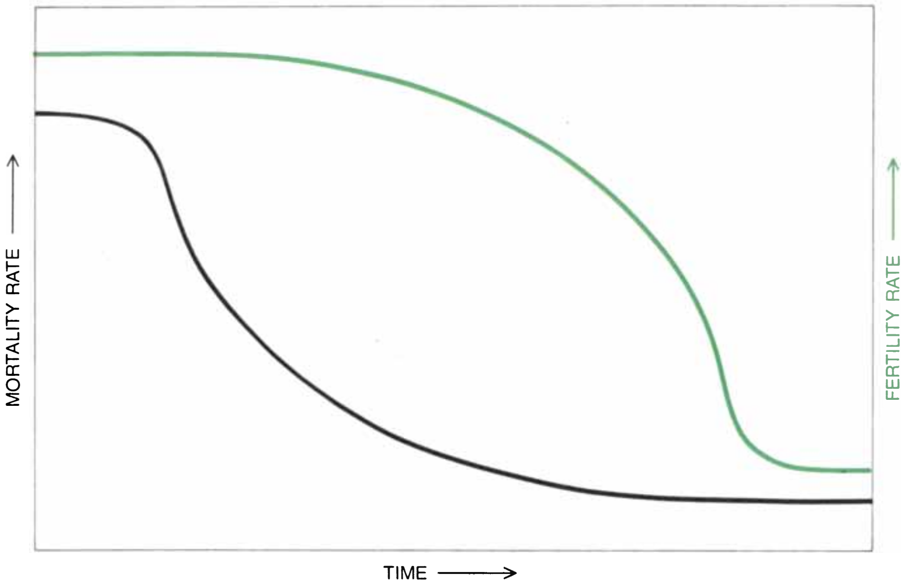
The rapid growth of population in the developing countries of Africa, Asia and Latin America is a consequence of the mortality declines that mark the first stage of the demographic transition. The transition is generally considered the central event in the recent history of the human population. During the transition a population evolves from a balance of high fertility and high mortality to a balance of low fertility and low mortality.

Before the demographic transition began the death rate was high enough to prevent significant long-term population growth in spite of high fertility. This situation prevailed throughout most of human history. In western Europe the death rate began to fall in the 17th and 18th centuries. In most countries, however, fertility remained high for several decades after mortality began to fall. As a result the population grew rapidly. The annual increase in most countries was between .5 and 1 percent. Eventually the death rate reached such a low level that further improvement became difficult. The birth rate began to fall in the course of the transition from an agricultural society to an industrial one. The growth of the population slowed when the birth rate and the death rate approached equality at a low level.

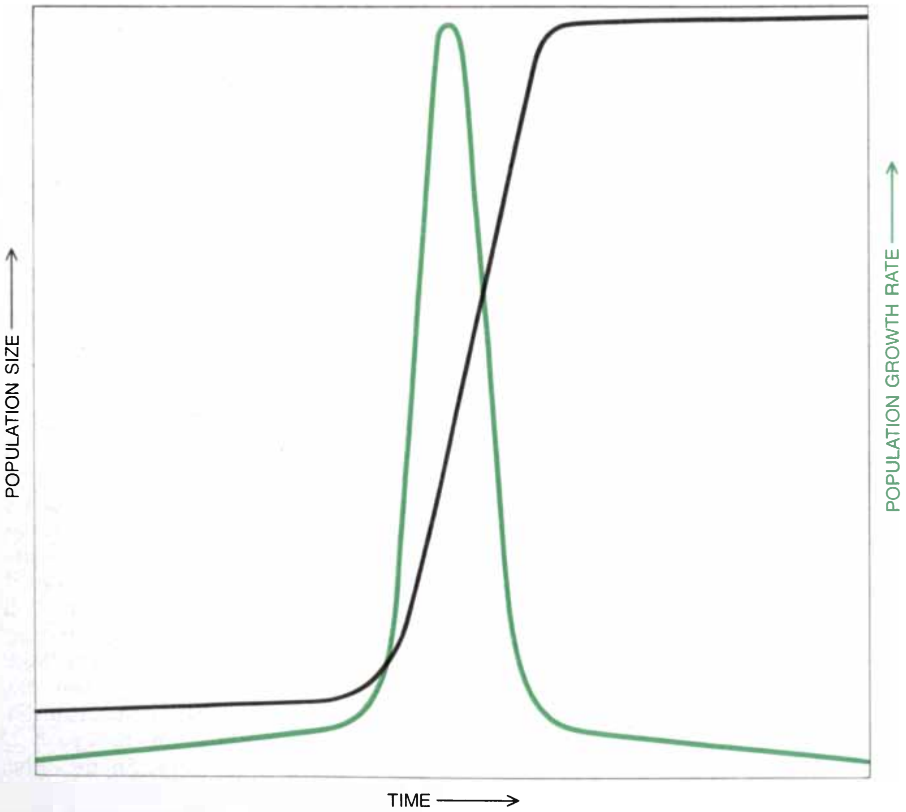
A widely used measure of mortality is life expectancy at birth. It represents the number of years a newborn infant can expect to live at the levels of mortality prevailing at the time of birth. Life expectancy at birth in most European countries is now between 72 and 73 years. This is more than twice what it was before the demographic transition began. A corresponding measure of births is the total fertility rate: the number of births a woman would have if she were to experience the prevailing age-specific birth rates throughout her reproductive lifetime. In Europe in the late 18th century the average total fertility rate was probably about five. It now ranges from about 1.4 to about 2.6 in European countries. Because not all children live to the age of childbearing an average of about 2.1 births per woman is required for a society with Europe's mortality level to maintain a constant population. Fertility is therefore at or below the replacement level in much of Europe.

A similar transition to low mortality and fertility has taken place in industrialized countries outside Europe, including Australia, Canada, Japan and the U.S.; it is well under way in places such as Argentina, Cuba, Hong Kong and Singapore, where economic development has reached a fairly high level. In most of the developing countries of Africa, Asia and Latin America, however, the demographic transition has not proceeded far. The death rate has been falling rapidly for the past 25 years or more. Fertility, on the other hand, has begun to decrease only recently. It is because of the disparity in the rates that the population is now growing rapidly. It remains to be seen how closely the course of the demographic transition in the Third World will resemble that in Europe.

Since World War II the declining mortality rate has led to a 50 percent increase in life expectancy in the developing countries: from less than 40 years to




**DEMOGRAPHIC TRANSITION,** often considered the central event in the recent history of the human population, is for the most part complete in the industrialized countries but is in mid-course in the developing ones, where its outcome is as yet uncertain. The effect of the transition on mortality and fertility is shown schematically. Before the transition a high birth rate is approximately in equilibrium with a high death rate. When the transition is complete, a low birth rate is balanced by a low death rate. The decrease in mortality generally precedes the decrease in fertility; during the interval the population can grow rapidly. Mortality declines in the developing countries have been particularly rapid since World War II. Widespread declines in fertility began only in the 1960's; so far fertility has been reduced by about 20 percent.




**POPULATION SIZE AND GROWTH RATE** show the influence of the demographic transition in this schematic representation. Before the transition the population is small and stable. When mortality decreases, the rate of growth increases. The population continues to grow rapidly for a considerable period after the reduction of fertility has caused the growth rate to begin to fall. At some time after the average woman is having about 2.1 children the population stabilizes again at a much larger size. In Europe in the 19th century the annual rate of population growth rarely exceeded 1 percent. In the Third World, where fertility was very high when mortality began to decline, annual growth rates as high as 4 percent have been recorded. The overall rate is now about 2.1 percent. Current projections suggest that by the time stability is achieved the population of the developing countries, now about three billion, will be about nine billion.

about 55 years. Until the mid-1960's fertility changed little. More recently it has begun to fall in most of the Third World, except for sub-Saharan Africa. By the late 1970's the average total fertility rate in the developing countries was about 4.7, some 20 percent less than it had been 15 years earlier.

A fertility decline of this magnitude, however, is not nearly large enough to compensate for the effect of the large mortality reductions of earlier years. The total fertility rate in the Third World is still about twice the replacement rate. In addition a decline in fertility takes some time to affect population growth. Hence the population of the Third World has continued to increase rapidly, and further rapid growth can be expected until well after the turn of the century even if fertility continues to fall. The present annual growth rate of slightly more than 2 percent yields a doubling of the population every 33 years. In some regions the growth rate is high enough to yield a doubling in less than 20 years.

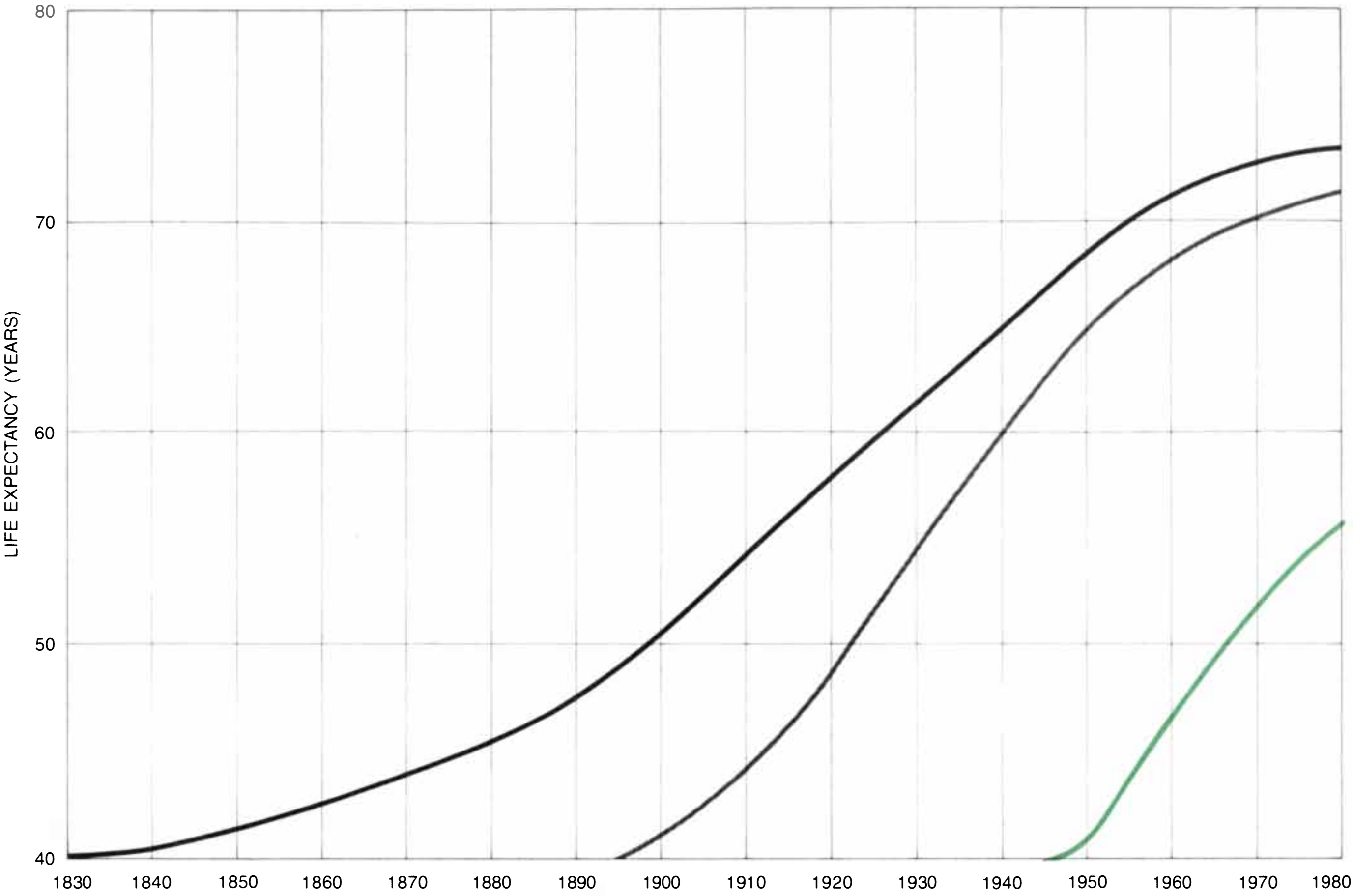
The current rate of population growth in the developing countries is several times that of the European countries at a comparable stage of the demographic transition. A principal reason for the disparity is the faster mortality decline in the Third World. Life expectancy has increased between two and five times as fast as it did in Europe in the early stages of the transition.

In the six western European countries for which adequate data are available, the average life expectancy rose from 40 years to 50 years between about 1830 and 1900. Further progress in this century raised it to about 65 years by the 1950's. Since then it has increased more slowly to the current level of about 73 years.

In southern and eastern Europe the increase in life expectancy began later. Shortly before 1900 life expectancy was still about 40 years. When the increase came, however, it was faster than that in western Europe. The advance from 40 to 50 years was accomplished in the first 25 years or so of the 20th century. Even more rapid increases followed, and the gap between southeastern Europe and western Europe was quickly closed. By the 1970's life expectancy in southeastern Europe was almost equal to that in western Europe.

Mortality did not begin to decline in the developing countries of Africa, Asia and Latin America until the process was already well advanced in the industrial countries. In a few developing countries, including India and some countries in Latin America, reductions began in the early 20th century. In most developing countries, however, substantial declines did not begin until after World War II.

In about 1950 the average life expectancy of an infant born in the Third




**INCREASE IN LIFE EXPECTANCY from 40 to 50 years occurs early in the demographic transition. The increase in average longevity was achieved in western Europe (*solid black line*) between about 1830 and 1900. A second wave of mortality declines in southern and eastern Europe (*gray line*) began later but proceeded faster. The increase from 40 to 50 years took place in the first 25 years or so of the 20th century. By the 1970's life expectancy was more than 70 years in almost all of Europe. The third wave of mortality declines, in the developing countries (*colored line*), was the last to begin but in its early stages was the fastest of all. In 1950 life expectancy in the developing countries was about 40 years. By about 1965 it had risen to 50, accomplishing in 15 years what had required 70 years or so in western Europe. Life expectancy in the Third World is now about 55 years, between 15 and 20 years less than in the developed countries.**

World was approximately 40 years. Thereafter an annual gain in life expectancy amounting to a year or more was recorded for as long as a decade in countries as diverse as Jamaica, Malaysia, Mauritius, Mexico, Sri Lanka and Taiwan. Such a rate apparently persisted for three decades in mainland China. Life expectancy in the developing world as a whole reached 50 years around 1965. What had taken about 70 years in western Europe in the 19th century was thus accomplished in the Third World in about 15 years.

Although the decrease in the mortality rate in the developing countries was rapid, it was not uniform; so far it has failed to reduce disparities among regions. According to the United Nations, in 1950 Latin America had the highest life expectancy of the world's three major developing regions: about 52 years. The annual increase there during the 1950's was between .5 and .6 year. In Asia life expectancy in 1950 was about 43 years, but the subsequent annual increase was greater: about .7 year. In Africa the initial life expectancy of 37 years was the lowest and the subsequent annual improvement was the smallest: between .4 and .5 year. By the mid-1970's this progress had brought life expectancy to about 47 years in Africa, compared with 57 in Asia and 62 in Latin America.

The average annual increase in life expectancy in all the developing countries during the 1950's was more than .6 year. In the next decade, however, it appears this rate was not maintained. According to UN estimates, by the late 1960's and the early 1970's the annual gain had fallen to about .4 year. The slowing of the improvement in life expectancy was observed in all three regions. As a result of the deceleration the gap between the developing countries and the developed ones, now about 17 years, is not being closed as quickly as it was earlier.

The reasons for the slowing of improvement in life expectancy are not fully understood; medical, social and economic factors seem to have played some role. Medical measures had only a limited influence on the increase in life expectancy in Europe in the 19th century. The better diet and living conditions that were made possible by increased agricultural productivity and economic development are thought to be largely responsible for decreases in mortality there. In the developing countries, on the other hand, drugs and insecticides formulated during World War II were introduced soon after the war ended. Such agents made it possible to control many communicable diseases and appear to be responsible for roughly half of the increase in life expectancy in the Third World. These measures succeeded in raising life expectancy above the level that might have been achieved through social and economic progress alone.

The available drugs and pesticides have apparently begun to exhaust their effects. The diseases they are capable of controlling have for the most part been brought under control. This may be one reason for the slowing of improvement in life expectancy. The diseases that remain have proved to be more difficult to control. A number of promising approaches under development could still make a significant difference. One approach consists in organizing nonprofessional personnel to administer simple medications in rural areas. Another involves teaching mothers to give children suffering from diarrhea an inexpensive liquid solution to combat dehydration. Work is being directed toward the development of vaccines against parasitic diseases such as malaria.

Diarrhea, pneumonia and malnutrition, which are the chief causes of death among the young in developing countries, are not amenable to medical solution alone. They are associated with a low standard of living, and improvements in diet, water supply, shelter, clothing and sanitary facilities will be needed to control them. The fact that the pace of development appears to have slackened in many parts of the Third World after the rapid progress in the 1950's and 1960's is probably another reason the increase in life expectancy has slowed.

Accelerating the pace of development is not easy. The available information, however, suggests that the potential gains in life expectancy could be great. Countries with an annual per capita income of $450, for example, appear to have a life expectancy about 16 years greater than countries with a per capita income of $220. Other things being equal, a poor country with an equitable distribution of income appears to have a life expectancy about six years greater than a country in which income is unfairly distributed. Education, particularly for women, is also important. Although increasing the pace of improvement in life expectancy in the Third World may require strenuous effort, there is little basis for thinking that only slow increases can be achieved. It may not be possible to surpass the rapid progress of the 1950's, but faster increases than those of the recent past are possible through social and economic progress and medical measures.

The greater annual gains in longevity achieved in the 1950's were a major cause of the current rapid population growth; the consequences of faster increases in life expectancy for population growth must therefore be a matter of concern. Even if life expectancy again begins to increase more rapidly, however, there are two reasons it is unlikely to cause population to grow as fast as past gains did.

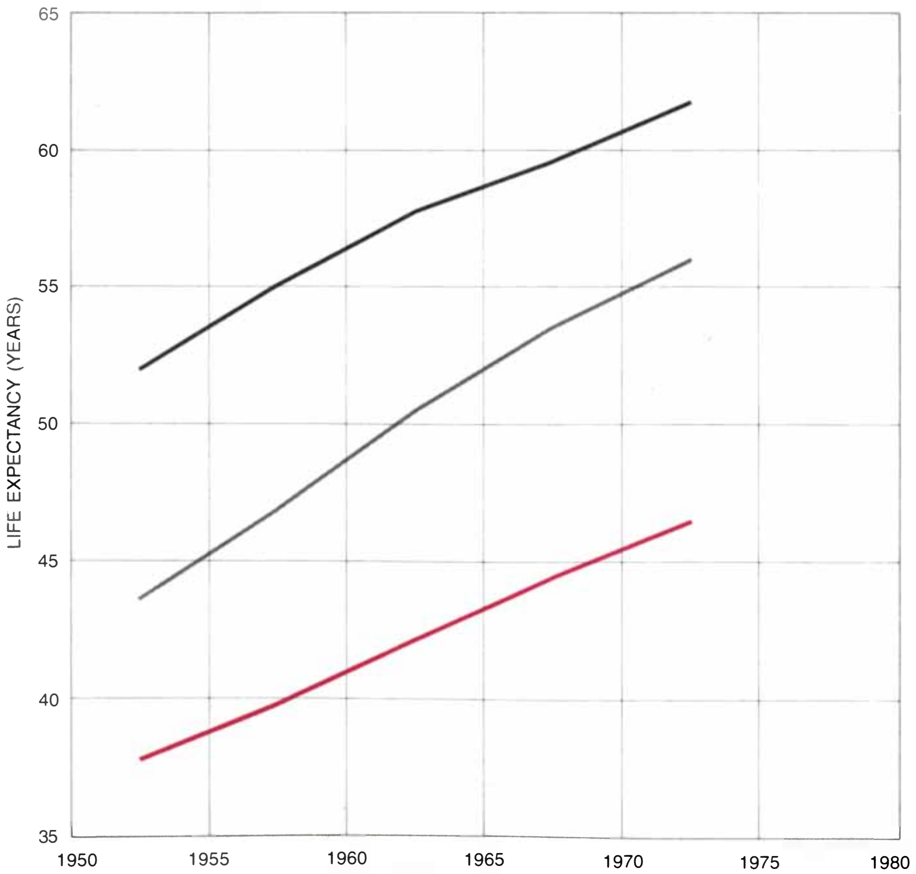
The first reason is that reductions in mortality have a diminishing influence on population growth as higher levels of life expectancy are achieved. The reduced effect is due to a shift in the age distribution of deaths. An infant saved from death from smallpox is enabled to live 50 or 60 years before dying of some other cause. A mother who would have died in childbirth gains another 30 or 40 years. A person 70 years old suffering from coronary insufficiency is granted another five years. As a result of such delays the average age at death and the proportion of deaths occurring among older people rise.

The average life expectancy by no means indicates the age at which most people die. When life expectancy is 35, for example, very few people live to 35 and then die. Instead, at the level of fertility that now prevails in the Third World, from 35 to 40 percent of all deaths are those of infants and children five or younger; 25 to 30 percent are those of people 60 or older. As life expectancy increases, the shift caused by delays in death produces a steady change in this pattern. By the time life expectancy reaches 75 at the current level of fertility only from 5 to 20 percent of deaths are those of infants and children 5 or younger, and from 70 to 80 percent are those of people 60 or older.

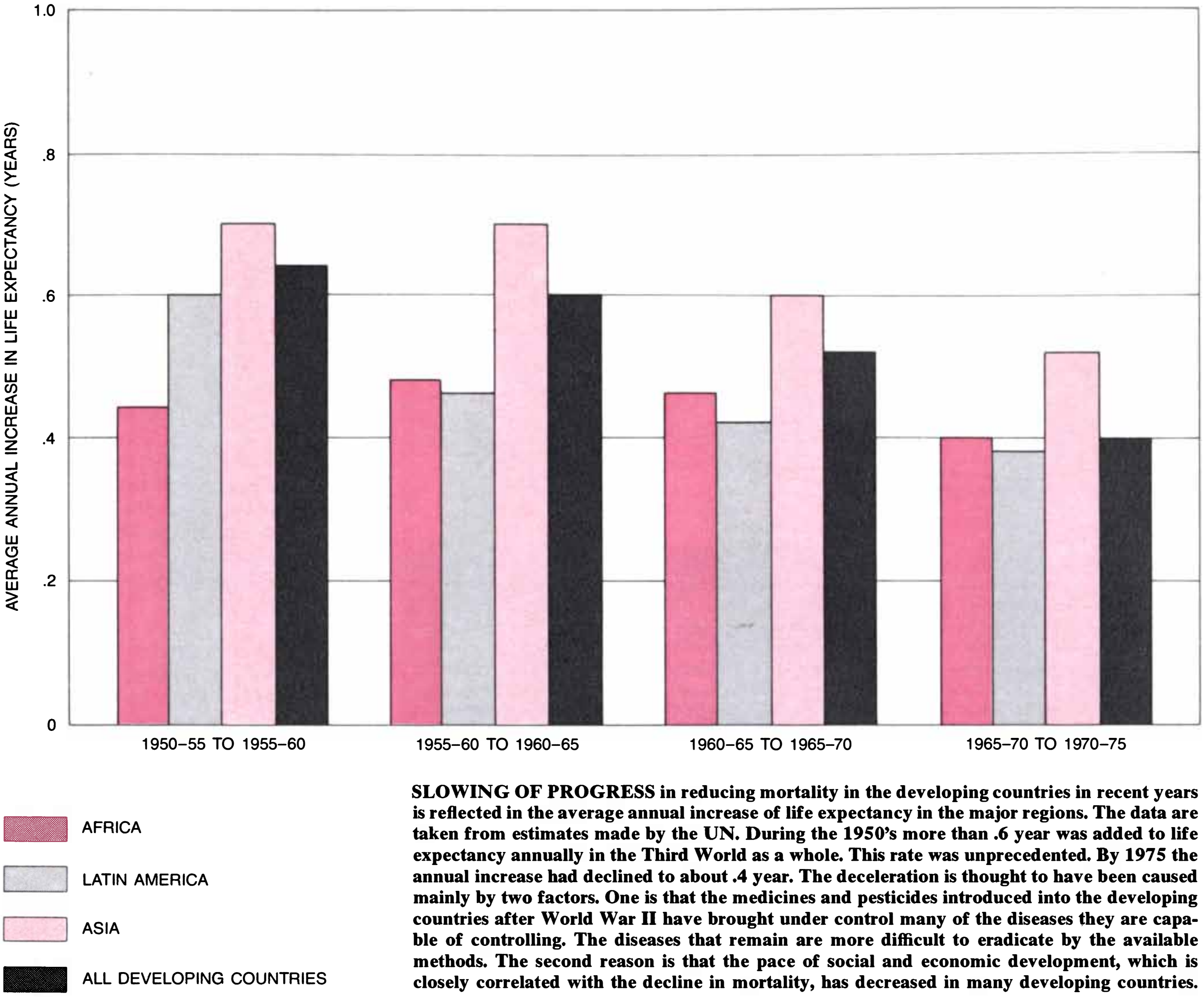
This shift has significant consequences for population growth because older people very rarely have children. Hence delaying the death of a 60-year-old adds only one person to the population. In contrast, when an infant is saved from death, the person saved is only the beginning. In later years that person's children, grandchildren and further descendants are added to the population in geometrically increasing numbers. In middle age some childbearing potential remains, and so the effect of delaying the death of a middle-aged person is greater than that of saving an older person but less than that of saving a child.

When life expectancy is low and a large proportion of those who die are young, an increase in the life span generally means that a large proportion of those saved are young. Because their childbearing years lie ahead, the effect on population growth is great. By the time higher levels of life expectancy are reached not many young people are dying. If life expectancy is to be increased further, it will be increased by extending the lives of older people, who are not likely to have children. Only slight population growth results.

For this reason, as the age of the people saved rises, further increases in life expectancy have progressively smaller effects on population. A gain of five years from 55 years (the average level of the Third World today) would yield an increase in the population only slightly more than half as large as a gain of five years from a life expectancy of 35 years. As Ansley J. Coale of Princeton University and later Nathan Keyfitz of Harvard University have shown, increasing life expectancy in a society from 70 to more than 100 years would have a negli-




**REGIONS OF THE THIRD WORLD show substantial variation in life expectancy. In the early 1950's the average person in Africa (*colored line*) could expect to live for about 37 years, the average person in Asia (*gray line*) for about 43 years and the average person in Latin America (*solid black line*) for about 52 years. The period between 1950 and 1975 was one of rapid increase in life expectancy in all three regions. Because the magnitude of the increase was similar in the three regions large disparities were maintained. In the early 1970's life expectancy in Africa was about 47 years, in Asia about 57 years and in Latin America about 62 years.**




**SLOWING OF PROGRESS in reducing mortality in the developing countries in recent years is reflected in the average annual increase of life expectancy in the major regions. The data are taken from estimates made by the UN. During the 1950's more than .6 year was added to life expectancy annually in the Third World as a whole. This rate was unprecedented. By 1975 the annual increase had declined to about .4 year. The deceleration is thought to have been caused mainly by two factors. One is that the medicines and pesticides introduced into the developing countries after World War II have brought under control many of the diseases they are capable of controlling. The diseases that remain are more difficult to eradicate by the available methods. The second reason is that the pace of social and economic development, which is closely correlated with the decline in mortality, has decreased in many developing countries.**

gible effect on long-term population growth.

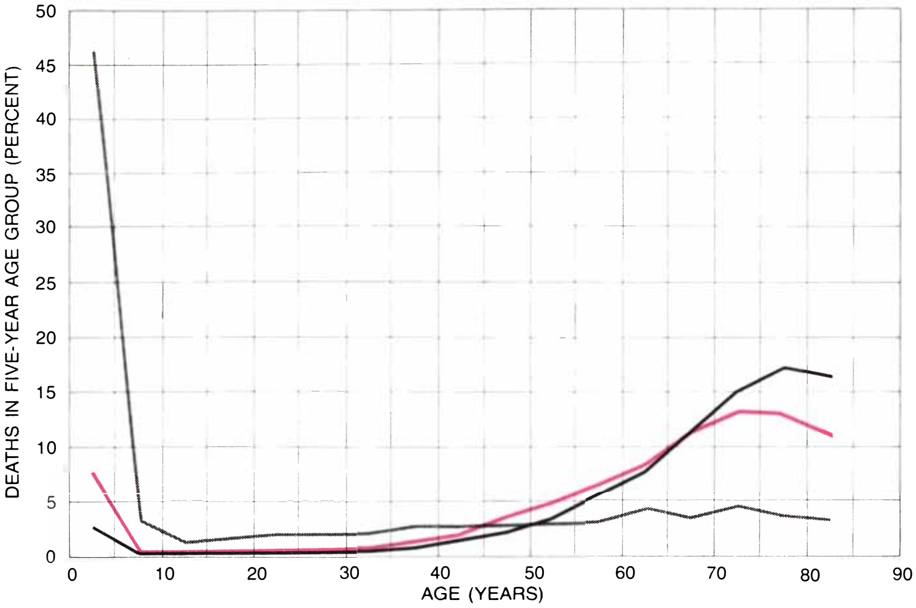
The second reason a faster increase in life expectancy would probably not have a large effect on population size is the reduction of fertility that has begun in the Third World and is expected to continue. In the 1950's and 1960's, when mortality decreased the most, there was little change in the high fertility of the developing countries. Since the mid-1960's, however, fertility in most developing countries has declined somewhat. Most demographers expect the decline to result in additional reductions of 20 to 35 percent in the Third World as a whole by the end of the century.

Lower fertility obviously has a direct effect on population growth. As we noted above, surviving infants and children affect population size not only because they are themselves added to the population but also because they have descendants. As fertility is lowered fewer children are born and each child leaves fewer descendants on the average, so that the consequences for population growth of enabling the children to survive become smaller. Moreover, lowered fertility contributes to the upward shift in the average age at death. If fewer infants are born, fewer die, and so a greater proportion of deaths are those of older people.

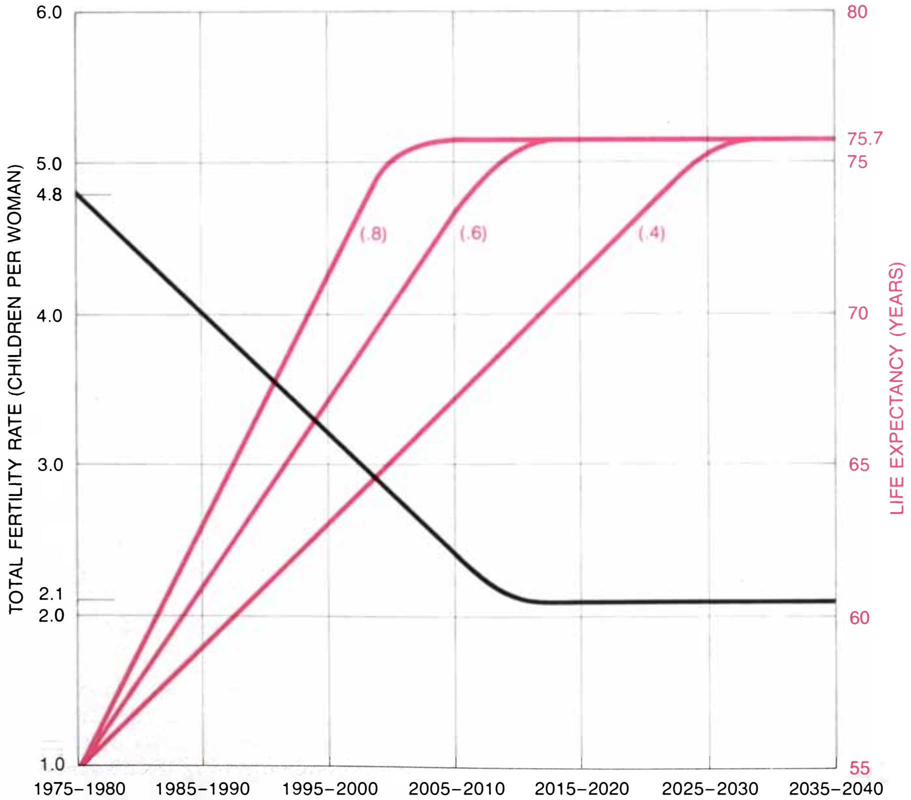
Two arguments thus suggest that the effect of a faster reduction in the death rate would be smaller than the effect of the rapid decline of the past. To see what the effect would be it is necessary to simulate mathematically the effects on population size of various rates of change in mortality. For comparative purposes it is useful to simulate the effect of changes in fertility as well. Such simulations require information or assumptions about the rates of change in fertility and mortality, the age pattern of mortality and fertility declines and the initial population size.

Useful starting points are the assumptions about future trends in mortality and fertility made by demographers at the UN, the U.S. Bureau of the Census and the World Bank. The most widely used estimates are probably those of the UN. The UN's 1978 medium projection of the world population is based on the assumption that the total fertility rate will decrease from about 4.7 to 3.4 and life expectancy will rise from 55 years to 63 years by the end of the century. This is equivalent to a decline in the total fertility rate of about .07 annually and an annual increase of about .4 year in life expectancy.

An age pattern of mortality decline can be selected from the standard reference work of Coale and Paul Demeny of the Population Council. From many sources of data on mortality in Europe and elsewhere Coale and Demeny constructed model life tables. The life tables are grouped in four families that represent different patterns of the distribution of deaths by age and sex at various levels of life expectancy. We utilized the "West" family, which is based on data from Australia, Canada, western Europe, Israel, Japan, New Zealand, the white population of South Africa, Taiwan and the U.S. In general it includes a somewhat faster decline in infant and child mortality than in adult mortality.




**AGE DISTRIBUTION OF DEATHS is quite different in the developed countries and the developing ones. The curves show the proportion of deaths in each five-year age group in Mexico (*gray line*), Sweden (*solid black line*) and the U.S. (*colored line*) in 1960. In the developing nations, as exemplified by Mexico, deaths of infants and children predominate. In 1960 more than 45 percent of all deaths were those of children younger than five. In the course of economic development the birth rate and the death rate fall and the age structure of the population changes. There are more older people and hence a larger number of deaths among them. Saving the lives of children has a much greater effect on population growth than extending the lives of older people. Increasing life expectancy from 35 to 55 years (which has already been done in the Third World) causes population to grow much more than increasing it from 55 to 75 years (which remains to be accomplished). The effect is reinforced by declining fertility.**




**ASSUMPTIONS made by the authors in projecting the size of the future population of the Third World are shown in graphic form. The number of children the average woman has is assumed to decline in a linear way from its initial level of 4.8 to the replacement level by about the year 2020. Three different annual rates of increase in life expectancy were employed to test their effect on population growth. The first rate is a continuation of the current annual increase of .4 year, reaching the maximum life expectancy of about 76 years in about 2025. The second annual rate is .6 year, one and a half times the current rate and just below the high rate of the 1950's; at this rate the maximum life expectancy would be reached in about 2010. The third rate, .8 year, or twice the current rate, yields the maximum longevity soon after 2000.**

It has been the most widely used of the four.

A model age pattern of fertility decline for the Third World and for each of its major regions has been formulated by the UN. The initial population of the developing countries assumed in our simulations was an estimate of 2.89 billion in 1975. Standard techniques yield a projection of the Third World population up to the year 2000 on the basis of these assumptions and estimates. When this is done, the population in 2000 is projected to be 4.80 billion.

With some further assumptions it is possible to extend the projections beyond 2000. A standard way of doing so is to assume that life expectancy will continue to rise until a maximum is reached and then remain constant. Fertility is assumed to fall until it reaches a level barely sufficient for the population to replace itself; thereafter fertility too remains steady. Through this procedure a stable population size is eventually achieved.

In such long-term projections the maximum life expectancy is usually taken to be about 75 years because the experience of the developed countries indicates that this level is attainable. If life expectancy continues to increase at the rate of .4 year annually, it will reach the maximum in about 2030. If fertility continues to decline at the rate assumed for the period up to 2000, it will reach thc replacement level in about 2020 and stabilize at the minimum replacement level of 2.1 in about 2030. The population will stop growing around 2100 at 8.46 billion, almost three times the 1975 population of 2.89 billion. The large increase results from past momentum as well as from fertility and mortality trends projected for the future.

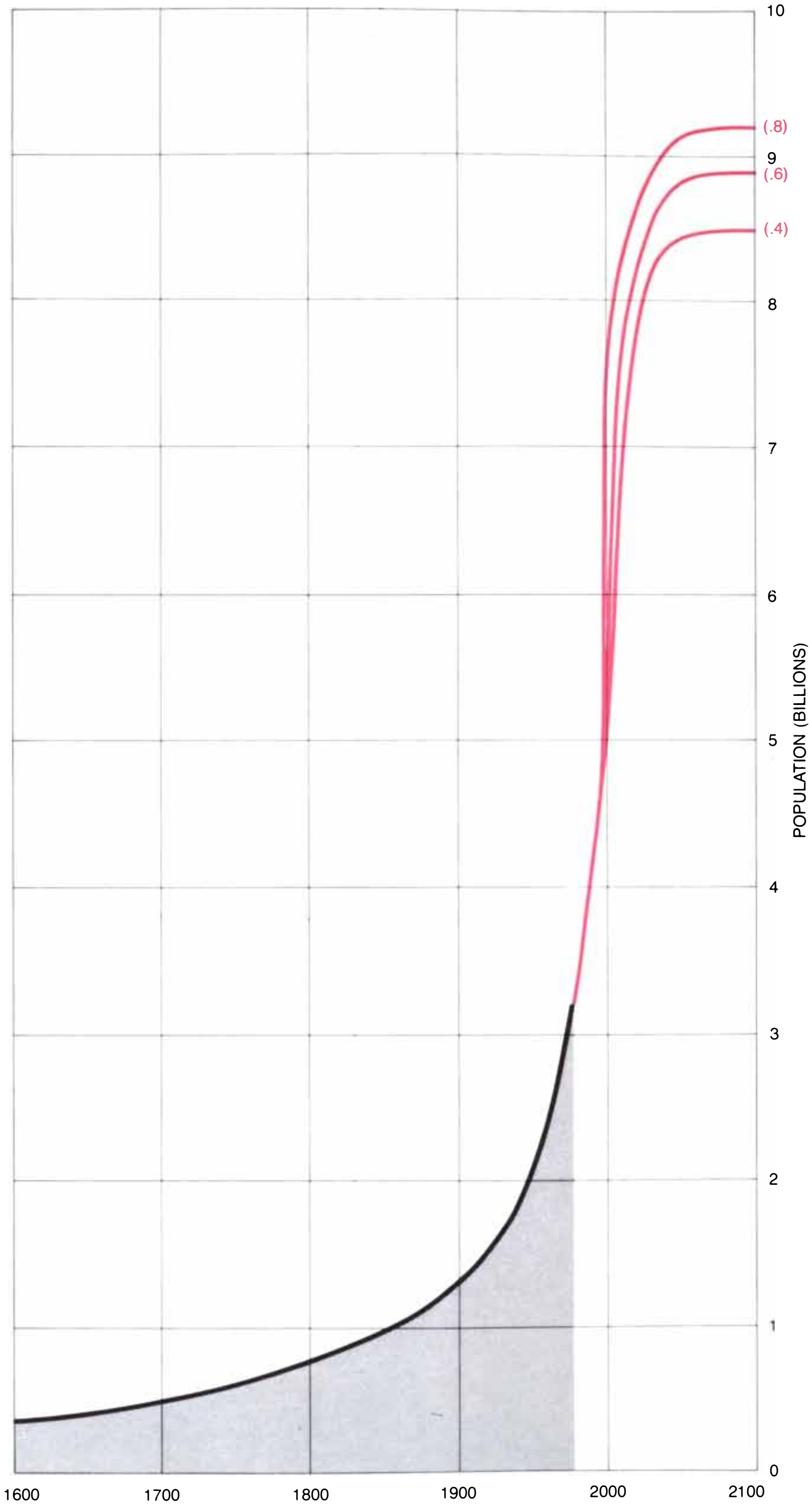
The population estimate based on this direct extrapolation from recent experience can serve as a base line with which other projections can be compared. In particular it is possible to ask what would happen if life expectancy were to increase considerably faster. The baseline estimate incorporates the assumption that life expectancy will increase .4 year each year. Such a rate implies that the deceleration of the past decade will be halted but that there will be no return to the higher level of the 1950's. Although it may not be possible for mortality to decline faster than it did in that decade, the considerations presented earlier suggest that through a strenuous effort the rate of the 1950's might be approached again. An annual gain of .6 year, half again as great as that of the base line, would approximate this rate.

If life expectancy in the Third World were to increase steadily by .6 year annually, maximum life expectancy would be reached in about 2015, or 15 years earlier than it would in the base-line model. Assuming the same initial population, the same rate and pattern of fertility decline and the same pattern of mortality decline as in the base-line estimate, the population in 2000 would be 4.90 billion. This is 100 million, or 2.1 percent, larger than the base-line estimate. When the population reached stability around 2100, it would include 8.94 billion people, some 480 million or 5.7 percent more than the base-line population. The difference is no larger because of the continued decrease in fertility and because the deaths delayed are increasingly those of older people.

If it were somehow possible to raise life expectancy by .8 year annually—twice the present rate—the ultimate increase in the population would be less than 10 percent larger than the stable population implied in the base-line estimate. The doubled rate yields a population of 5.00 billion in 2000, 4.2 percent larger than the base line, and 9.16 billion in 2100, 8.3 percent larger.

The effect of a faster increase in life expectancy would vary from region to region in the developing world partly because life expectancy is greater in some areas than in others and partly because fertility is lower or projected to fall faster in some regions. The greatest difference would be in Africa, where fertility is highest, life expectancy is lowest and these quantities are not changing as rapidly as they are in other areas. An annual increase in life expectancy in Africa half again as large as the base-line rate would yield a population 9.2 percent larger when growth stops. In Latin America the population in 2100 would be 1.8 percent larger. In East Asia the difference would be 1.6 percent and in South Asia 7.1 percent.

Of course there are aspects of future mortality trends other than the rate of increase in life expectancy that need to be considered. One factor is the maximum life expectancy that could be attained. Another is the age pattern of mortality decline. Neither factor, however, is likely to influence population growth much more than the rate of increase in life expectancy. If life expectancy in the Third World never rose to more than 70 years, for example, the stable population would be less than 5 percent smaller than the base-line population of 8.46 billion. Even the immediate decline of infant and child mortality to the low levels of northern Europe would yield a stable population less than 10 percent larger than the base line. It is possible to imagine a combination of the three factors that would affect future population size substantially. One such set of circumstances would be a very slow increase (or even a decline) in life expectancy, a low maximum life expectancy and gains in life expectancy resulting almost exclusively from reductions in infant and child mortality. Any such combination, how-




**STABLE POPULATION** ultimately reached in the Third World will be much larger than the current population but will vary little with the rate of increase in life expectancy. The graph displays the authors' projections of the future population based on three rates of increase in longevity. The shaded region shows historical data. The current population of the Third World is about 3.2 billion. According to the projections, it will stabilize by 2100. If the current annual rate of increase in life expectancy (.4 year) persists, the size of the stable population will be 8.5 billion. If the rate increases to .6 year, the population will be 8.9 billion. If the rate increases to .8 year, the population will be 9.2 billion. The rate of .6 year is probably the highest that can be achieved under current conditions. It yields an ultimate population only about 6 percent larger than the one associated with the base-line rate. The authors suggest that the actual difference might be smaller, because a decrease in mortality could lead to faster reductions in fertility.

ever, almost certainly lies well outside the range of possibilities considered likely by informed observers.

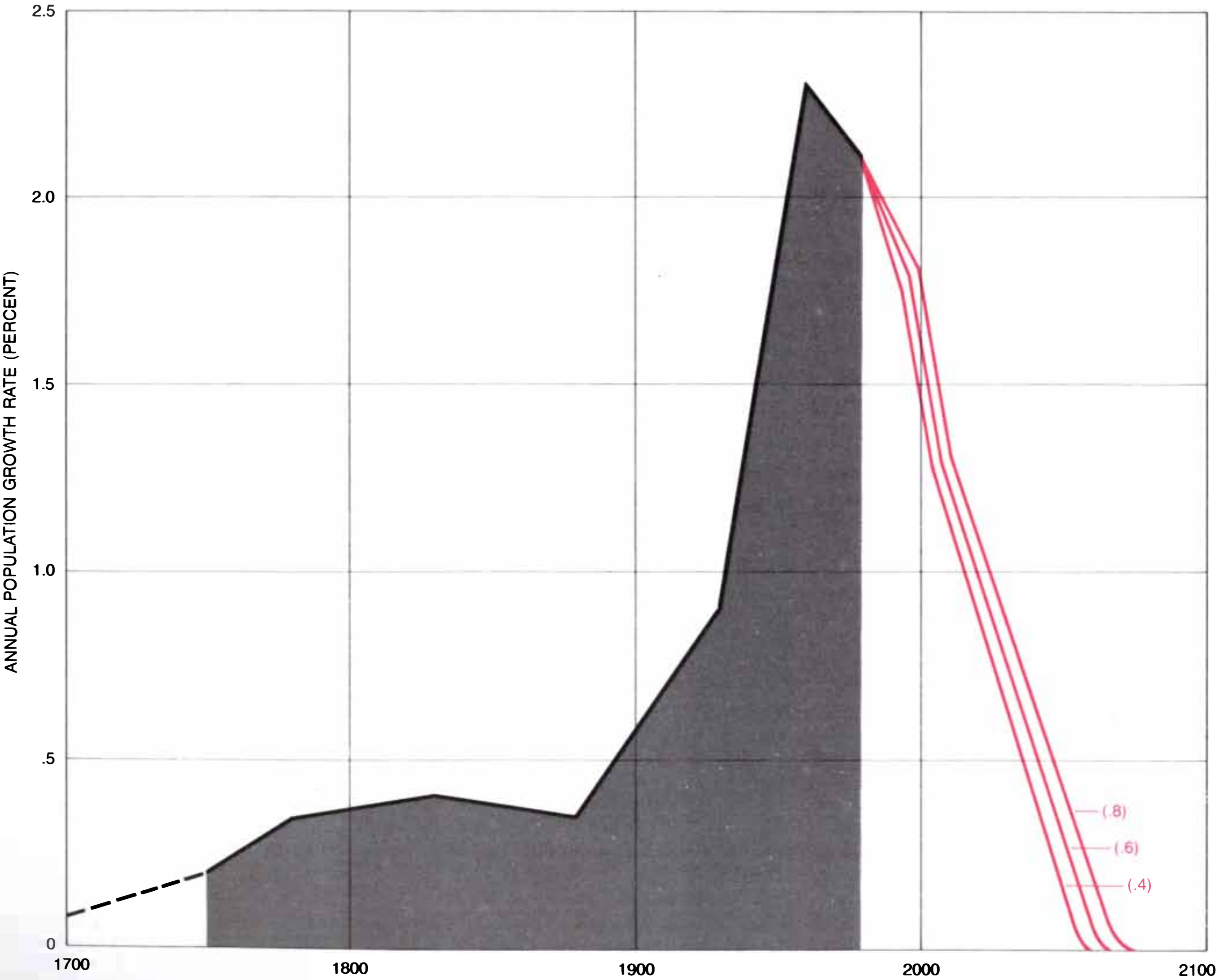
Compared with the increase in population caused by a faster decline in the death rate, the decrease caused by a faster reduction of fertility might be substantial. Suppose fertility were to decline half again as fast as is assumed in the base-line estimate. If that happened while life expectancy rose at the base-line rate of .4 year, the population in 2100 would be 18.8 percent smaller. This is more than three times the 5.7 percent difference caused by a 50 percent rise in the base-line rate of life expectancy increase.

Our main finding so far—that the population of the Third World at stabilization would be about 6 percent larger if life expectancy were to increase half again as fast—is only a rough approximation. It is probable that the actual increase in population would be smaller. The most important reason is that mortality and fertility are not independent of each other, contrary to the simplifying assumption employed in our projections. There is a great deal of information from the developing countries to suggest that faster reductions in mortality, particularly in infant and child mortality, would cause fertility to decline more rapidly.

The connection between fertility and infant mortality is particularly strong in societies where most women nurse their babies. Breast-feeding confers protection from pregnancy by inhibiting ovulation. The death of an infant removes the protection and exposes the mother to the possibility of another pregnancy. If a greater proportion of infants survive past the end of the period of breast-feeding, fertility is reduced.

In addition many observers have argued that when infant mortality is high couples have more births than the number of children they ultimately want in order to ensure that the desired number survive. When mortality declines, the number of births needed to yield a particular number of surviving children falls. Interest in family limitation grows and fertility is lowered.

Although the existence of such a relation has been established, its strength is a matter of controversy. The fact that the population of the Third World is growing very rapidly implies that the decline in fertility resulting from a lower death rate does not compensate fully




**RATE OF GROWTH of the population of the developing nations reached its peak in the 1960's and has now begun to decline. The historical record is represented in the shaded region. (Reliable data are available only for the period after 1750.) The growth rate was highest between 1940 and 1975, when mortality declines were fastest and fertility had not yet begun to fall. At its maximum the rate of population growth in the Third World was about 2.3 percent. Since the 1960's moderate reductions in marital fertility have reduced the rate to 2.1 percent. Future reductions in growth will be brought about primarily by changes in the birth rate. Increases in life expectancy will play a small role, as demonstrated by the close agreement of the curves representing population growth rates for three rates of improvement in life expectancy. Under all three assumptions the rate of population growth will return to zero several decades before 2100.**

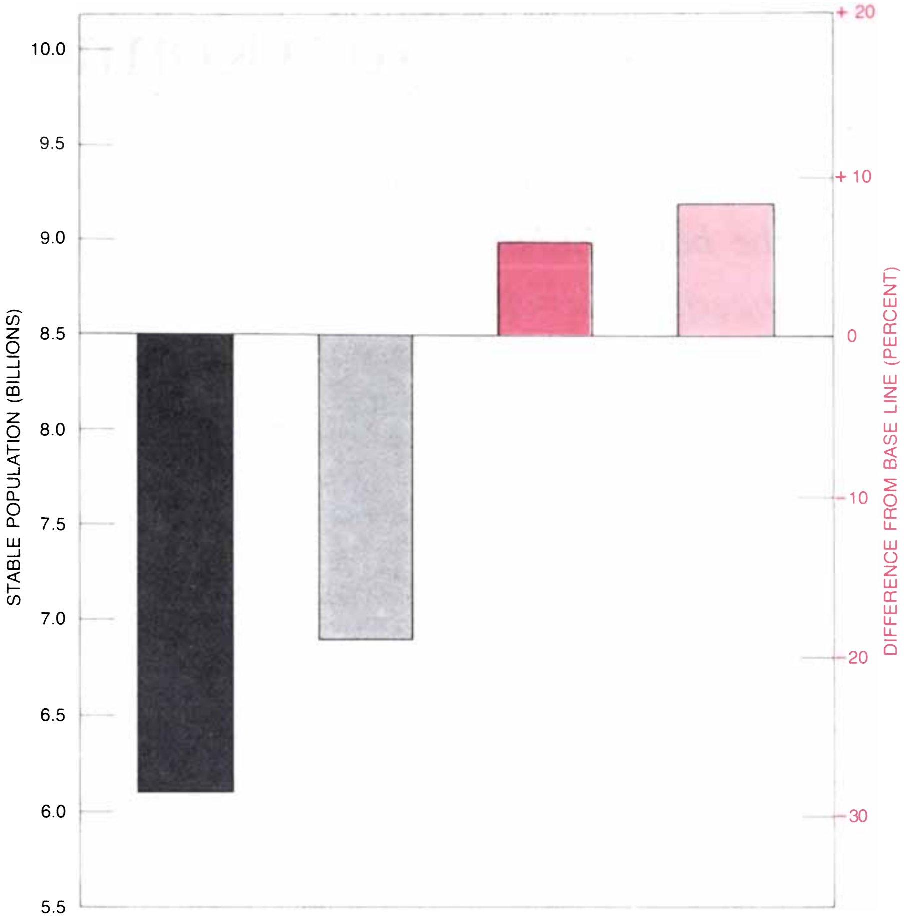
for the growth caused by reduced mortality. As Samuel H. Preston of the University of Pennsylvania has noted, there would be no population-growth problem in the Third World if the effect were fully compensating.

A partial effect, however, appears to be important in many situations. Because of the partial effect the 5.7 percent increase in stable population size caused by raising the rate of gain in life expectancy from .4 to .6 is almost certainly the extreme of any reasonable group of estimates. In light of the work done thus far on the relation between mortality and fertility a difference of between 3 and 4 percent is probably more accurate.

Neither this estimate of the additional population nor any other should be considered inevitable, however. The size of the population increase brought about by a faster mortality decline is not fixed. On the contrary, it will vary according to the policy measures by which it is brought about. One reason is that different measures will affect the mortality of various age groups differently. For the reasons discussed above programs that reduce the mortality of the young would have a somewhat greater impact than those that primarily affect the old. Even this effect is not as important as it might appear. As we noted above, the ultimate, stable population of the Third World would still be less than 10 percent larger than the base line even if most infant and child deaths were eliminated immediately. After taking into account the influence of a faster mortality decline on fertility the difference would probably be about 5 percent.

Much more significant is the influence on fertility of the particular programs by which mortality is reduced. Some measures can be expected to have only a relatively small and indirect effect on fertility. When infant mortality is reduced through immunization or the administration of antibiotics, for example, fertility can be expected to decline as couples realize they need fewer births in order to raise a given number of children. With more surviving children there would be more nursing mothers, which would also lower fertility. The fertility decline caused by such changes, however, would probably be fairly slow.

Other measures can exercise a stronger and more direct effect. Education for women can serve as an example. Increased literacy and education among women in rural societies has often been shown to be closely related to lowered infant and child mortality. From the resulting mortality decline would come the same motivation for lower fertility produced by an immunization program. At the same time the changes resulting from improved education could in many settings have an additional influence on fertility. Although there are exceptions, in many societies education for women has been shown to raise their aspirations for themselves and for their children. The aspirations can often best be fulfilled by delaying marriage, spacing births and reducing the number of children born.




**EFFECT OF FERTILITY on the size of the Third World population at stabilization by 2100 is likely to be far greater than the effect of mortality. The bars show the difference between the base-line population and the population that would result from a faster decline in mortality or fertility. If life expectancy rose half again as fast as it is now rising, the stable population (*dark color*) would be about 6 percent larger than the base-line population. If fertility fell half again as fast as it is assumed to fall in the base-line estimate, the stable population (*light gray*) would be 19 percent smaller than the base-line population. Doubling the two rates yields analogous results (*light color and dark gray*). The effect of such a decrease in fertility on future population size thus appears to be roughly three times the effect of a more rapid decrease of mortality.**

Family-planning services are another example. Family planning serves to reduce infant and child mortality by enabling couples to lengthen the intervals between births, which results in healthier children, and by reducing the number of additional births to older women who have many children. (The mortality rate is particularly high for children born in such circumstances.) In this way family planning influences fertility directly.

To emphasize measures such as these in efforts to reduce mortality would not necessarily be without cost. Nor is it now possible to state with certainty to what extent population growth would be affected by doing so. From what is now known, however, it does not seem unrealistic to believe that careful planning could create a package of measures designed to reduce mortality that would result in compensating fertility declines or perhaps even larger ones.

Because a difference of even 5 percent in the population of the developing world in 2100 is likely to represent between 400 and 500 million people, it would be highly advantageous if programs to increase life expectancy were designed to minimize their consequences for population growth. Even if this is not accomplished, however, what has been said suggests that the pace of mortality declines in the future is not likely to be the principal factor influencing the rate of population growth. Past trends have already created a momentum leading toward a Third World population nearly three times its present size before stabilization is achieved. Faster reductions in fertility could change this outlook somewhat, but faster reductions in mortality would have no more than a limited impact. Reservations about efforts to reduce mortality faster because of a concern for population growth are therefore misplaced.

# Quarkonium

*An "atom" made up of a heavy quark and an antiquark provides the best available system for examining the forces that bind together the elementary constituents of subnuclear particles*

by Elliott D. Bloom and Gary J. Feldman

The most convenient context for investigating the forces of nature is a system of two objects bound together by mutual attraction. The earth and the moon, for example, constitute the most readily accessible system in which to observe the gravitational force. The hydrogen atom, consisting of an electron and a proton, has long been an essential testing ground for theories of the electromagnetic force. The deuteron, made up of a proton and a neutron, represents a model system for studies of the forces in the atomic nucleus. Now there is a bound system in which to investigate the force that acts between quarks, the constituents of protons, neutrons and many related particles. The system is called quarkonium, and it consists of a heavy quark bound to an equally massive antiquark. The force at work in quarkonium is the strongest one known; it has come to be called the color force, and it is now thought to be the basis of all nuclear forces.

Of the various two-body systems the simplest in some respects is the artificial atom called positronium. It is made up of an electron bound to a positron (the antiparticle of the electron). Like the hydrogen atom, positronium is held together by the attraction of opposite electric charges, but it is a more symmetric structure. Whereas the proton of the hydrogen atom is larger and heavier than the electron, the two component particles of positronium are identical in size and mass. Several states of positronium have been observed; in each state the electron and the positron have a unique mode of motion, and as a result each state has a distinctive energy. From the spectrum of energy states it is possible to deduce certain characteristics of the electromagnetic force.

Quarkonium is closely analogous in structure to positronium. In quarkonium the bound quark and the antiquark are again identical in size and mass, and the allowed modes of motion in quarkonium are similar to those in positronium, so that an equivalent spectrum of energy states can be expected. In the past several years a major effort to detect the quarkonium states has been undertaken, and about a dozen have been found. In a qualitative sense they correspond exactly to the positronium states. On the other hand, a quarkonium system is smaller than a positronium system by a factor of 100,000, and its total mass or energy is larger by a factor of from 3,000 to 10,000. Moreover, from the details of the quarkonium spectrum it is apparent that the color force is not only stronger than electromagnetism but also more complex.

Quarks were introduced into physics in 1964 by Murray Gell-Mann and George Zweig, both of the California Institute of Technology. The aim of the quark model was to explain the diversity of the particles designated hadrons, which up to then had been considered elementary. Quarks can combine in two ways to form a hadron. Three quarks bound together make up a baryon, a member of the subclass of hadrons that includes the proton and the neutron. A bound system of a quark and an antiquark constitutes a meson. The lightest and commonest of the mesons is the pion; the quarkonium systems we shall describe are also mesons, but they are much heavier.

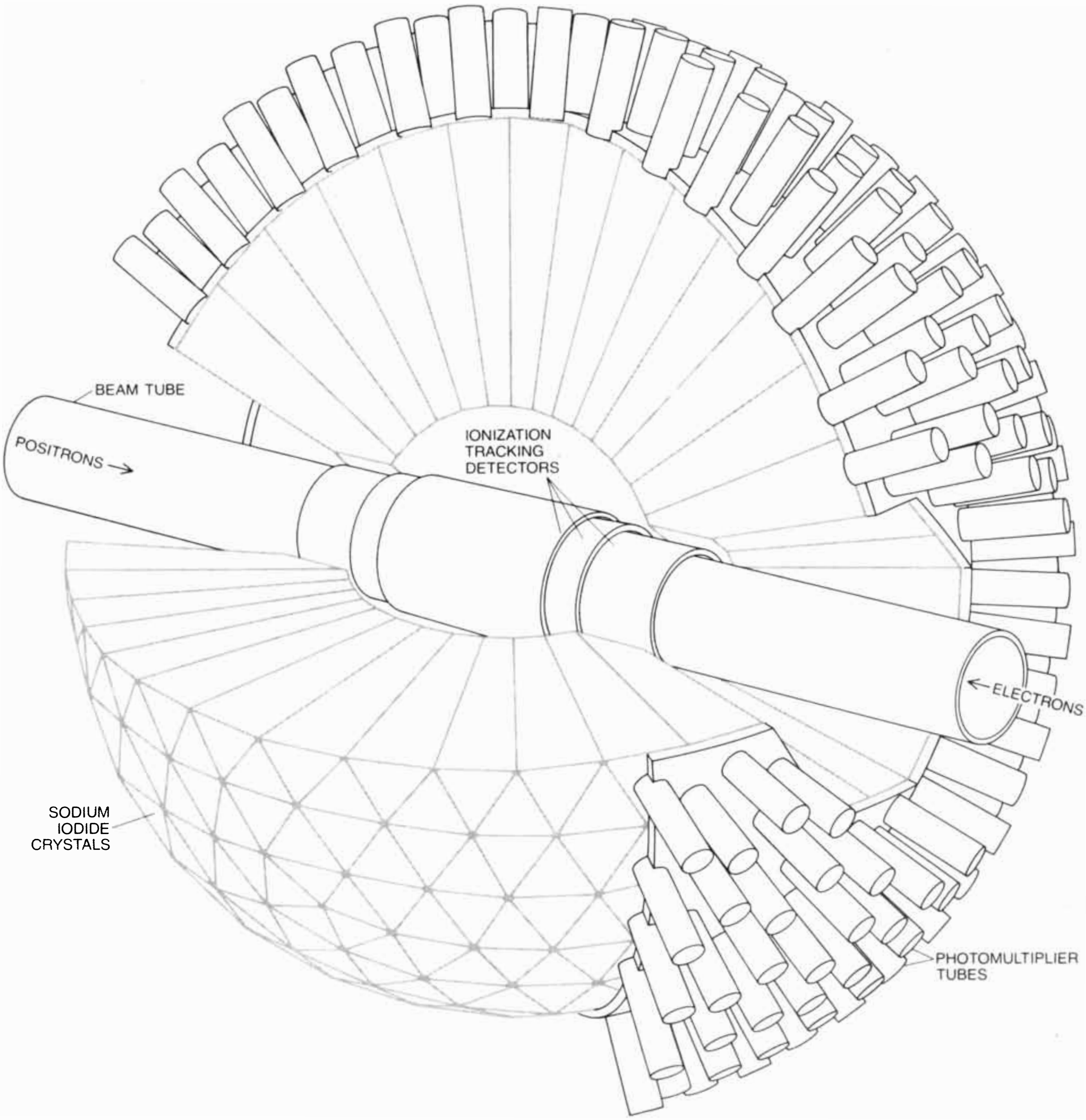
Originally there were supposed to be just three kinds of quark, designated up, down and strange, or *u, d* and *s*. All the hadrons known could be interpreted as combinations of these three quarks (and the antiquarks $\bar{u}$, $\bar{d}$ and $\bar{s}$). For example, the quark composite *uud* has all the properties of the proton, and the quark-antiquark aggregate $u\bar{d}$ can be identified as the positively charged pion. It might seem that this rich variety of hadrons would offer ample opportunities for exploring the force between quarks. Actually experiments with protons, neutrons, pions and other "ordinary" hadrons can yield only indirect information about the interquark force. The reason is that the *u, d* and *s* quarks are quite light; indeed, their mass, when it is expressed in energy units, is comparable to the binding energy that holds the quarks together in the hadron. As a result the quarks in an ordinary hadron move with a speed close to the speed of light, and calculations of their properties must be done with the complicated methods of the special theory of relativity. In general such calculations are too difficult to be practical.

What was needed was a bound system of heavier quarks, in which the binding energy would be small compared with the quark mass. The quarks would then move much slower than the speed of light, and the complications of the theory of relativity could be ignored. Such a nonrelativistic quark system was found in 1974 with the discovery of an extraordinary meson having a mass (in energy units) of 3,095 million electron volts (MeV). The meson was found almost simultaneously by two groups of experimenters. One group, at the Brookhaven National Laboratory, named the new particle *J*. The other group, at the Stanford Linear Accelerator Center (SLAC), called it $\psi$ (the Greek letter psi), and that is the name we shall use here.

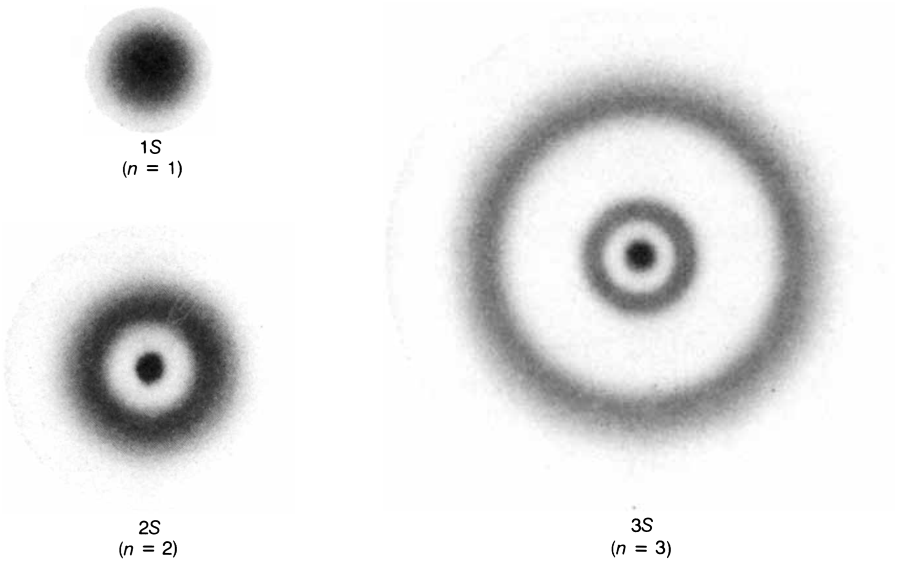
What is the $\psi$ meson? It seemed unlikely to be any combination of *u, d* or *s* quarks, in part because all the combinations with appropriate properties were already listed among the known hadrons. Some 10 years earlier, however, James D. Bjorken, who was then at SLAC, and Sheldon Lee Glashow of Harvard University had speculated that there might be a fourth quark flavor, which they had fancifully named charm (*c*). In 1970 Glashow and his colleagues John Iliopoulos and Luciano Maiani had argued on theoretical grounds that the charmed quark must exist and should be substantially heavier than the other quarks. Not long before the discovery of the $\psi$ Thomas W. Appelquist and H. David Politzer, who were then also at Harvard, had pointed out that a charmed quark and a charmed antiquark might form a nonrelativistic bound state. They had named the bound state charmonium, in analogy to positronium. The $\psi$ was soon recognized as a form of charmonium, that is, a meson with the quark composition $c\bar{c}$.

The discovery of charmonium stimulated a search for still heavier quarks. There was reason to think they would come in pairs, and the first two quarks after the known ones were designated bottom (*b*) and top (*t*). In 1976 Estia Eichten and Kurt Gottfried of Cornell University suggested that bottomonium (the meson with the quark constitution $b\bar{b}$) and toponium ($t\bar{t}$) should form nonrelativistic systems similar to charmonium but with a considerably richer spectrum of bound states. The first bottomonium state was discovered in 1977 at the Fermi National Accelerator Laboratory (Fermilab) near Chicago; it is called the Υ (the Greek letter upsilon), and it has a mass of 9,460 MeV. The toponium system has not been detected. If it exists, its mass must be greater than 36,000 MeV.

In the past eight years the charmonium spectrum has been surveyed in detail and a start has been made on the exploration of the bottomonium states. In addition a number of particles made up of a charmed quark in combination with a quark of another kind have been identified; such particles are said to exhibit naked charm. Recently indications of naked bottom have also been reported.




**CRYSTAL BALL DETECTOR is one of the instruments employed in exploring the structure of quarkonium, the state of matter made up of a quark and an antiquark of the same kind. The detector is mounted at a particle-storage ring; beams of high-energy electrons and positrons (or antielectrons) collide at the center of the detector and give rise to new particles, including quarkonium. The quarkonium then decays to yield still further particles, whose passage is recorded. The Crystal Ball is notably sensitive to photons, or quanta of electromagnetic radiation. The main elements of the detector are 732 crystals of sodium iodide, which surround the collision zone in an arrangement based on the symmetry of an icosahedron. (Additional crystals at the ends of the detector are not shown.) A crystal that absorbs a photon gives off light that can be registered by a photomultiplier tube. The direction of the photon and the amount of energy it deposits in the crystals provide information on the state of quarkonium that emitted the photon. Ionization tracking detectors aid in characterizing charged particles. The Crystal Ball was operated for about three years at SPEAR, a storage ring at the Stanford Linear Accelerator Center (SLAC). It is now being moved to DORIS, a higher-energy ring at the Deutsches Elektronen-Synchrotron (DESY) in Hamburg.**




**STATES OF A BOUND SYSTEM made up of a particle and an antiparticle are described by wave functions, which give the probability of finding the particle or the antiparticle at any point in space. The probability-density distribution is given for three states of positronium, the bound system consisting of an electron and a positron. Darker regions signify higher probability. Each state is identified by its principal quantum number, *n*, and by a letter code that indicates the orbital angular momentum of the system. The three configurations shown are *S* states, which have zero orbital angular momentum. Quarkonium forms analogous states.**

Here we shall be concerned primarily with the systems in which a heavy quark is bound to an antiquark of the same kind, as in $c\bar{c}$, $b\bar{b}$ and the conjectured $t\bar{t}$. It is to these three systems that we apply the generic term quarkonium.

In order to understand the forces at work in quarkonium it is instructive to begin with a discussion of the hydrogen atom and positronium. The chief force binding an electron to a proton or to a positron is described by Coulomb's law. The law states that the force is directly proportional to the product of the electric charges of the particles and inversely proportional to the square of the distance between them. The constant of proportionality, $\alpha$, is a measure of the inherent strength of the electromagnetic force; its numerical value is approximately 1/137.

Classical physics predicts that the particles in positronium or the hydrogen atom should fall toward each other until they collide. The prediction disagrees with experiments and was troublesome in early attempts to formulate a theory of the atom; it was resolved only with the development of quantum mechanics in the 1920's. An important step in this development was the model of the hydrogen atom put forward by Niels Bohr in 1913. Bohr simply postulated that the electron in a hydrogen atom can occupy only certain discrete orbits. In each allowed orbit the electron has a definite energy, and it can change its energy only by making an abrupt jump to another allowed orbit. In going from one orbit to another the electron either emits or absorbs a photon, or quantum of electromagnetic radiation, with an energy exactly equal to the difference in energy between the two orbits.

Bohr showed that the binding energy of the electron in a given orbit is equal to $E_0/n^2$, where $E_0$ is the binding energy of the smallest orbit and $n$ is a positive integer called the principal quantum number. The binding energy is the energy that must be supplied to separate the electron from the proton completely. For the hydrogen atom $E_0$ is about 13.6 electron volts. Values of $n$ are assigned to the orbits in sequence, beginning with the smallest orbit. For the smallest orbit $n$ is 1 and the binding energy is 13.6 electron volts; for the next orbit $n$ is 2 and the binding energy is $13.6/2^2$, or 3.4, electron volts, and so on.

The Bohr model can be applied to positronium, but because of the difference in mass between the proton and the positron the radius of a given positronium orbit is twice the radius of the hydrogen orbit with the same principal quantum number. The binding energy of positronium is therefore half as great: 6.8 electron volts for the smallest orbit. The sequence of states is the same, but all energies are reduced by a factor of two.

For a complete description of a bound system it is necessary to specify not only the energy but also the angular momentum. In a two-body system there are three contributions to the total angular momentum: the spinning of each body on its internal axis and the orbital motion of the two bodies around their center of mass. Each of these quantities can be represented by a vector oriented along the axis of rotation and having a length determined by the magnitude of the angular momentum. One could describe the angular momentum of a system by giving the magnitude and the direction of the three vectors. Other methods of representation, however, convey the same information and better reflect the symmetries of the system. In describing the states of positronium and quarkonium it is more illuminating to specify the orbital angular momentum, the sum of the two spin angular momenta and the total angular momentum.

In a large-scale system where the laws of quantum mechanics can be neglected the angular-momentum vectors can have any magnitude and orientation. In some respects the description of angular momentum in a quantum-mechanical system is simpler. Quantum mechanics does not allow an observer complete knowledge of an angular-momentum vector; one can measure only the absolute magnitude of the vector and its projection along one axis in space (although any axis may be chosen). Furthermore, the measured quantities can have only certain discrete values: an angular momentum must be either an integer or a half integer when it is expressed in units of the smallest possible quantity of angular momentum (which is equal to Planck's constant divided by $2\pi$). A final constraint is that the orbital angular momentum must always be less than $n$, the principal quantum number.

Orbital angular momentum can take on values given by the integers in the series 0, 1, 2, 3 and so on. In a compact notation introduced to describe atomic spectra some years before Bohr developed his model, the successive values of orbital angular momentum are designated by the letters *S, P, D* and *F*. The Bohr orbit and the orbital angular momentum of a state can therefore be specified by appending the appropriate letter to the principal quantum number. The lowest-energy state of positronium, for example, is designated 1*S;* it is the state with $n$ equal to 1 and no orbital angular momentum. The symbol 2*P* designates a state with $n$ equal to 2 and one unit of orbital angular momentum.

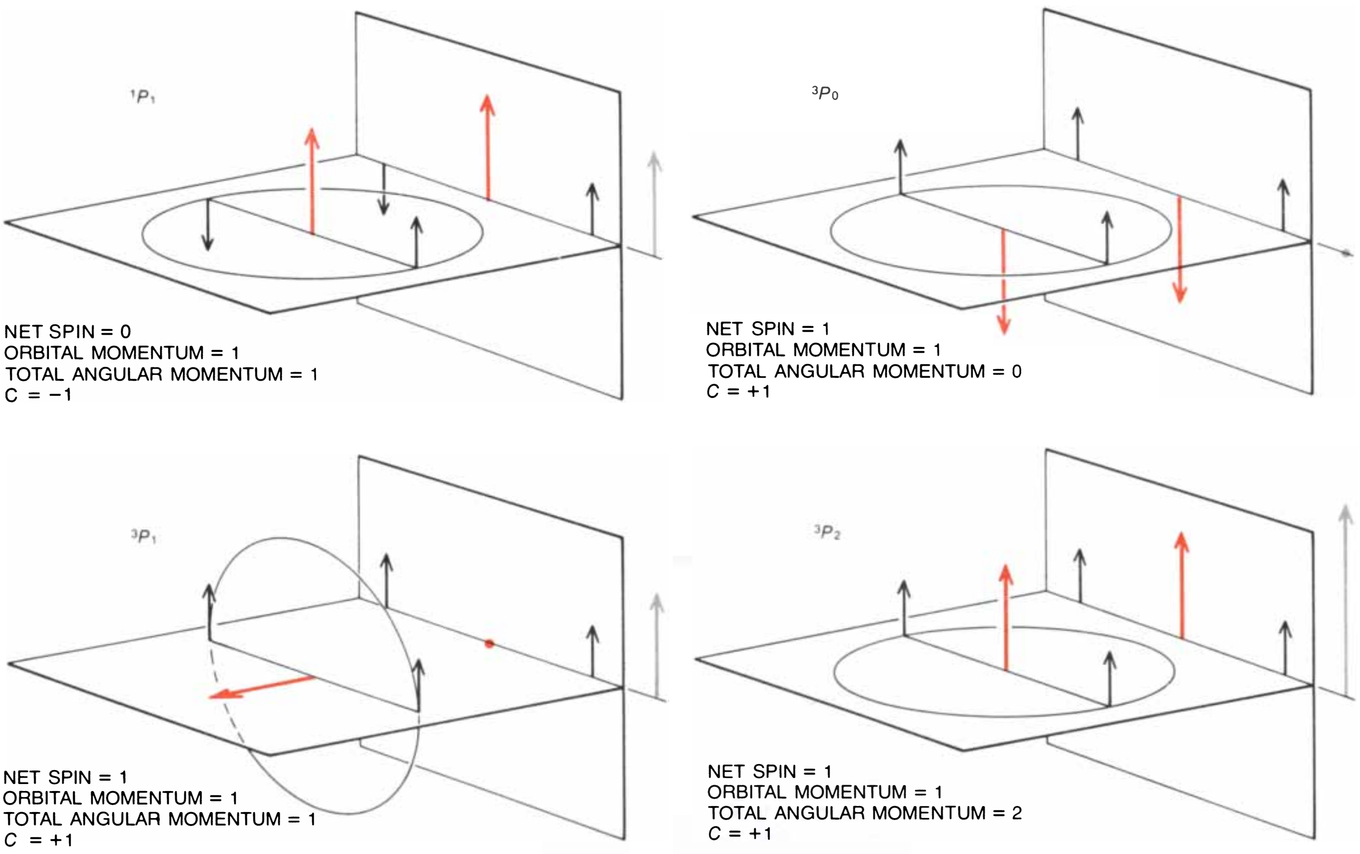
In both positronium and quarkonium the individual constituent particles have an intrinsic spin angular momentum of 1/2 unit. The spins can combine in just two ways. If they point in opposite directions, they cancel each other and the system has a net spin of zero. If they are parallel, they add and the net spin is 1. The total angular momentum of the system depends on the relative orientation of the orbital angular momentum and the net spin; to be more precise, it depends on the projection of the net spin vector along the axis of the orbital vector. When the net spin is zero, it can clearly have only one possible projection, namely zero. Such a state is called a singlet, since for a given value of orbital angular momentum there is just one possible value of total angular momentum. When the net spin is 1, it has three possible projections along the orbital axis: +1, 0 and −1. These values correspond to the three possible orientations parallel, perpendicular and antiparallel. A state of this kind is called a triplet, since for a given value of orbital angular momentum there are three distinguishable states that differ in total angular momentum.

The net spin and the total angular momentum are incorporated into the notation for quantum states by a system of superscripts and subscripts. For singlet states a superscript 1 is prefixed to the letter designation of the state; for triplet states the superscript is a 3. The total angular momentum is given by a subscript appended to the symbol. Thus $2^1P_1$ designates the state of positronium with $n$ equal to 2, one unit of orbital angular momentum and a net spin of zero. Since there is no net spin, the state is a singlet and the total angular momentum is necessarily equal to the orbital momentum. The triplet of states $2^3P_0$, $2^3P_1$ and $2^3P_2$ have the same principal quantum number and the same orbital angular momentum, and they all have a net spin with a magnitude of 1. They are distinguished by the orientation of the net spin with respect to the orbital axis, giving rise to the three values of total angular momentum.

The Coulomb force between an electron and a positron depends only on the charges of the particles and the distance between them; it is independent of their angular momentum. If the Coulomb force were the only one at work in positronium, all angular-momentum states with the same value of $n$ would have the same energy. Actually there are other forces. Both the spinning and the orbital motion of the electric charges give rise to magnetic fields, which in turn cause attractions and repulsions that can alter the energy of a state. For example, in the $1^1S_0$ state the two spins are antiparallel and the resulting magnetic interaction is an attractive one. In the $1^3S_1$ state the spins are parallel and the magnetic interaction is repulsive. As a result the former state has a slightly lower energy than the latter one.

In positronium the energy shifts associated with the angular momentum of the constituent particles are exceedingly small, at most a five-thousandth of the binding energy. Analogous effects can be observed in quarkonium, where the spin and the orbital motion of the quarks give rise to color magnetic fields. The energy differences in quarkonium are much larger, however. The overall binding energy of quarkonium is about 100 million times the binding energy of positronium, and the angular-momentum states are typically separated by a fifth of the binding energy.

The analogy between positronium and quarkonium can be extended to one additional phenomenon: the decay of the bound systems. When an electron and a positron come together, they annihilate each other, and the energy equivalent of their mass appears in the form of electromagnetic radiation. A quark and an antiquark of the same kind can annihilate each other in a similar way, although the energy initially takes the




**ANGULAR MOMENTUM of a bound system can be analyzed by means of vectors projected onto a plane. There are two contributions to the angular momentum: the spin of the particles and their orbital motion. Both electrons and quarks have an intrinsic spin of 1/2 (*black arrows*); in the states shown the orbital momentum is equal to 1 (*colored arrows*). If the projections of the spin vectors point in opposite directions, the net spin of the system is zero and the total angular momentum (*gray arrows*) is simply equal to the orbital momentum. The resulting state is called a singlet because the projections of the spin and of the orbital vectors can have only one possible sum. The state is designated $^1P_1$, where *P* signifies one unit of orbital angular momentum, the superscript 1 identifies the state as a singlet and the subscript 1 is the total angular momentum. If the spin vectors are parallel, the net spin is 1, and it can have three possible orientations with respect to the orbital momentum. Hence there is a triplet of states, designated $^3P_0$, $^3P_1$ and $^3P_2$. The superscript 3 denotes the triplet, and the subscripts 0, 1 and 2 give the total angular momentum when the projections of the net spin and of the orbital vector are respectively antiparallel, perpendicular and parallel. Another property of the bound system, called the charge-conjugation number (*C*), is determined by adding not the projections of the net spin and the orbital angular momentum but rather their absolute magnitudes. When the sum is an odd number (as it is in the state $^1P_1$), *C* has a value of −1; when the sum is even (as it is in the $^3P$ states), *C* is set equal to +1.**

form of radiation associated with the color force. For the description of these events, however, the Bohr model of the atom is not adequate.

The Bohr model has been supplanted, of course, by a more refined version of quantum mechanics. In the modern view the electron and the positron do not have a definite orbit and indeed cannot be assigned a definite position at any given instant. Instead one can calculate only the probability of finding a particle at a given point in space. The probability is defined by a mathematical function called a wave function, which is different for each energy state. For the $S$ states of positronium the rate of decay by mutual annihilation is proportional to the probability of finding the two particles at the same point.

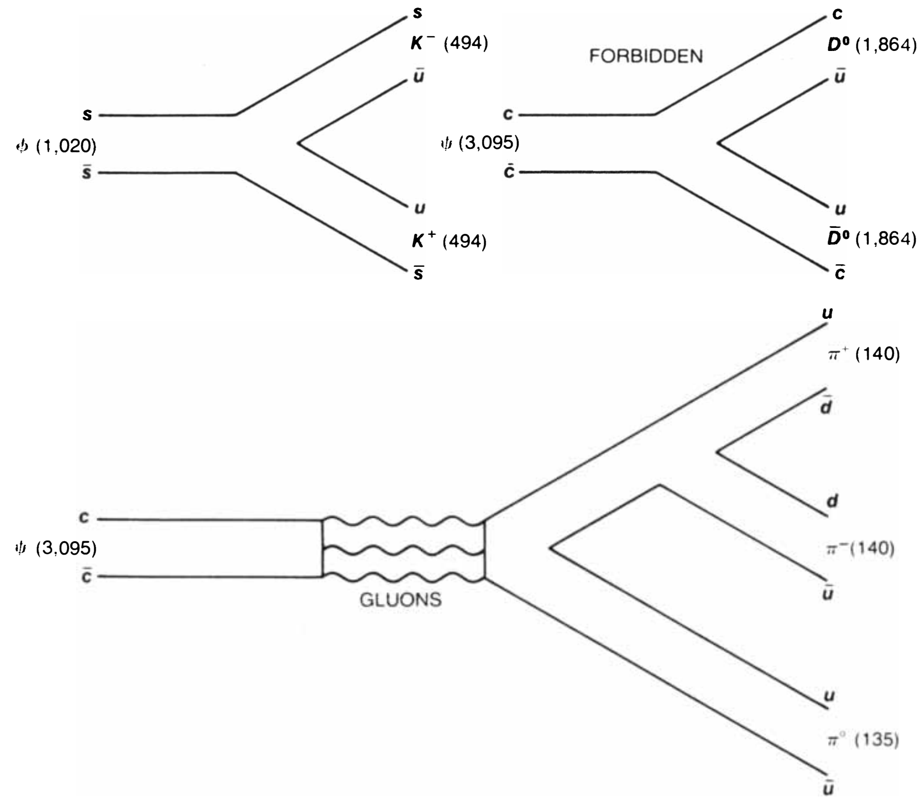
The decay is governed by a number of conservation laws. For example, the total electric charge of all the particles after the decay must be the same as it was before the decay. Energy and linear momentum must also be conserved. The energy of the positronium system is roughly 1 MeV; the linear momentum, since the particles can be thought of as moving in opposite directions when they collide, is zero. The photon can carry any quantity of energy, but because its mass is zero its linear momentum is invariably equal to its energy. If positronium were to decay to yield a single photon, energy and linear momentum could not both be conserved. For this reason the decay always gives rise to at least two photons. If the two photons are emitted in opposite directions, each one can carry half of the energy of the system and their linear momenta cancel.

A property of a wave function called the charge-conjugation number leads to a further constraint on the number of photons emitted. Charge conjugation is an imaginary operation in which all particles are converted into their antiparticles. When the operation is applied to positronium, the electron becomes a positron and the positron becomes an electron. Hence the system is still an atom of positronium, but the constituent particles have exchanged identities. In some states of positronium such an exchange is of no consequence; in other states, however, it changes the sign of the wave function. There is a method for keeping track of how a given state responds to charge conjugation. When the charges are interchanged, the wave function of the state is multiplied by the charge-conjugation quantum number, which has a value of $+1$ for the states that remain unaltered and $-1$ for the others. The charge-conjugation number, or $C$, is conserved in all interactions mediated by either the electromagnetic force or the color force.

The photon, which is its own antiparticle, has an intrinsic charge-conjugation number of $-1$. For a system of several photons or other particles the values of $C$ for all the particles are multiplied to yield the total charge-conjugation number. It follows that a state of positronium with $C$ equal to $+1$ must decay to yield an even number of photons; the actual number is almost always two. A positronium state with $C$ equal to $-1$ must yield an odd number of photons. As we showed above, the decay into a single photon is forbidden by the conservation of energy and momentum, and so the minimum odd number is three.

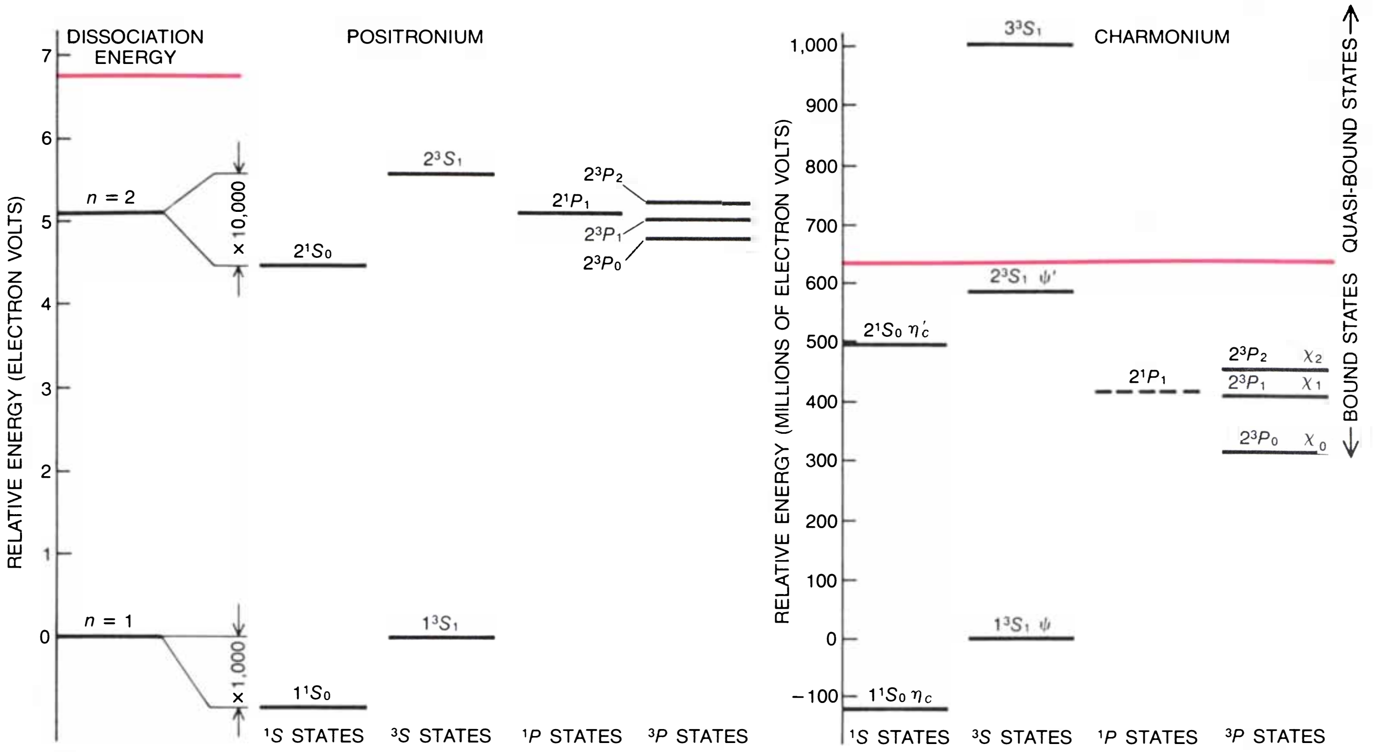
What determines the charge-conjugation number of a positronium state? It can be determined by adding the absolute magnitudes of the orbital angular momentum and the net spin. If the sum of the absolute magnitudes is an even number, $C$ is $+1$; if the sum is odd, $C$ is $-1$. Thus for the $2^1P_1$ state, with one unit of orbital angular momentum and no net spin, the sum is odd and $C$ is $-1$. In comparison the $2^3P_0$, $2^3P_1$ and $2^3P_2$ states all have one unit of orbital angular momentum and one unit of net spin, yielding an even sum and a charge-conjugation number of $+1$.

The various properties of the wave function of a positronium state influence the lifetime of the state. For an $S$ state the intensity of the wave function at the point corresponding to an interparticle separation of zero determines the probability of the particles' colliding, which is a precondition to their annihilation. For states whose wave functions have the same intensity at zero separation, such as the $1^1S_0$ and $1^3S_1$ states, the number of photons that must be emitted is an important factor in determining the lifetime. In effect it takes longer to emit more photons. The $1^1S_0$ state, which has $C$ equal to $+1$, decays into two photons in about $10^{-10}$ second. The $1^3S_1$ state has the same probability of electron-positron collision, but because $C$ is $-1$ for this state it must decay into three photons; the lifetime is about 1,000 times longer.

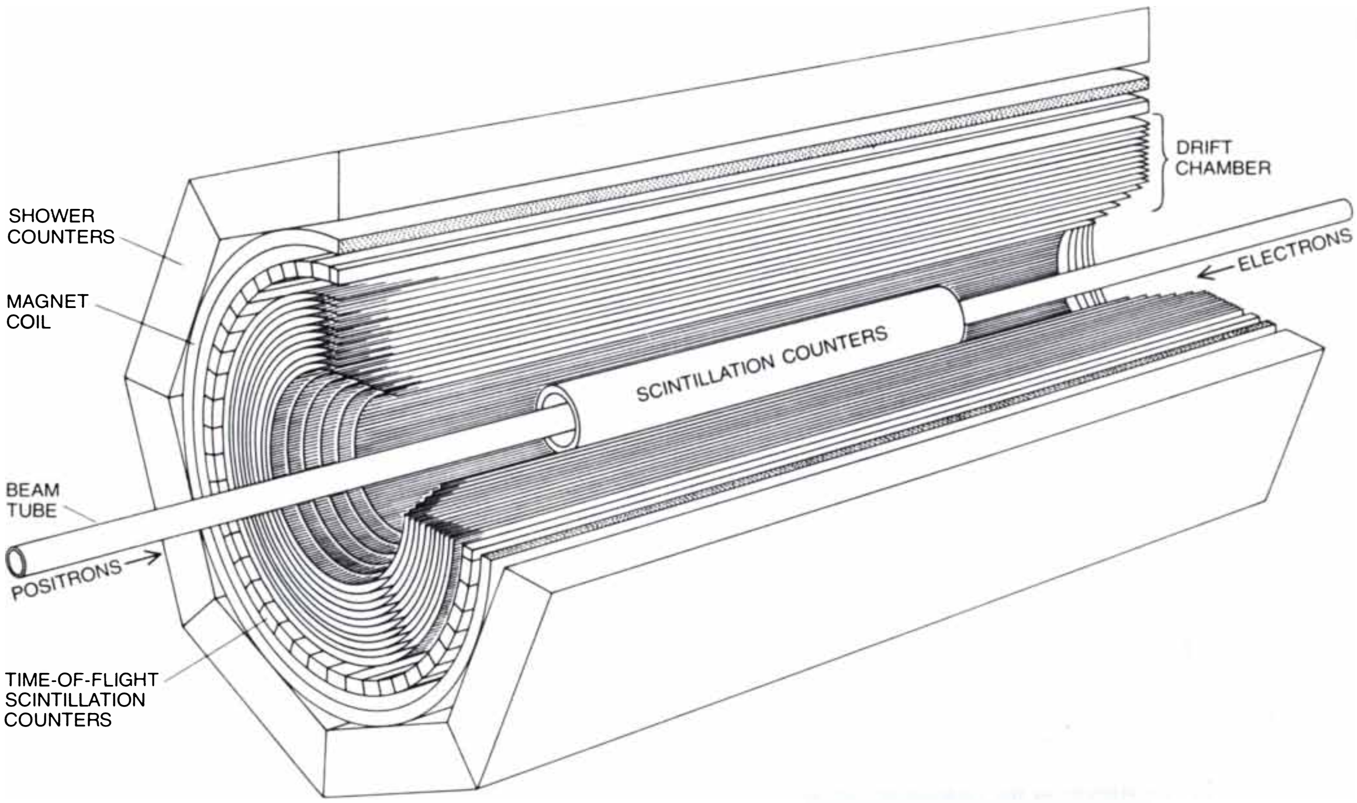



**QUARK-ANTIQUARK SYSTEMS can be classified as bound or quasi-bound according to their mode of decay. The particle called the $\phi$ (phi) meson is quasi-bound. It is the $1^3S_1$ state of a strange quark $s$ and a strange antiquark $\bar{s}$ and has a mass of 1,020 million electron volts (MeV). When the $\phi$ decays, the $s$ and the $\bar{s}$ merely separate and a new quark-antiquark pair is formed; the products are the mesons designated $K^+$ and $K^-$, with a combined mass of 988 MeV. The remaining 32 MeV is converted into kinetic energy. For the $\psi$ (psi) meson the analogous decay process is forbidden by the conservation of energy. The $\psi$ is the $1^3S_1$ state of a charm quark $c$ and a charm antiquark $\bar{c}$, and its mass is 3,095 MeV. The lightest particles incorporating charm quarks are the $D^0$ and $\bar{D}^0$ mesons, with a combined mass of 3,728 MeV. Making such a pair from the $\psi$ would leave a deficit of 633 MeV. The $\psi$ decays by annihilation into three gluons, which give rise to lighter quark-antiquark pairs and ultimately to ordinary hadrons. The $\psi$ is a bound state of matter called charmonium, analogous to positronium.**

The positronium system can be described completely and with remarkable accuracy by the theory called quantum electrodynamics, or QED. The theory attributes the force between two electrically charged particles to the exchange of virtual photons emitted by one particle and absorbed by the other. The photons are said to be virtual because they can never be detected direct-




**SPECTRUM OF ENERGY STATES** is similar in positronium and charmonium, but the scale of the energy differences in charmonium is greater by a factor of roughly 100 million. The energy of a state is determined by the principal quantum number $n$ and by the orientation of the particle spins and the orbital angular momentum. In positronium the various combinations of angular momentum cause only minuscule shifts in energy (shown by expanding the vertical scale), but in charmonium the shifts are much larger. All energies are given with reference to the $1^3S_1$ state. At 6.8 electron volts positronium dissociates. At 633 MeV above the energy of the ψ charmonium becomes quasi-bound because it can decay into $D^0$ and $\bar{D}^0$ mesons. The $2^1P_1$ state of charmonium has not been detected experimentally.




**MARK II DETECTOR** at the SPEAR storage ring is most effective in measuring the momentum of electrically charged particles. The main volume is a gas-filled chamber in which some 12,800 electrically charged wires are arranged in 16 concentric cylinders. A charged particle passing through the chamber ionizes atoms of the gas; liberated electrons drift to a positively charged wire, giving rise to a current that can be measured electronically. The entire chamber is permeated by a magnetic field that causes a charged particle to follow a curved path; from the radius of curvature the momentum of the particle can be deduced. Devices such as scintillation counters and shower counters aid in the identification of charged particles; they also detect photons, but with lower energy resolution than the Crystal Ball.

ly in the laboratory. The probability of an electron's emitting a virtual photon is proportional to $\alpha$, the constant in Coulomb's law, which again has a numerical value of about 1/137.

To calculate the force between an electron and a positron one can begin by determining the force that results from the exchange of a single photon; the result is proportional to $\alpha$. For more precision one can include the possibility of two photons being exchanged; this calculation yields a small correction to the initial result, proportional to $\alpha^2$. The possible exchange of three photons gives rise to a still smaller correction proportional to $\alpha^3$. To attain an exact result would require making an infinite series of such calculations, but because $\alpha$ is quite small the corrections quickly become negligible; considerable accuracy can be achieved by including only the first few terms in the series.

In recent years a theory of the color force between quarks has been developed; it is patterned after QED and is called quantum chromodynamics, or QCD. Just as QED describes the interactions of electrically charged particles, so QCD describes the forces that arise between particles that bear a color charge. Whereas there is just one kind of electric charge, however, there are three kinds of color charge, usually called red, blue and green. (The names, of course, have nothing to do with colors in the everyday sense.) Furthermore, whereas there is just one carrier particle for the electromagnetic force (the photon), there are eight carriers of the color force; they are called gluons. Quarks and antiquarks have color charges, and the forces between them come about from an exchange of gluons.

Perhaps the most important difference between QCD and QED is that the gluons themselves carry a color charge, whereas the photon is electrically neutral. A quark can continually emit and then reabsorb gluons, which effectively spreads its color charge over a region of space. When two quarks or a quark and an antiquark approach each other closely, the extended regions of color charge begin to overlap. Because of the overlapping the force between the particles is not as great as it would be between two point charges separated by the same distance. As the particles approach each other the color force between them, which is inherently quite strong, becomes weaker and takes on the same form as the Coulomb force: its strength is inversely proportional to the square of the distance. This property of QCD, called asymptotic freedom, was discovered by Politzer and David Gross and Frank Wilczek of Princeton University. It is what led Appelquist and Politzer to expect the charmonium system to form nonrelativistic bound states.

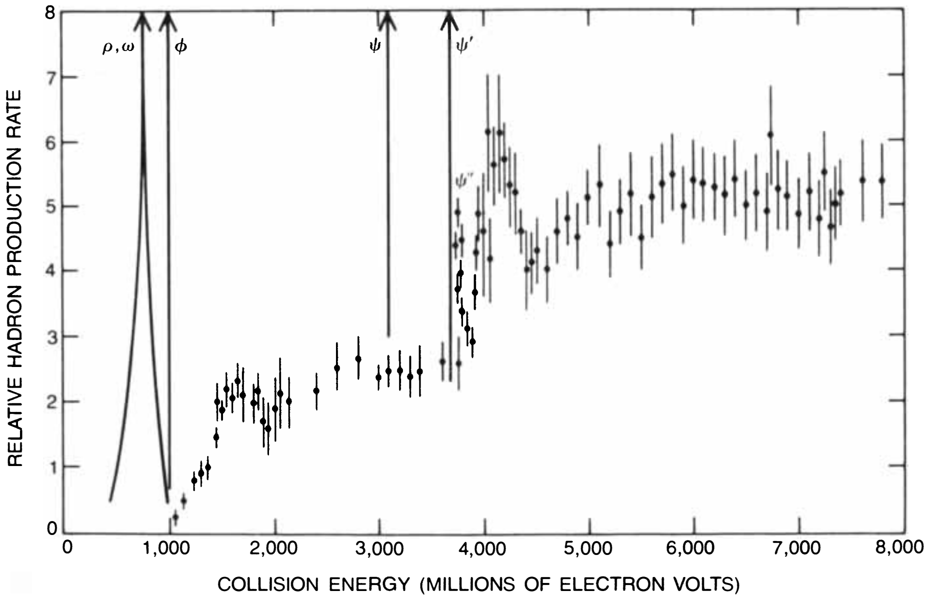
The force between a quark and an antiquark when they are far apart is not well understood, but there is reason to think the force becomes constant, or independent of distance. If it does, the energy needed to separate a quark from an antiquark would increase without limit as the particles moved apart. Such a force law could explain why particles with a color charge (that is, quarks and gluons) have not been seen in isolation.

One reason the long-range color force has not been characterized in detail is that QCD, unlike QED, is not a theory in which calculations are currently practical. The constant $\alpha$ in QED is replaced in QCD by $\alpha_s$, which has a substantially larger value; moreover, the value increases with the distance between the particles. Hence the probability that a single gluon will be emitted is large, and the probability for two or three or more gluons is not small enough for it to be neglected. The impracticality of doing calculations in QCD makes the quarkonium system all the more important for studies of the color force. QCD can serve as a guide in constructing models of the force; the models can then be tested experimentally on the quarkonium system. The results may in turn reflect on the validity and properties of the theory.

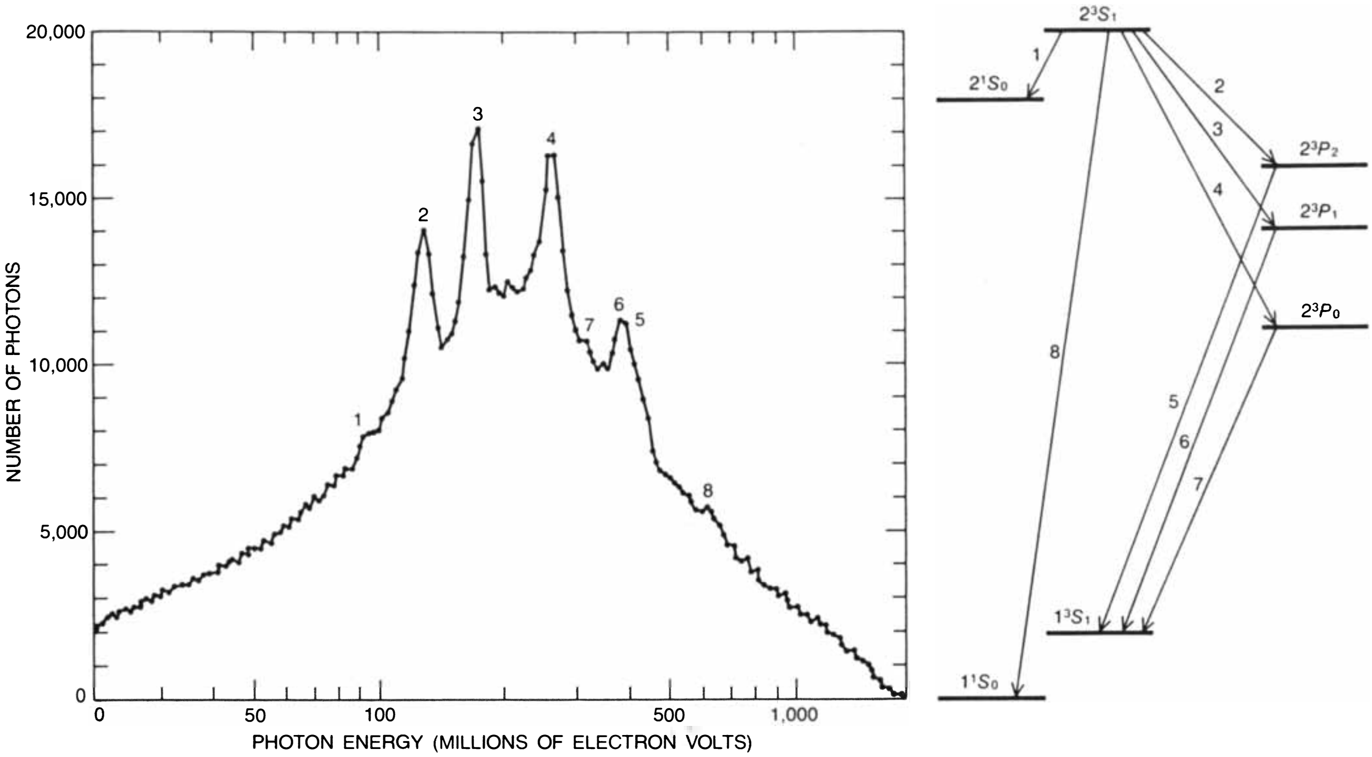
In order to study quarkonium it is necessary to create it. It seems the most effective way to do this is through the annihilation of high-energy electrons and positrons. The process is conceptually identical with the decay of positronium, but because of the higher energy it can proceed through a somewhat different mechanism. The electron and the positron annihilate to produce a single photon, an event forbidden in the decay of positronium because it cannot conserve both energy and momentum. These quantities must still be conserved in the high-energy annihilation, but the uncertainty principle of Werner Heisenberg in effect allows a momentary violation of energy conservation. Before the photon has existed long enough for its presence to be registered it decays into two or more new particles. It is a virtual photon.

In the final state created from the decay of the virtual photon all conservation laws must be obeyed. This condition can be met in a simple way: the virtual photon gives rise to a particle and its corresponding antiparticle. In some cases the particle-antiparticle pair is merely another electron and positron, which can then move apart and be detected. The products of the collision can also be a quark and an antiquark, however, which do not escape unencumbered. Instead additional quark-antiquark pairs materialize and become bound to the original pair, so that what is ultimately observed is in all cases a set of hadrons.

The high-energy annihilations take place in the device known as a storage ring, where beams of electrons and positrons circulate in opposite directions within a toroidal vacuum chamber. Most of the work with charmonium we shall describe was done with the storage ring called SPEAR, which is a facility of SLAC. There are two interaction




**SHARP ENHANCEMENTS in the emission of hadrons (the generic term for the particles made up of quarks) signal the existence of quarkonium systems. The data for energies above 2,400 MeV were recorded at SPEAR; the lower-energy data come from storage rings in France, Italy and the U.S.S.R. The peaks at the left marked ρ (rho), ω (omega) and φ represent quasi-bound state of the lighter quarks. The ψ and ψ′ are the $1^3S_1$ and $2^3S_1$ states of charmonium. The enhancement near 3,770 MeV is the $3^3D_1$ state of charmonium, a quasi-bound state designated ψ″. The increase at about 4,000 MeV is due to the creation of charmed *D* mesons.**




**TRANSITIONS BETWEEN STATES of the charmonium system have been mapped with the Crystal Ball detector. Only the ${}^3S_1$ states can be made directly in electron-positron collisions; the others are formed when one of the ${}^3S_1$ states decays by emitting a photon, which is detected with the Crystal Ball. The peaks in the spectrum of photons are keyed by number to the transitions in the diagram at the right.**

regions at SPEAR, and so two detectors can be operated simultaneously. The two we have employed are called Mark II and the Crystal Ball. They provide largely complementary information.

The Mark II detector was built by workers from SLAC and the Lawrence Berkeley Laboratory of the University of California. It excels in measuring the energy of electrically charged particles, which are registered by a device called a drift chamber. In the chamber fine parallel wires extend through a gas-filled cylindrical volume. An electric potential of a few thousand volts is applied to adjacent wires. When a charged particle passes through the chamber, it ionizes the gas, and the liberated electrons then drift toward the nearest positively charged wire; the resulting current is detected electronically. By timing the current pulses the trajectory of the particle as it passes through 16 concentric cylinders of wires can be determined to within about .2 millimeter. The entire drift chamber is immersed in a strong magnetic field, which causes a charged particle to follow a curved path. From the radius of curvature the momentum of the particle can be deduced.

The Mark II also has devices sensitive to high-energy photons, but these are detected with much greater precision by the Crystal Ball. Since the photon is neutral, it is not deflected by a magnetic field, and the Crystal Ball has no field. Instead the photon is absorbed in a dense crystal of sodium iodide, where it gives rise to a cascade of electron-positron pairs. The cascade causes the crystal to scintillate, or give off many photons at visible wavelengths, which are detected by a photomultiplier tube.

The Crystal Ball was designed and built by workers from Cal Tech, Harvard, Princeton and Stanford. The main component of the detector is an array of 732 crystals, which are arranged symmetrically around the interaction zone. The energy of a photon can be determined to within 2 or 3 percent, and the direction to within one or two degrees.

The discovery of the $\psi$ meson at SLAC was made with the Mark I detector, the predecessor of the Mark II. At the time the detector was being employed to measure the rate of hadron production as a function of the energy of the incident electron and positron. If the quarks generated by the virtual photon initially act as free particles, the rate of hadron production should be proportional to the sum of the squares of the electric charges of all the quarks that can be created at a given energy. Over much of the energy range explored this assumption seemed to be confirmed by the measurements. When the colliding beams were set to an energy of precisely 3,095 MeV, however, the number of hadrons emitted abruptly increased a hundredfold. It was soon apparent that this sharp enhancement signaled the existence of a new kind of quark. The exceptional height and narrowness of the peak indicated something else as well. At 3,095 MeV the assumption that the quark and the antiquark act as free particles fails badly. The charmed quark and antiquark that make up the $\psi$ are not born free; they are tightly bound from the moment of their creation.

The $\psi$ meson was not the first example of a sharp peak in the hadron production rate. Three other short-lived mesons had been observed in a similar way at lower energies: the $\rho$ (rho) at 776 MeV, the $\omega$ (omega) at 782 MeV and the $\phi$ (phi) at 1,020 MeV. One might suppose these mesons are bound states of the up, down and strange quarks, although they would presumably be relativistic bound states. This explanation is only partly correct: the $\rho$, the $\omega$ and the $\phi$ are considered quasi-bound rather than fully bound states. The distinction lies in whether or not the quarks in the hadron can eventually appear in its decay products. Consider the $\phi$ meson, composed of an $s\bar{s}$ pair. It can decay by mutual annihilation, but that is not the usual course of events. Instead the $s$ quark and the $\bar{s}$ antiquark merely move apart and a new $u$ quark and $\bar{u}$ antiquark materialize, forming two new mesons with the quark composition $\bar{s}u$ and $s\bar{u}$. These mesons are designated the $K^+$ and the $K^-$, and their total mass is 32 MeV less than the mass of the $\phi$. The excess 32 MeV is converted into the kinetic energy of the $K$ mesons.

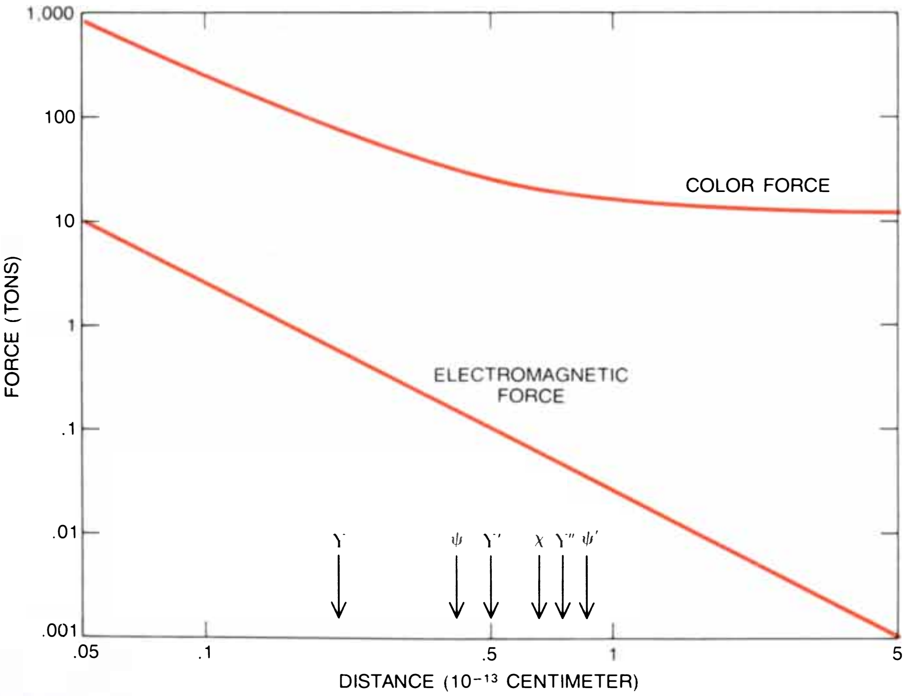
The $\psi$ might decay in a similar way except for an accident of nature. The meson composed of a $c$ quark and a $\bar{u}$ antiquark is the $D^0$; its antiparticle, with the quark constituents $\bar{c}u$, is the $\bar{D}^0$. The mass of the neutral $D$ mesons is 1,864 MeV, and so the mass of the particle-antiparticle pair is 3,728 MeV, which is 633 MeV greater than the mass of the $\psi$. Thus the $\psi$ cannot decay into two $D$ mesons because the decay would violate the conservation of energy. The $c$ and the $\bar{c}$ cannot escape from each other but can decay only by annihilating each other to form gluons, which are then transformed into quarks and finally into ordinary hadrons. This process is comparatively slow and makes the lifetime of the $\psi$ several hundred times longer than it would be otherwise.

In analogy with the lowest-lying state of positronium the $\psi$ was assumed to have zero orbital angular momentum. Because it is created from a virtual photon, however, its total angular momentum must be the same as that of the photon, namely one unit. Hence the $\psi$ is the $1^3S_1$ state of charmonium, the state with $n$ equal to 1, one unit of net spin and no orbital angular momentum.

Immediately after the discovery of the $\psi$ it was suggested that the $2^3S_1$ state, with the same quantities of net spin and orbital angular momentum but with $n$ equal to 2, might also be a bound state. A search for it was undertaken by increasing the energy of the storage ring in increments of 2 MeV. Two weeks later the particle named the $\psi'$ was discovered at an energy of 3,684 MeV; it is now understood to be the $2^3S_1$ state of the $c\bar{c}$ system. If its energy were 44 MeV greater, it could decay into $D$ mesons and would be only quasi-bound.

If the charmonium model is correct, there must be at least five additional bound states of charmonium that might be detected in the decays of the $\psi$ and the $\psi'$. Two of these states are the singlet $S$ states that differ from the $\psi$ and the $\psi'$ only in that they have zero net spin as well as zero orbital angular momentum. Their designations are $1^1S_0$ and $2^1S_0$, and they are called charmed $\eta$ (eta) mesons. To be precise, they have been given the names $\eta_c$ and $\eta_c'$ in analogy with the $\eta$ meson, which is a $^1S_0$ state of the light quarks. The $\eta_c$ and the $\eta_c'$ were expected to have masses about 100 MeV lower than those of the $\psi$ and the $\psi'$ respectively.

The other three expected states were the triplet states with one unit of orbital angular momentum: $2^3P_0$, $2^3P_1$ and $2^3P_2$. They were given the name $\chi$ (chi) mesons by the experimenters who undertook to find them. They were expected to have an average mass about 200 MeV less than the mass of the $\psi'$ and to be separated from one another by between 50 and 150 MeV. A sixth state is also predicted by the charmonium model; it is the singlet state with zero net spin and one unit of orbital angular momentum, namely $2^1P_1$. A low expected rate of conversion between this state and the others, however, makes it unlikely to be detected, and indeed it has not yet been observed.




**FORCE BETWEEN QUARKS, called the color force, seems to obey a law significantly different from the one that describes the electromagnetic force. The strength of the electromagnetic force between two particles varies inversely as the square of the distance between them; such a force law corresponds to a straight, sloping line on this logarithmic graph. At exceedingly short distances the color force also seems to follow an inverse-square law, but beyond about $10^{-13}$ centimeter the force may have a constant value independent of the distance. If the long-range force is constant, data from quarkonium suggest its magnitude is about 16 tons. The force law shown is based on a model devised by John Richardson, who was then at SLAC.**

The singlet $S$ states and the triplet $P$ states of charmonium cannot be created directly in electron-positron annihilations for two reasons. First, with the exception of the $2^3P_1$ configuration, they do not have the same total angular momentum as the virtual photon has, so that their direct production would violate the conservation of angular momentum. Second, for all five states the sum of the absolute magnitudes of the net spin and the orbital angular momentum is an even number (zero for the $S$ states and 2 for the $P$ states). For all the states $C$ is therefore $+1$, in contrast with the virtual photon's value of $-1$. For the $\psi$ and the $\psi'$, on the other hand, $C$ is equal to $-1$, and so they can be formed directly. The $C = +1$ states appear only when the $\psi$ or the $\psi'$ decays into a photon and a lower-energy charmonium configuration, a process that obeys all conservation laws.

The first evidence of the $\chi$ mesons came from an experiment done at DORIS, a storage ring similar to SPEAR at the Deutsches Elektronen-Synchrotron (DESY) in Hamburg. In the years since then several other experiments based on a variety of detection methods have also seen the $\chi$ mesons. They have masses 135, 180 and 270 MeV less than the $\psi'$ mass, in general agreement with the predictions of the charmonium model.

At SPEAR a group from the University of Maryland, the University of Pavia in Italy, Princeton, the University of California at San Diego and Stanford measured the masses of the $\chi$ states and the rate at which the $\psi'$ decays to yield them. These workers made their measurements with an earlier detector based on sodium iodide crystals, detecting not the $\chi$ particles themselves but the photon emitted during a transition from the $\psi'$ to a $\chi$. The energy of such a photon is sharply defined and is nearly equal to the difference in mass between the states. Measurements were also made with the Mark I, searching for events in which a $\chi$ decays into several charged particles. For example, a $\chi$ occasionally breaks down into two positive and two negative pions; from the energy and the angular distribution of the pions the mass of the $\chi$ state can be deduced. More recently higher-resolution measurements have been made by the Crystal Ball and the Mark II groups.

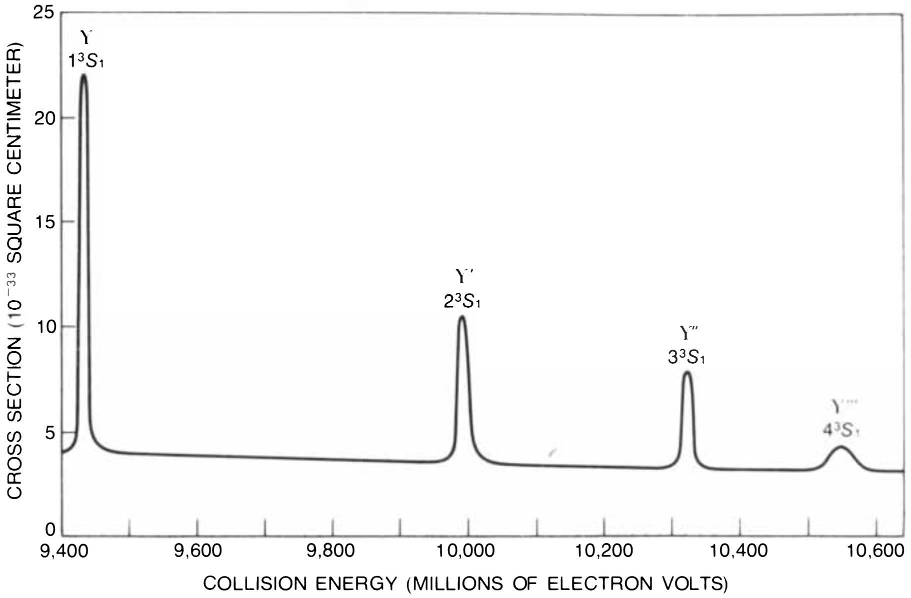
The angular-momentum quantum numbers of the $\chi$ mesons can be determined from the nature of the hadrons emitted in the decay of the particles or from the angular distribution of the photons. The early experiments showed that none of the three particles could be the $^1S_0$ states $\eta_c$ or $\eta_c'$. Furthermore, if it was assumed that one meson had a total angular momentum of 2, another an angular momentum of 1 and the third an angular momentum of zero, there was only one way to match the angular-momentum states with the observed masses. The most massive state had to be the one with an angular momentum of 2, the next state the one with an angular momentum of 1 and the least massive the one with an angular momentum of zero. This is the sequence of the $2^3P$ states of positronium. A more detailed study done with the Crystal Ball later confirmed the angular-momentum assignments.

The story of the $\eta_c$ and the $\eta_c'$ mesons is not as straightforward. Experimenters working at DORIS reported evidence for an $\eta_c$ candidate in the early years of the exploration of the charmonium spectrum. They measured an anomalously large number of three-photon decays of the $\psi$ meson in which two of the photons had an aggregate mass of 2,830 MeV; this mass is 265 MeV less than the mass of the $\psi$ itself. They interpreted these events as the emission of a comparatively low-energy photon by the $\psi$, which was thereby transformed to the $\eta_c$ state. The latter meson then gave up its mass of 2,830 MeV in forming the two additional observed photons. The signal seemed impressive, but there was a large background of extraneous events, and no other experiment was able to provide a convincing confirmation of the DORIS findings.

There was also a hint of an $\eta_c'$ candidate based on four events recorded with the Mark I detector and an equal number from DORIS experiments. In these events the $\psi'$ emitted a photon to become the $\eta_c'$, which then emitted a second photon in the course of decaying to the $\psi$ level. The mass of the $\eta_c'$ candidate was 3,454 MeV, or 230 MeV less than the mass of the $\psi'$.

Neither candidate could be fitted easily into the charmonium model. The problem was that the masses were too low. In 1978 the expected mass of the $\eta_c$ was calculated by Michael Shifman, Arkady Vainshtein, Michael Voloshin and Valentin Zakharov of the Institute of Theoretical and Experimental Physics in Moscow. From theoretical principles based on QCD they showed that the $\eta_c$ must have a mass about 100 MeV less than the mass of the $\psi$, within an error of plus or minus 20 or 30 MeV. On this basis they concluded unequivocally that either the $\eta_c$ candidate (with a mass 265 MeV less than that of the $\psi$) was not the true $\eta_c$ meson or the simple charmonium model was not valid.

At this point the status of the charmonium hypothesis was uncertain. The three $\chi$ mesons had been found and their properties were in good agreement with the predictions of the model, but the candidate $\eta_c$ states were at variance with the model in their mass and in certain other properties. It was not clear that the model was understood well enough for it to be trusted where it seemed to contradict the experimental evidence.




**BOTTOMONIUM SYSTEMS, made up of a bottom quark and a bottom antiquark, are heavier than the corresponding states of charmonium and may reveal more clearly the nature of the interquark forces. Whereas two of the $^3S_1$ states of charmonium (the $\psi$ and the $\psi'$) are fully bound systems, there are three analogous bound states of bottomonium. They are designated Υ (upsilon), Υ′ and Υ″. A fourth bottomonium system, the Υ‴, is quasi-bound. The graph is based on data from the storage ring called CESR at Cornell University. The height of the peaks is given in terms of a cross section, which measures the probability of creating a particle.**

In 1978 the experimenters working with the Crystal Ball began collecting data, and by 1979 they had repeated the measurement of the decay of the $\psi$ meson into three photons, the only decay mode in which the $\eta_c$ candidate had been observed. The Crystal Ball was able to detect many more of these decays than had been recorded by the apparatus at DORIS and to resolve the energy of the photons with much greater precision. The candidate was simply not there. The earlier experiment was almost certainly in error, although the reason is still not clear. It is possible there was a statistical fluctuation in the number of background events or a misunderstanding of the detector's response to them.

The $\eta_c'$ candidate died a similar death. Experiments at DORIS and with both the Mark II and the Crystal Ball detectors at SPEAR searched the energy region where the candidate had been reported. Although all the new experiments were more sensitive than the earlier ones, they found no evidence of the $\eta_c'$. The particle with a mass of 3,454 MeV did not exist.

The state of uncertainty finally ended late in 1979 with the discovery by the Crystal Ball collaboration of the particle that appears to be the true $\eta_c$. The new $\eta_c$ candidate has a mass of 2,980 MeV, or about 115 MeV less than the mass of the $\psi$, in accordance with the predictions of the charmonium model. The 2,980-MeV state was first observed as a small but significant peak in the spectrum of photons emitted in the decay of the $\psi'$. Later a peak with the same energy was noted in the photon spectrum from the decay of the $\psi$. The existence of the $\eta_c$ particle was soon confirmed by the Mark II detector, which was able to record several of its decays into hadrons.

In August, 1981, the last of the five charmonium states with $C$ equal to $+1$ was found close to where it had been predicted to be. The Crystal Ball collaboration reported the observation of another small peak in the spectrum of photons emitted by decaying $\psi'$ mesons. The peak is at an energy of 3,592 MeV, some 92 MeV less than the mass of the $\psi'$; this is well within the range of masses predicted for the $\eta_c'$. With this discovery the $\eta_c$ and $\eta_c'$ puzzles seem to be solved, and the way is clear to apply the quarkonium model to the understanding of the color force.

Probably the best quarkonium system now available to experimenters is the $b\bar{b}$ system, or bottomonium. It should be superior to charmonium because it is appreciably heavier and is therefore a better approximation to a true nonrelativistic system.

When the Υ was discovered at Fermilab in 1977, it was interpreted as the $1^3S_1$ state of bottomonium, directly analogous to the $\psi$, which is the same state of charmonium. The discovery was accompanied by reports of the probable observation of the $2^3S_1$ state, named the Υ′, and the tentative observation of the $3^3S_1$ state, the Υ″ state. The existence of three bound states of bottomonium (compared with two bound states of charmonium) was an important prediction of the model put forward by Eichten and Gottfried.

Because of the greater binding energy of bottomonium the array of associated bound states should be even richer than it is in charmonium. For each of the Υ particles there should be a singlet $S$ state of somewhat lower mass: they are the $1^1S_0$, $2^1S_0$ and $3^1S_0$ states, which have been labeled $\eta_b$, $\eta_b'$ and $\eta_b''$. Similarly, three $\chi_b$ particles should exist, representing the triplet of $2^3P$ states, that is, $2^3P_0$, $2^3P_1$ and $2^3P_2$. In the bottomonium system, however, the $3P$ states should also be bound, whereas they are not in charmonium. Furthermore, the lowest-energy $D$ states, with two units of orbital angular momentum, also ought to exist as bound systems. They are the $3^3D_1$, $3^3D_2$ and $3^3D_3$ configurations.

As yet none of the bottomonium systems other than the Υ particles has been observed. The experiments are more difficult than those with charmonium, largely because the $b\bar{b}$ systems are formed at a much lower rate than the $c\bar{c}$ systems. The search is now under way in earnest at several laboratories. An electron-positron storage ring called CESR at Cornell has a nearly optimum energy range for creating bottomonium systems, and new storage rings capable of still higher energy have been built at both SLAC and DESY. In this connection the Crystal Ball has been removed from SPEAR and is being installed at an interaction region of the higher-energy ring at DESY. The coming generation of experiments should be able to observe the $\chi_b$ states and perhaps even the $\eta_b$.

What has been learned from quarkonium about the nature of the interquark force? Because the various states of quarkonium differ in the average separation between the quark and the antiquark, the energies of these systems convey information about the strength of the force over a range of distances. The experiments done so far constitute measurements of the color force at distances ranging from $10^{-13}$ centimeter, which is roughly the size of an ordinary hadron, down to about $2 \times 10^{-14}$ centimeter, a distance a fifth as large. From the measurements it is possible to construct models of how the force varies as a function of interquark distance.

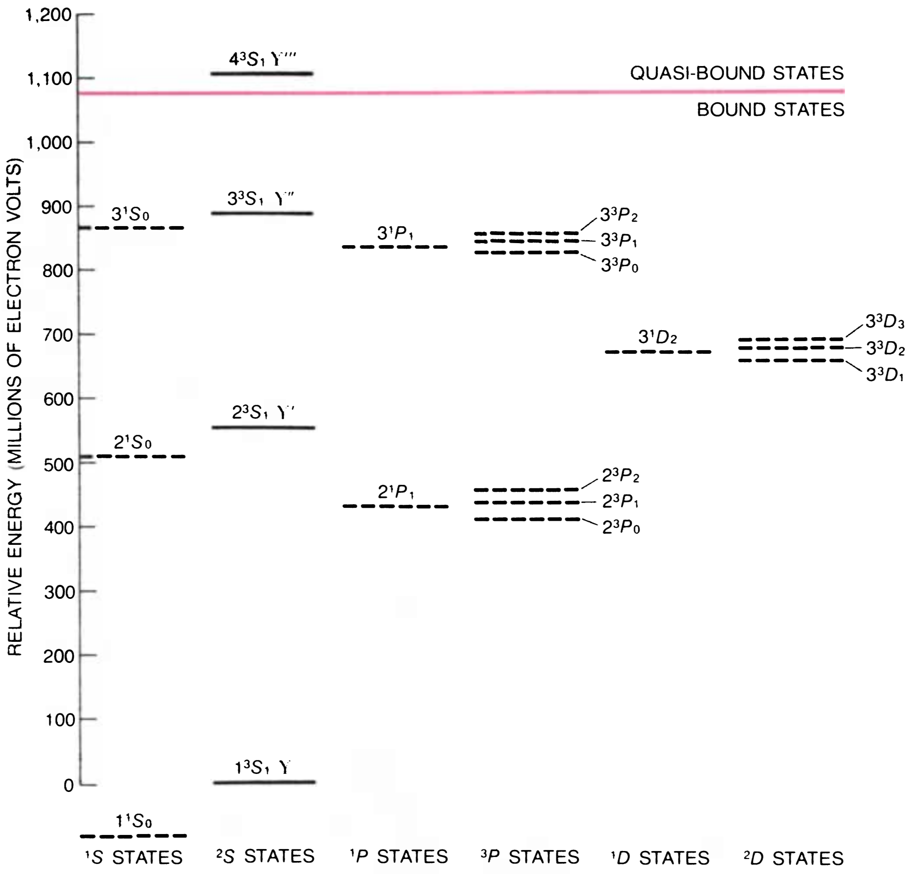
QCD suggests that the color force may vary inversely as the square of the distance when the quarks are close together but may assume a constant strength for quarks that are far apart. A plausible way to approximate such a force is to assume that it is merely the sum of an inverse-square force and a constant force. This model force law can be expressed mathematically by the equation $F = a/r^2 + b$, where $F$ is the force, $r$ is the interquark distance and $a$ and $b$ are constants to be determined by experiment. At very small distances the $a/r^2$ term is large and makes the dominant contribution to the total force. At large distances, on the other hand, $a/r^2$ is negligible and the force is essentially equal to the value of $b$.

The model force law can be tested by attempting to fit its predictions to the data for quarkonium. If the resulting curve has the right form, one can then ask what values of $a$ and $b$ give the best fit. It turns out that the form of the curve is consistent with the data, and the value of $b$ is about 16 tons. In other words, two quarks attract each other with a force of at least 16 tons regardless of how far apart they are. With this in mind it becomes easier to understand why a quark has never been extracted from a hadron. It should be emphasized that we have not proved by these arguments that a constant force between quarks exists; a proof is hardly possible when the data extend only to distances of about $10^{-13}$ centimeter. What we have shown is that the existing data are consistent with a constant long-range force and that if it exists, its magnitude is about 16 tons.

The value of $a$ in the equation for the force law is more difficult to specify. The reason is that $a$ is not actually a constant but depends to some extent on $r$. The value of $a$ is related to the probability that a quark will emit a gluon, the quantity we have designated $\alpha_s$. Unlike $\alpha$ (the probability that an electron will emit a photon), $\alpha_s$ varies with distance, becoming smaller as the quark and the antiquark move closer together. For the range of distances explored in the quarkonium system $\alpha_s$ seems to have a value of approximately 1/5, compared with 1/137 for $\alpha$.

Much of the interest in quarkonium stems from the close parallels between quarkonium and positronium and between QCD and QED. There is one phenomenon in systems governed by the color force, however, that has no equivalent in electromagnetic systems. It is a particle composed exclusively of gluons, bound together by their color charges. The analogous electromagnetic entity would be a bound state of two photons, but such a system cannot exist because the photon has no charge.

The hypothetical gluon bound states are generally called glueballs, but we shall call them gluonium systems to make plain their similarity to quarkonium. If they exist at all, the decay of heavy quarkonium is one of the processes most likely to create them. Just as the $1^3S_1$ state of positronium decays into three photons, the $1^3S_1$ state of charmonium (that is, the ψ particle) usually gives rise to three gluons. The gluons in turn create quark-antiquark pairs, which combine to form hadrons. About 10 percent of the time, however, charmonium decays into a photon and two




**ENERGY LEVELS OF BOTTOMONIUM are expected to include more bound states than those of charmonium. In addition to the $S$ and $P$ bound states, certain bound $D$ states, with two units of orbital angular momentum, may be detectable. Only the $^3S_1$ states have been observed.**

gluons; it is in these events that gluonium might be most likely to form. The two gluons emitted in the decay have opposite values of color charge and could form a color-neutral bound system directly.

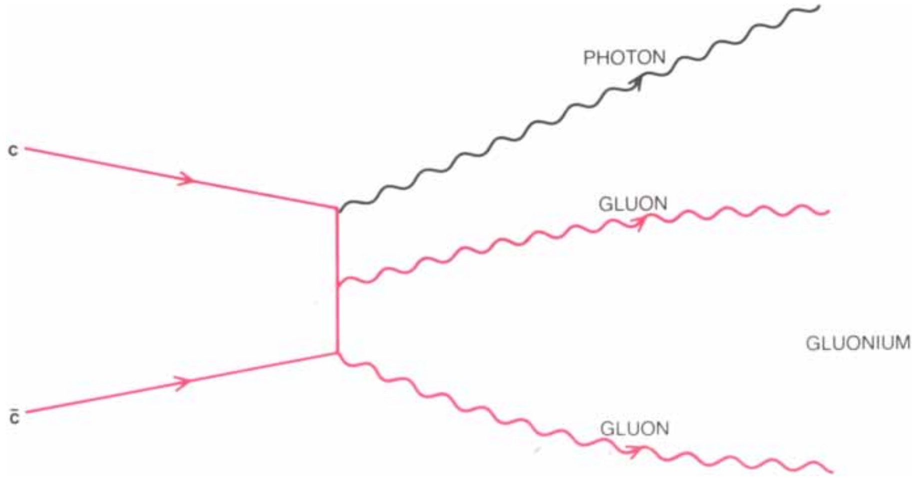
It is not at all clear how to recognize gluonium if it is created. It is even possible that a particle might vacillate continually between quark-antiquark states and gluon states, in which case the concept of gluonium would not be very meaningful. If gluonium does have a fixed identity, one of its distinguishing characteristics might be the spectrum of angular-momentum states. Because the gluons have a spin of 1 rather than 1/2 the gluonium state of lowest energy is expected to be an $S$ state with a total angular momentum of either zero or 2. This is in contrast to quarkonium, where the $S$ states have a total angular momentum of zero or 1. Another clue is that gluonium should be produced more abundantly in processes that are rich in gluons (such as the decay of quarkonium into a photon and two gluons) than in other interactions of hadrons. Gluonium might also be made conspicuous by its failure to fit into any of the established families of mesons.

Two years ago Daniel Scharre of SLAC discovered evidence in data from the Mark II detector of a curious mode of decay of the $\psi$ particle. In roughly three decays in every 1,000 the $\psi$ gave rise to a photon and a meson with a mass of about 1,420 MeV. Scharre identified the meson as the $E$, which is one of the $2^3P_1$ states of the light quarks and therefore has a total angular momentum of one unit. The finding was surprising because the $E$ meson is quite rare; although three per 1,000 is a small proportion of all $\psi$ decays, it is considered a remarkably high rate of production for the $E$ meson.

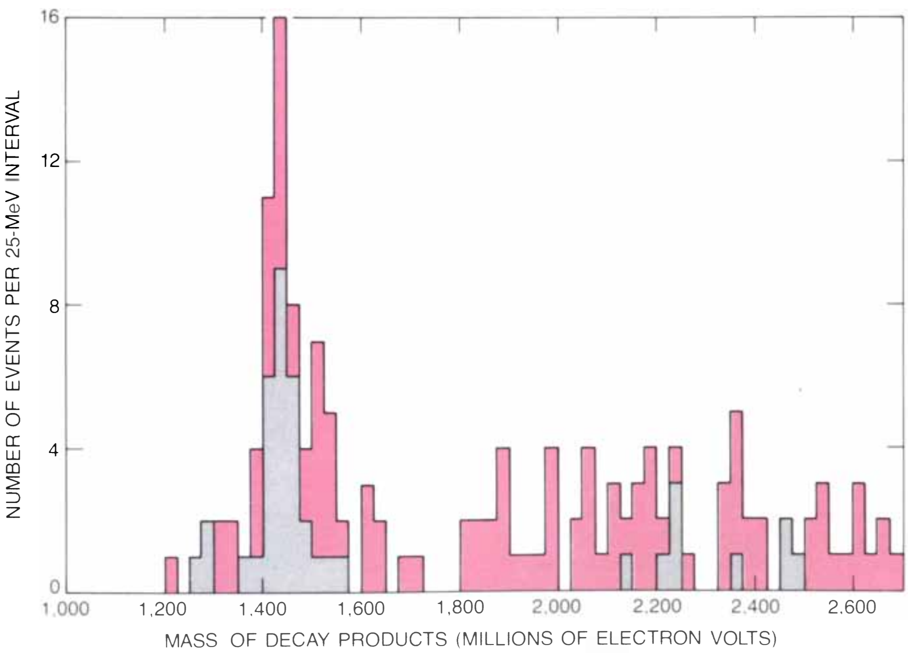
Michael S. Chanowitz of the Lawrence Berkeley Laboratory and others suggested that the solution might lie in a case of mistaken identity. The particle observed in the decay of the $\psi$ had been identified on the basis of its mass and its most prominent mode of decay, which were the same as those of the $E$ meson. Chanowitz pointed out that the new particle differed from the $E$ in certain details of its production and decay. He suggested that what was being seen was not the $E$ but a new particle, one with a total angular momentum of zero.

Chanowitz' arguments were proved to be correct last August when the Crystal Ball collaboration reported a measurement of the angular momentum of the 1,420-MeV particle; it was found to be zero. Since the particle could not be the $E$ meson, it was given a new name: $\iota$ (iota). At the same time the Crystal Ball collaboration announced the discovery of yet another new particle among the decay products of the $\psi$. Like the $\iota$, the second new particle appears in conjunction with a photon, but its mass is greater: 1,640 MeV. The second particle has been given the name $\theta$ (theta), and it has been found to have two units of total angular momentum.

What are the $\iota$ and the $\theta$ particles? The question is not easily answered. If they are quark-antiquark states, the $\iota$ must belong to the $^1S_0$ family of mesons and the $\theta$ to the $^3P_2$ family. They cannot be any of the $n = 1$ states of these families because those states are already fully occupied with known particles. They could be $n = 2$ states. They could be exotic "double" mesons, composed of two quarks and two antiquarks. Of course, they could also be gluonium states. Each of the possibilities entails a distinctive spectrum of accompanying new particles. If experiments show that the spectrum is the one characteristic of gluonium, the $\iota$ and the $\theta$ mesons will be recognized as the first examples of a new form of matter: particles embodying the interquark force in the absence of quarks.




**GLUONIUM is a hypothetical bound system made up entirely of gluons, the particles that ordinarily act as intermediaries transmitting the color force between quarks. Because gluons carry color charges the gluons are themselves subject to the color force and may form bound states. If gluonium exists, it is likely to be made in events that yield gluons in abundance. One plausible mechanism is the decay of a $\psi$ meson into a photon and two gluons; under these circumstances the gluons would necessarily have opposite color charges and could become bound.**




**POSSIBLE EVIDENCE OF GLUONIUM may lie in the $\iota$ (iota) particle, discovered last year in work at SPEAR. The $\iota$, which has a mass of 1,420 MeV and zero angular momentum, does not readily fit into any of the well-established families of quark-antiquark mesons. It could be a bound state of two gluons, although other explanations are possible. Here decays of the $\psi$ meson recorded with the Mark II detector have been selected to include only those events having a certain set of three particles among the decay products. The events were assigned to 25-MeV intervals according to the combined mass of the particles. The peak from 1,400 to 1,450 MeV is caused by the $\iota$. Gray bars represent a subset of the data selected by more restrictive criteria.**

# NMR Imaging in Medicine

*Nuclear magnetic resonance, or NMR, can reveal the distribution of atoms in a sample of material. It can do the same in the body, generating images of internal structure without the use of X rays*

by Ian L. Pykett

The medical need to see inside the human body from the outside has been met for many decades by recording the differential absorption of X rays. A major deficiency of the standard method of radiography is its inability to discriminate among overlapping structures. This deficiency has been remedied in recent years by the development of X-ray computerized tomography, or CT scanning, a technique in which X-ray data recorded from many different directions are reconstructed mathematically to yield cross-sectional views of selected regions of any part of the body. Although CT scanning has proved to be an extremely useful diagnostic tool, the information its images provide is basically anatomical; they tell little about the functional or physiological state of the internal organs. Moreover, some pathological lesions have X-ray absorption properties so similar to those of the surrounding tissues that the lesions can go undetected in a CT scan unless they are large enough to change the size or shape of the organ. Beyond that X rays, even in small doses, carry a finite risk of doing physiological harm.

A new technique for obtaining cross-sectional pictures through the human body without exposing the patient to ionizing radiation is on the threshold of clinical application: nuclear-magnetic-resonance imaging. NMR imaging not only yields anatomical information comparable in many ways to the information supplied by a CT scan but also promises to discriminate more sensitively between healthy and diseased tissue. The latter prospect is founded on the well-established ability of NMR spectroscopy to elucidate the intricate conformation of organic molecules and to provide insight into dynamic chemical processes. For several years biochemists have exploited NMR techniques to monitor metabolic reactions in experimental animals and human beings. It is the recent development of methods for presenting NMR information in pictorial form that is now providing clinicians with a powerful new diagnostic tool.

The experimental foundations of NMR spectroscopy were laid by Felix Bloch of Stanford University and Edward M. Purcell of Harvard University more than three decades ago, work for which they were awarded a Nobel prize in 1952. It had been known since the 1920's that many atomic nuclei have an angular momentum arising from their inherent property of rotation, or spin. Since nuclei are electrically charged, the spin corresponds to a current flowing about the spin axis, which in turn generates a small magnetic field. Each nucleus of nonzero spin therefore has a magnetic moment, or dipole, associated with it. Only nuclei with an odd number of nucleons (protons or neutrons) exhibit a net spin and therefore lend themselves to NMR spectroscopy.

In general the magnetic dipoles of the nuclei with spin will be pointing in random directions. When they are placed in a magnetic field, however, they will orient themselves with the field's lines of induction, or lines of force. For nuclei of the spin designated 1/2, such as protons (hydrogen nuclei, $^{1}H$), the only allowed orientations of the dipoles are parallel to the field or antiparallel to it (in the opposite direction). The two orientations have slightly different energies, described as a Zeeman splitting of the energy levels. In the case of protons the difference between the number of protons with spin "up" (parallel) and spin "down" (antiparallel) is very small: only about one part in $10^8$, with a slight excess in the lower energy state (spin up).

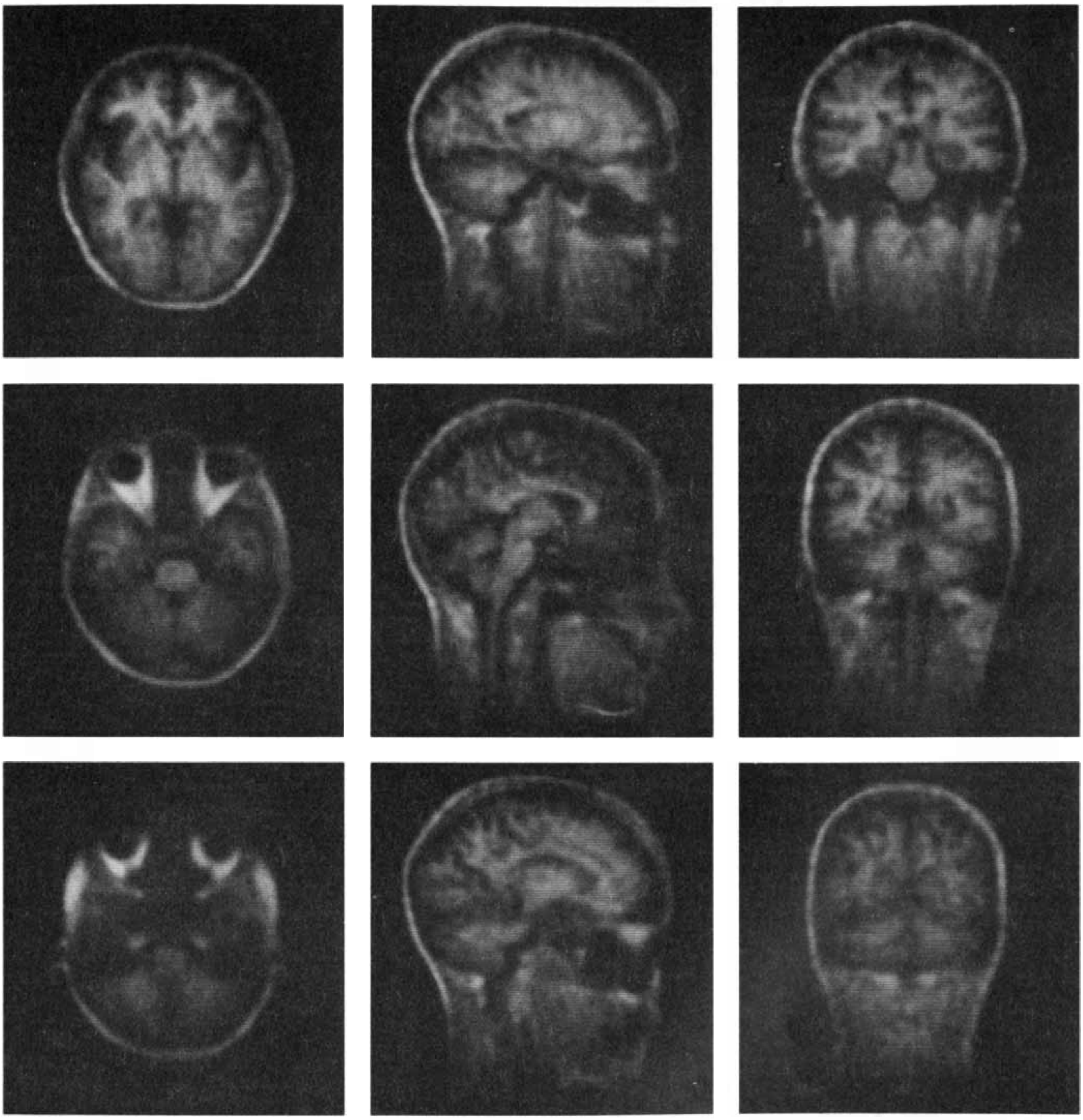
The magnetic behavior of the entire population of nuclei can be predicted by defining a macroscopic, or bulk, magnetization vector, $M$, that represents the net effect of all the magnetic moments of the nuclei of a given nuclear species in the sample of material being examined. In the absence of an external magnetic field the bulk magnetization is of course zero. When a magnetic field is imposed on the sample, however, the nuclear dipoles become oriented to yield a finite equilibrium bulk magnetization that will point in a direction parallel to the applied magnetic field. This direction conventionally defines the $z$ axis.

Spinning nuclei behave rather like tiny tops or gyroscopes. If the axis of a spinning gyroscope is tipped away from the vertical, the gyroscope will rotate about its former axis in a motion describing the wall of a cone. It is the motion called precession. Similarly, if the bulk magnetization $M$, corresponding to an assembly of spinning nuclei in a magnetic field, is tipped away from the $z$ direction, $M$ will precess about the $z$ axis. Such a tipping can be achieved by applying a much smaller magnetic field that is rotating in the $x$-$y$ plane, at right angles to the static (nonrotating) field. In practice the rotating magnetic field is applied by surrounding the sample with a coil connected to a source of radio-frequency power. In order to tip the macroscopic spin vector away from the $z$ axis the frequency of the applied electromagnetic radiation must match the natural precessional frequency of the nuclei of the sample, hence the term nuclear magnetic resonance.

A simple mathematical relation links the resonance frequency, often called the Larmor frequency, to the value of the externally applied static magnetic field. The frequency is equal to the strength of the field multiplied by the "gyromagnetic ratio," which is unique for each nuclear species of nonzero spin. For hydrogen nuclei (protons) in a magnetic field of one tesla (10,000 gauss) the resonance frequency is 42.57 megahertz (MHz), or 42.57 million cycles per second. For nuclei of the isotope phosphorus 31 ($^{31}P$) in the same field the resonance frequency is 17.24 MHz; for nuclei of sodium 23 ($^{23}Na$) it is 11.26 MHz. These frequencies are in the radio-frequency band of the electromagnetic spectrum. Such frequencies, far below those of X rays or even visible light, are powerless to disrupt the molecules of living systems.

It is evident, therefore, that by the proper choice of frequency one can "tune in" to specific nuclear species and observe their response in isolation. All medical NMR images produced so far, however, have been obtained with the resonances only of hydrogen nuclei. Other nuclei not only have a lower intrinsic NMR sensitivity but also are found in much lower concentrations in biological material.

From the quantum-mechanical viewpoint the tipping of the bulk magnetization vector of an assembly of nuclei away from its equilibrium position is equivalent to a transition from a lower energy state to a higher one. The transition is effected only when the energy of the quanta carried by the radio-frequency field exactly equals the difference in magnetic energy between the two energy states. The fact that medical NMR images can be most easily generated from the resonance of hydrogen nuclei is fortunate, because the human body is 75 percent water, each molecule of

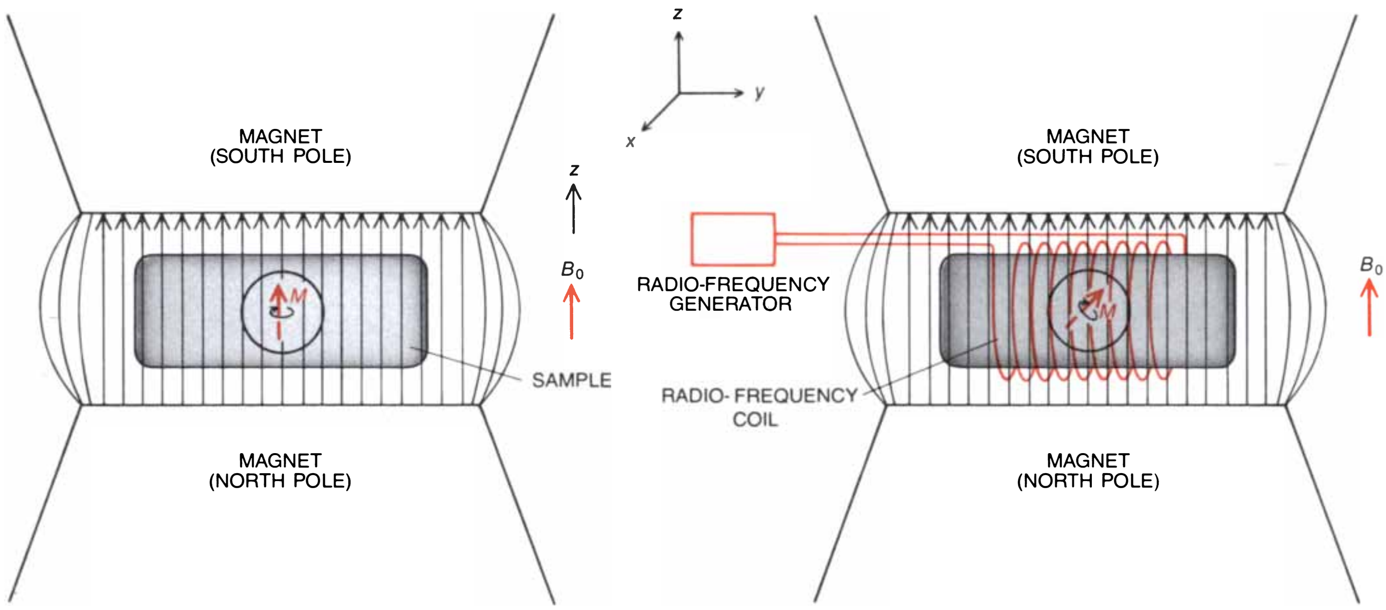


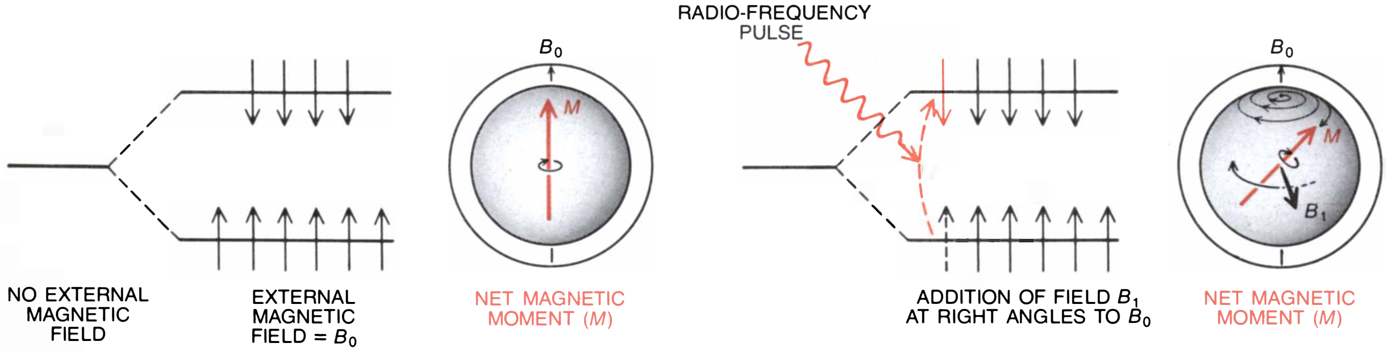
**NMR IMAGES are cross-sectional pictures of thin slices through the body obtained by using radio waves to interrogate susceptible atomic nuclei that have been precisely oriented in a magnetic field. The susceptible nuclei are those that have an odd number of nucleons (protons or neutrons) and therefore exhibit a net spin. Hydrogen nuclei (protons) are the most ubiquitous in living matter. After radio excitation the nuclei reveal their location by emitting a signal of precise frequency for a brief period. With computer techniques pictorial images can be reconstructed from the emitted signals. Here are nine images of a human head, reconstructed from a single three-dimensional data collection, showing primarily the distribution of proton-containing water and lipid molecules in blood and tissue. The three transverse images at the left show sections at three levels: through the center of the brain (*top*), at eye level (*middle*) and just below eye level (*bottom*). The three images in the middle show a section at the midline (*center*) and parallel sections about three centimeters to the left and right. The three images at the right show the head from the front at the brain's widest point (*top*) and at two regions farther to the rear. The images, which were made in the NMR research laboratory of the Massachusetts General Hospital, were generated with NMR apparatus built by the Technicare Corporation of Solon, Ohio. The apparatus had a magnet operating at .15 tesla (1,500 gauss).**

which has two hydrogen nuclei. Moreover, the distribution of water, together with that of various other small, hydrogen-rich molecules (for example lipids), is known to be altered by many disease states.

The displacement angle between the nuclear magnetization vector *M* and the direction of the static magnetic field continues to increase for as long as the rotating field is applied to the sample, and the rate of increase depends on the power of the field. A pulse long and strong enough to tip *M* from its initial position until it is just rotating in the *x-y* plane is termed a 90-degree pulse. But how does one know that there has been any tipping? Immediately after the application of a 90-degree pulse the magnetization vector continues to rotate freely in the *x-y* plane, and in so doing it generates a small electromotive force that can be detected either by the same coil that transmitted the pulse or by a separate receiver coil. The emitted signal is called the free induction signal or free induction decay. In quantum-mechanical terms the signal is generated as the nuclei drop back from the excited energy level to the ground state, the lowest energy level.

After the excitation pulse ends, the magnetization vector of the nuclei eventually returns to its original position directed along the *z* axis. The return to equilibrium is characterized by two principal "relaxation" times, $T_1$ and $T_2$. To discuss $T_2$ first, it is called the spin-spin relaxation time or the transverse relaxation time. Spin-spin relaxation phenomena influence the natural lifetime of the free induction signal, during which the various components of magnetization in the *x-y* plane remain more or less in phase. When the excitation pulse ends, each nucleus continues to "feel" not only the external static field but also local fields associated with the magnetic properties of neighboring nuclei. The nuclei will therefore acquire a range of slightly different precessional frequencies, causing the free induction signal to go out of phase. In a liquid, where the atoms and their nuclei are continuously in motion, the internuclear magnetic fields responsible for spin-spin relaxation tend to average out. As a result the signal decays much more slowly than it would in a solid, where the nuclei




**IN THE PRESENCE OF A MAGNETIC FIELD, $B_0$, nuclei with nonzero spin orient themselves with the magnetic field lines, which point in the *z* direction (*left*). A net magnetic moment *M*, aligned parallel to the field, is therefore generated within the sample. In quantum-mechanical terms the alignment corresponds to the creation of multiple magnetic energy levels. For nuclei with the spin designated 1/2, which include protons, two levels are established, with a small excess of nuclei in the lower energy state. The vector *M* can be made to precess about the magnetic field direction as if it were a falling top. This is done (*right*) by applying a second magnetic field, $B_1$, which has a rotating component in the plane *x-y* at right angles to the static (nonrotating) field. Only when the frequency of the rotating field exactly matches the Larmor frequency (the natural resonance frequency) of the nuclei is *M* tipped toward the *x-y* plane. $B_1$ is generated by placing around the sample a coil that applies radio-frequency energy to the spin system. In quantum-mechanical terms quanta of radio energy that exactly match the gap between energy levels will cause some nuclei to "flip" from the lower energy state to the higher one.**

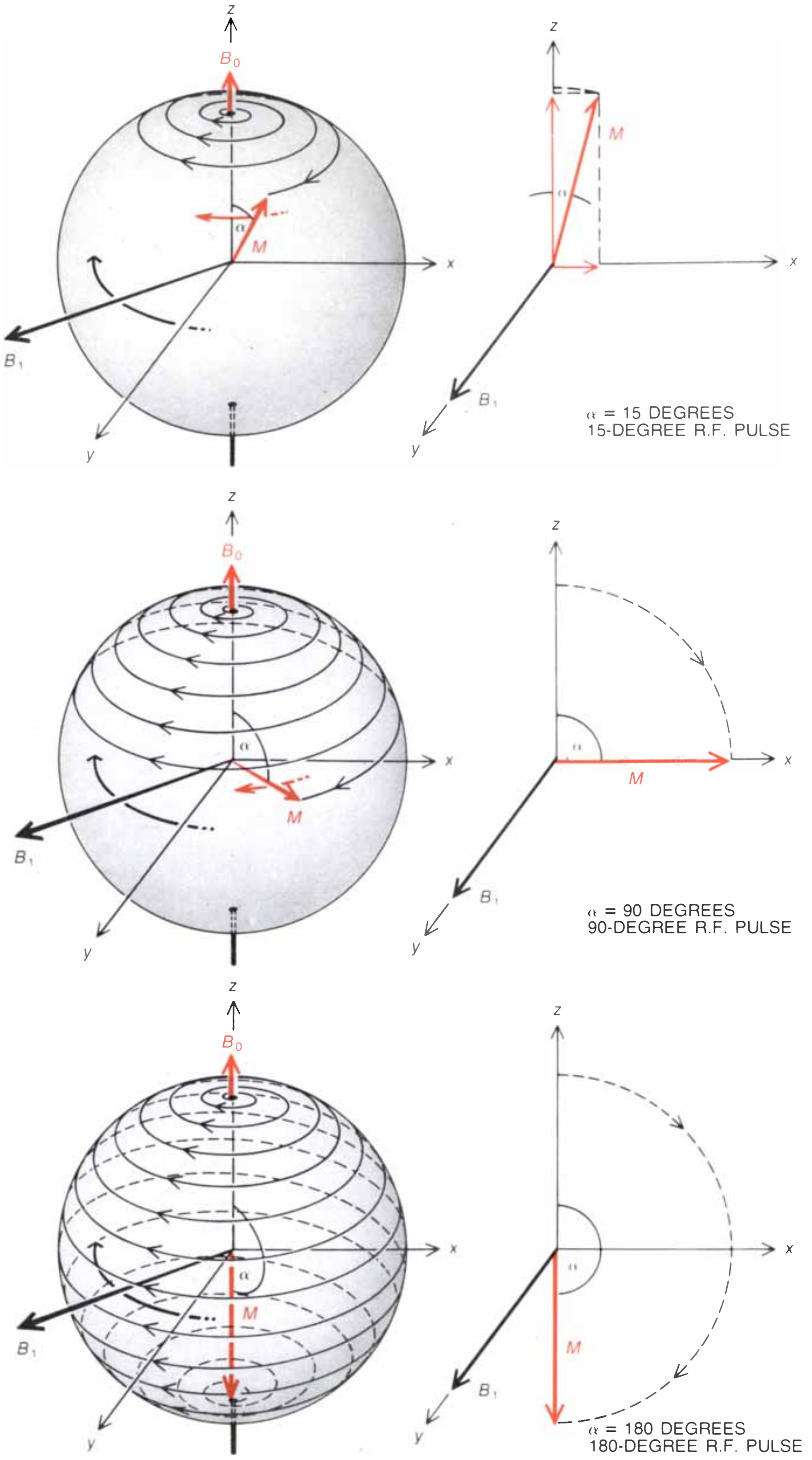
remain essentially fixed in space. In a pure liquid $T_2$ can be as long as several seconds. In a solid it is usually only a few microseconds; indeed, the signal decays so fast that it is generally undetectable.

If the static magnetic field were perfectly uniform, $T_2$ could be determined simply by measuring the rate of decay of the free induction signal. The fields generated by real magnets, however, are always less than perfect. Even though the imperfections in the best magnets used for NMR spectroscopy are extremely small, they cause the free induction signal to decay faster than it would if the magnetic field were perfectly homogeneous. The time constant defining the actual rate of signal decay in an imperfect field is designated $T_2^*$ ("$T$ two-star") to distinguish it from the true relaxation time $T_2$.

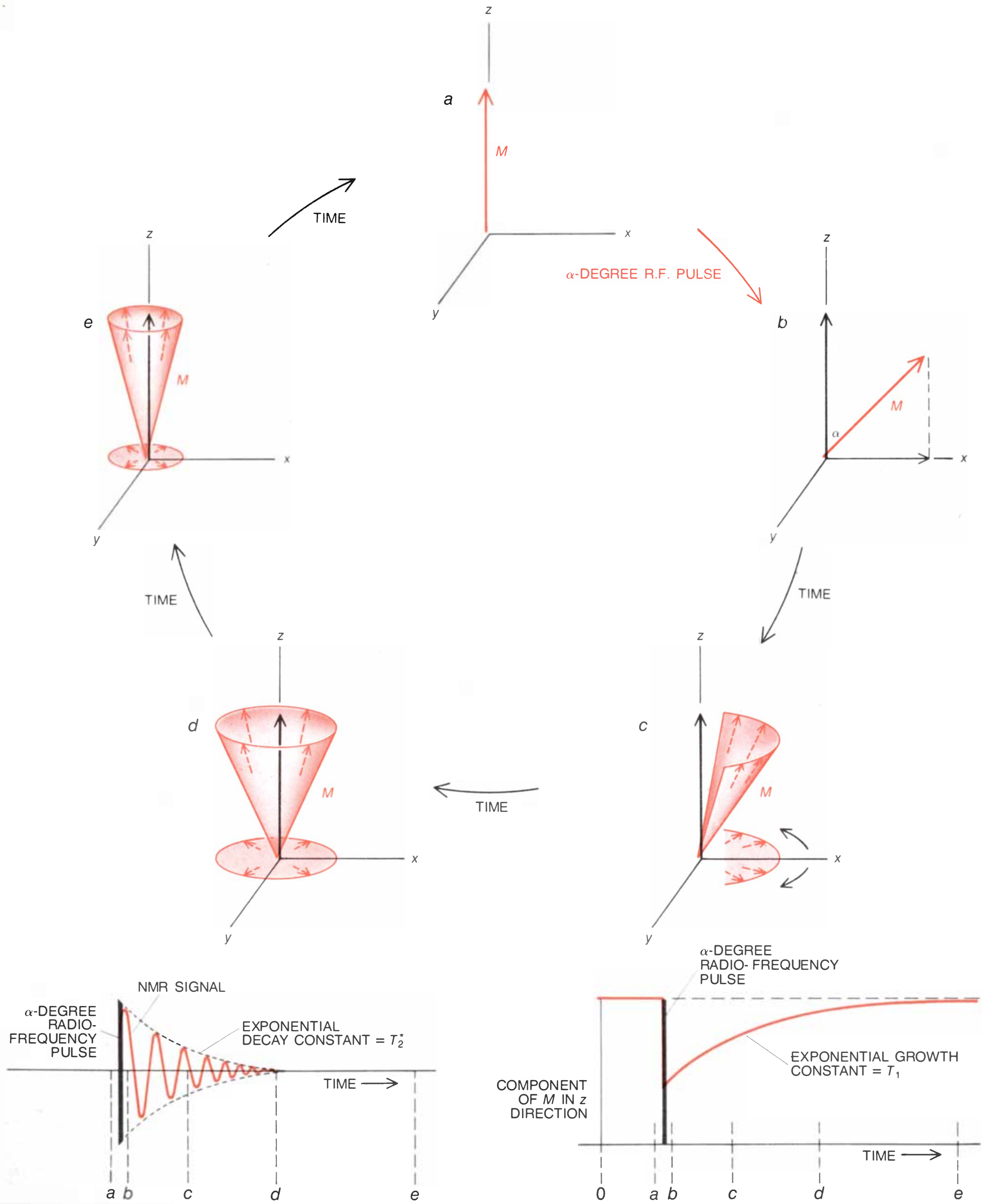
It is nonetheless possible to determine the intrinsic $T_2$ value for a sample of material even when it is in an imperfect field, because the nonuniformities in the magnetic field are constant and can in effect be identified and canceled. For this purpose one can recall the signal in the form of a "spin echo," or a series of echoes, by applying a special pattern of radio-frequency pulses called the Carr-Purcell pulse sequence. In such a sequence the initial free induction decay signal and each of the individual spin echoes decay with a time constant $T_2^*$, but the peak heights of successive spin echoes decay with a time constant equal to the intrinsic $T_2$ value of the sample.

The other relaxation time, $T_1$, is known as the spin-lattice relaxation time or longitudinal relaxation time. $T_1$ is characteristic of the time required for the spin system to return to thermal equilibrium with its surroundings (the "lattice") after the excitation pulse ends. In quantum-mechanical terms the fluctuating magnetic fields of the nuclei making up the lattice must have appropriate frequency components to stimulate transitions from the upper magnetic energy level back down to the ground state. In solids or in samples of material at low temperatures, where the atoms and molecules move about very little, there will be few components at the right frequency, and $T_1$ can last for hours. For protons in pure, simple liquids, such as distilled water, $T_1$ and $T_2$ are approximately equal (a few seconds), an indication of the mutual dependence of the relaxation times on the fluctuating, internal magnetic fields.

In liquid and liquidlike materials, then, the ratio of $T_2$ to $T_1$ tends to approach unity, but in solid materials the ratio is very small. The internuclear ("dipole-dipole") magnetic interactions that cause $T_2$ to be extremely short in solids can be effectively canceled by resorting to complex multiple-pulse cy-




**PRECESSIONAL ANGLE, α, continues to increase as long as the rotating magnetic field, $B_1$, is applied. The precessional frequency, however, remains constant, being fixed by the inherent properties of the nuclei and the strength of the static field, $B_0$. The radio-frequency pulse needed to tip the vector of the net magnetic moment *M* through an angle of 15 degrees is described as a 15-degree pulse. Pulses of 90 and 180 degrees cause corresponding increases in the precessional angle. To an imaginary observer rotating with the magnetization vector at the Larmor frequency (*right*) the increase in the precessional angle would appear to be a simple rotation of *M* about the applied field $B_1$, which would appear to be stationary. Whenever a net component of *M* exists in the *x-y* plane, an electromotive force is generated that can be detected by a coil surrounding the sample. This electromotive force is the origin of the NMR signal.**




**IN A PULSED NMR EXPERIMENT** the emitted signal is observed after the radio-frequency energy that induces precession is turned off. In these diagrams the frame of reference is assumed to be rotating at the average Larmor frequency. After a radio-frequency pulse tips the vector of net magnetic moment $M$ through some angle, $\alpha$, there will be a component of $M$ in the $x$-$y$ plane (*b*). For a brief instant the NMR signal is at a maximum (*curve at bottom left*). Nuclei immediately begin to precess, however, at slightly different rates because of magnetic interactions between nuclei and slight nonuniformities in the magnetic field. The net component of $M$ in the $x$-$y$ plane therefore diminishes, and the signal amplitude decays exponentially with a time constant $T_2^*$ (*c, d*). If the magnetic field $B_0$ is perfectly uniform, the signal decay time is longer, and the time constant is $T_2$, called the spin-spin relaxation time. Simultaneously the longitudinal component of magnetization increases as $M$ returns to its equilibrium position, aligned with the $z$ axis (*e*). This relaxation, designated $T_1$ (spin-lattice relaxation time), measures the time needed for the spin system to return to thermal equilibrium (*curve at bottom right*).

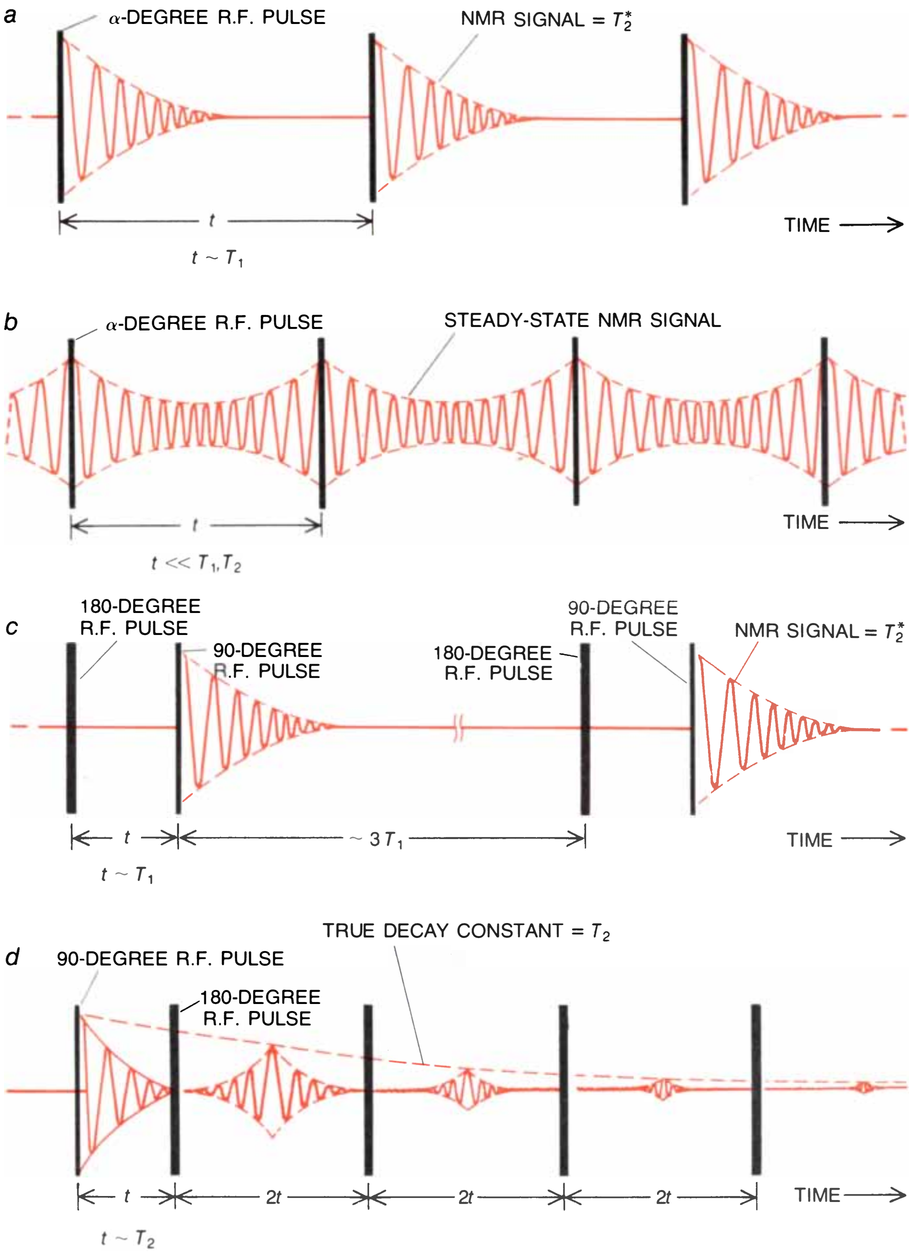
cles. Such cycles, however, have not yet been employed for medical NMR imaging; only liquidlike regions in biological materials yield any appreciable signal.

We have seen that it is possible to determine the material's intrinsic relaxation time $T_2$. It is also possible to determine the intrinsic value of $T_1$. This is done by applying a second pulse after the signal has decayed but before there has been full spin-lattice relaxation. The second pulse will evoke another free induction signal, but the amplitude in this case will be smaller. The explanation is that the strength of the evoked signal is proportional to the longitudinal component of magnetization along the field direction $z$ just before the pulse is applied. Before the initial pulse all the nuclei are aligned with the field, and the evoked signal is therefore maximal. The second pulse, however, is applied while the component of magnetization is still in the process of returning to the aligned position, so that the evoked signal is not as strong as the original one. Since the component of magnetization returns to its equilibrium value exponentially with a time constant $T_1$, the respective amplitude of the two signals is a measure of that constant.

Variations in the value of $T_1$ can be exploited in NMR imaging to increase the contrast between different regions in samples of soft tissue. Differences in $T_1$ values can be brought out by means of special pulse sequences, for example the "saturation-recovery" sequence. Here the sample is exposed to a train of radio-frequency pulses with a constant spacing between pulses. Provided that complete spin-spin relaxation ($T_2$) occurs in each pulse interval, the amplitude of free induction decay signals will reach a steady-state value that is dependent on both the value of $T_1$ and the density of liquidlike nuclei contributing to the signal. The relation to signal strength is such that for areas in the image where $T_1$ is long compared with the interval between pulses the image intensity is weak.

This imaging procedure yields images that are said to be $T_1$-weighted. If two or more images are obtained with different spacings between pulses, it is possible to calculate from the data a $T_1$ value for each pixel, or picture element, in the image, independent of spin density. The result is a "$T_1$ map." The $T_1$ differences identified in this way are extremely useful in proton NMR imaging because proton $T_1$ differences among various soft tissues are much greater than corresponding differences in mobile proton density.

If the pulses are applied so rapidly that the signal does not decay to zero between successive pulses, a different steady-state condition arises in which

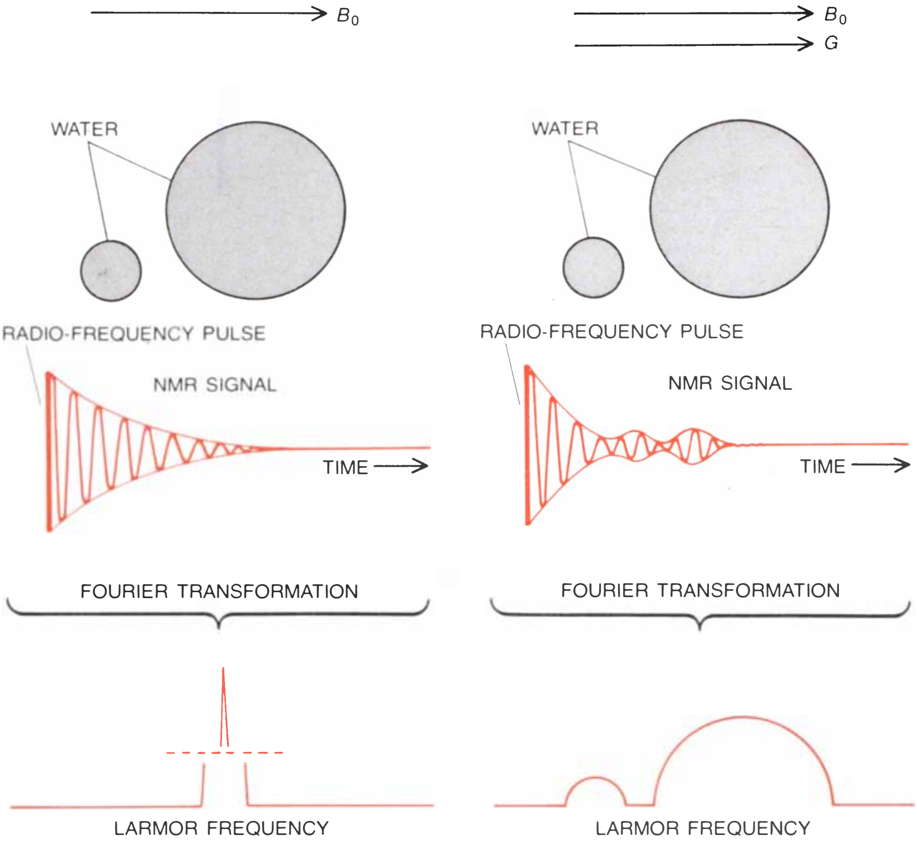



**VARIOUS PULSE PATTERNS that emphasize different aspects of the evoked NMR signal can be exploited to alter the contrast of NMR images. In a "saturation-recovery" pulse sequence (*a*) a series of 90-degree pulses is applied with an interpulse spacing, *t*, longer than the decay time $T_2$ and roughly the same length as $T_1$. When the value of *t* is changed, variations in $T_1$ in different parts of a sample will show up as differences in image intensity and a "$T_1$ map" can be generated. If *t* is much shorter than $T_2$, the signal will not decay to zero between successive pulses, creating a condition known as steady-state free precession, or SSFP (*b*). Now image contrast can be altered by changing the pulse angle α. In the "inversion-recovery" pulse sequence (*c*) the magnetization vector is inverted by first applying a 180-degree pulse and then a 90-degree "read" pulse. This sequence also emphasizes $T_1$ variations in the sample. The "Carr-Purcell spin echo" sequence (*d*) consists of a 90-degree pulse followed by a series of 180-degree pulses. Here the resulting images are strongly dependent on the spin-spin relaxation time $T_2$.**

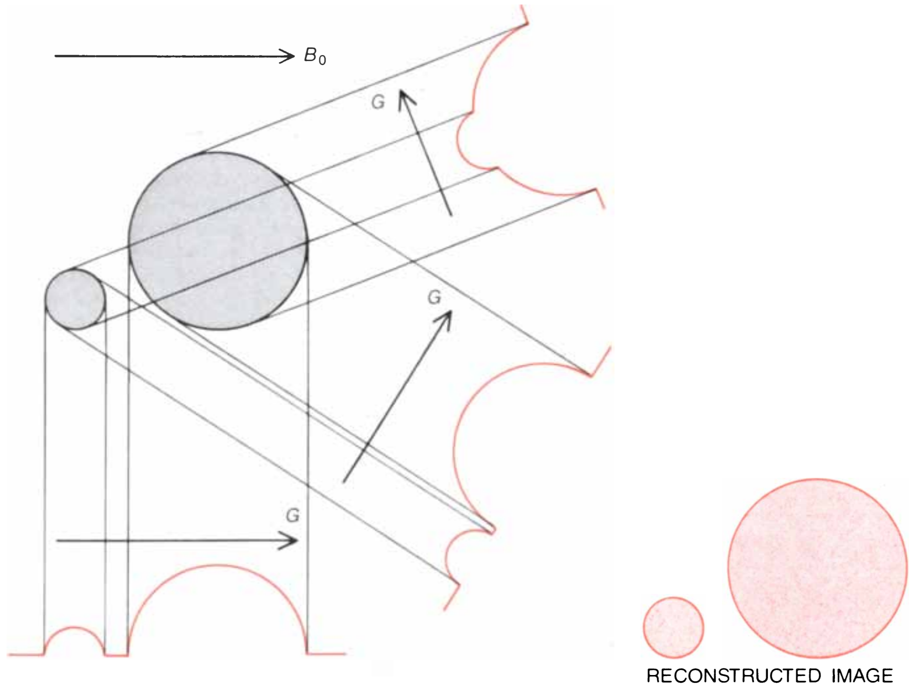
the signal intensity is dependent on $T_2$ as well as $T_1$. This pulse sequence is called steady-state free precession, or SSFP. Although images of very high quality have been generated with the SSFP sequence, it is not easy to separate the individual contributions of $T_1$ and $T_2$ to the resulting image intensity.

Another frequently adopted pulse pattern is the "inversion-recovery" sequence, which resembles the saturation-recovery sequence in that $T_1$ variations in the sample are emphasized. In this pattern the bulk magnetization vector is first completely inverted by applying a 180-degree pulse to the sample. $T_1$ relaxation proceeds during a selected pulse interval, after which a 90-degree "read" pulse is applied. The free induction signal that follows the read pulse serves to generate the image.

The inversion-recovery sequence gives




**METHODS OF CREATING NMR IMAGE call for spatial "encoding" of the NMR signal. Here the sample is water in two cylinders, viewed from above. The hydrogen NMR signal obtained in NMR spectroscopy (*left*) appears as a single tall spike after Fourier transformation, a mathematical process that converts a curve representing signal strength v. time into one representing signal strength v. frequency. If a linear magnetic field gradient, *G*, is now added to the original static field, $B_0$, the evoked signal, after Fourier transformation, takes the form of a curve that is representative of the shape of the sample (*right*). The area under the curves represents the total number of protons in the sample, and therefore the area is equal for both.**




**IN PROJECTION-RECONSTRUCTION IMAGING a magnetic field gradient is rotated to obtain "snapshots" of the sample at many different angles covering an arc of at least 180 degrees. From such a set of data a computer can reconstruct a cross-sectional image of the sample. The method is analogous to the one in X-ray computerized tomography (CT scanning).**

an image of higher contrast than the saturation-recovery sequence, but a penalty is paid in the form of either longer imaging time or reduced spatial resolution. The reason is that if errors in $T_1$ determinations are to be avoided, a delay time equal to at least three times the value of $T_1$ should be allowed to elapse before repetition of the 180- and 90-degree pair of pulses. $T_1$ maps can also be derived from inversion-recovery pulse patterns. The spin-echo pulse sequence mentioned above is useful for creating images that are primarily or solely dependent on the spin-spin relaxation time, $T_2$.

Therefore by choosing an appropriate pulse sequence, the intensity of an NMR image can be made to reflect one or more of several NMR parameters inherent to the tissue being examined. Such parameters in turn are sensitive to the physicochemical environment of the nuclei and underlie the conviction that NMR imaging holds promise for detecting disease early and for monitoring its progress.

The foregoing methods for manipulating the NMR response apply not only to NMR imaging but also to NMR spectroscopy, in which the signal elicited represents a summation of the NMR response from the entire sample of material. But if the nuclear signal comes from the entire sample, how can it be encoded with spatial information? As long ago as 1951 Robert Gabillard of the École Normale Supérieure in Paris was puzzled by the reverse question: he noted that the NMR signal could be distorted according to the shape and size of the sample. This was correctly attributed to nonuniformities in the static magnetic field. The degree of distortion depends on how much of the sample is in nonuniform parts of the field and on the magnitude of the nonuniformities.

A continuing effort in NMR instrumentation has therefore been to remove the influence of sample shape on the NMR signal by making magnets with fields of ever increasing homogeneity and stability. The quest may conceivably have delayed the advent of NMR imaging, because in imaging it is necessary to make the field nonuniform deliberately, albeit in a controlled manner, usually by superimposing on it a linear magnetic field gradient.

The primary motivation for perfecting the static field of magnets has come from the desire of NMR spectroscopists to measure the subtle "chemical shift" in samples incorporating molecules of complex structure. One might expect all nuclei of the same species in a homogeneous magnetic field to have the same resonant frequency; that is to say, one would expect to observe a single, narrow peak in the NMR frequency spectrum. For nuclei in samples consisting of very simple molecules (such as for the hydrogen nuclei of pure water) this is indeed so. For more complex molecules, however, the magnetic field is subtly altered around some nuclei as a result of "shielding currents" that are associated with the distribution of electrons around adjacent atoms. Such alterations cause shifts in the resonance frequency that are representative of particular molecular conformations and therefore directly aid in the determination of chemical structure.

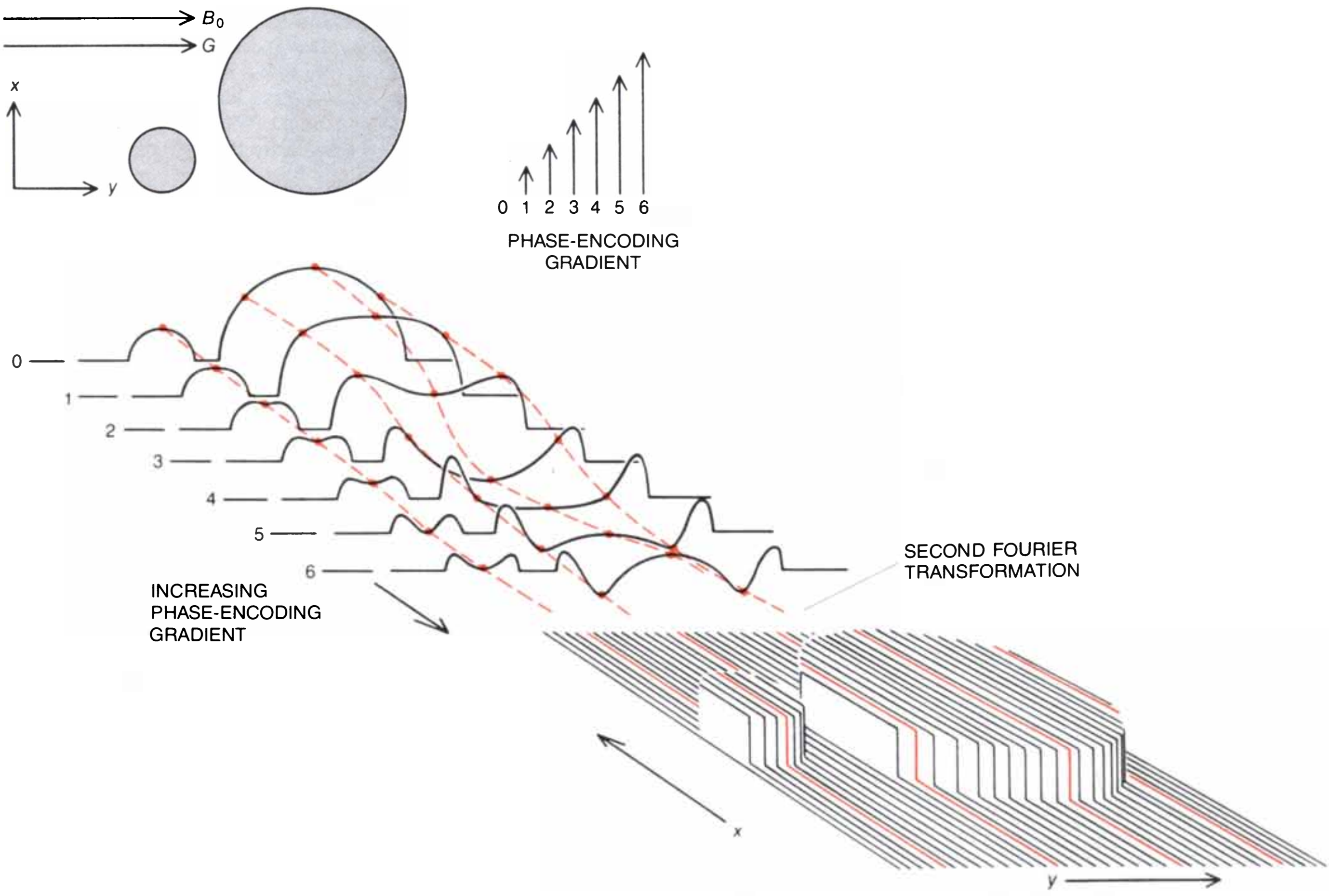
These chemical shifts are very small, being measured in parts per million with respect to the strength of the static magnetic field, and this explains why such emphasis has been placed on the construction of magnets with a highly uniform field. In the current generation of NMR imaging systems the static magnetic field is not uniform enough to show chemical shifts with a useful degree of accuracy. In any case the imaging procedure itself (specifically the application of the field gradients) often precludes the extraction of such information.

The first published NMR image can be credited to Paul C. Lauterbur of the State University of New York at Stony Brook. In a 1973 article he presented images of two water-filled capillary tubes, obtained with a modified NMR spectrometer. The sample can be regarded as a medium that couples the radio-frequency field to the gradient of the magnetic field, and so Lauterbur called his imaging method zeugmatography, from the Greek *zeugma,* that which joins together.

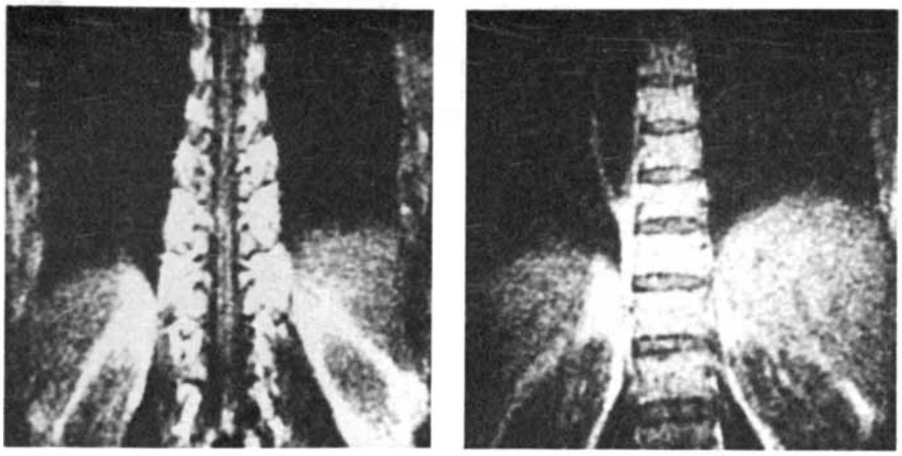
NMR imaging systems can be designed to receive data from a single point in the sample, from a line, from a plane or from the complete three-dimensional volume all at once. In point- or line-scanning methods images can be generated by electronically moving the selected point or line through the sample in a raster, or sequential pattern. Such methods, however, have been almost entirely superseded by two- and three-dimensional methods, which are much more efficient.

Most of the two- and three-dimensional methods call for the application of a linear magnetic field gradient to supply one dimension of spatial information. Resolution in the second and third dimensions can be achieved by sequentially stepping gradient magnitudes or directions through a predetermined set of values. There are other approaches. Rotating-frame zeugmatography, first reported by David Hoult of the National Institute of Dental Research, is unique in that it exploits a gradient in the radio-frequency field as well as one in the static magnetic field. Peter Mansfield of the University of Nottingham has devised a very high-speed method, echo-planar imaging, that generates a complete, spatially resolved image following a single radio-frequency pulse.

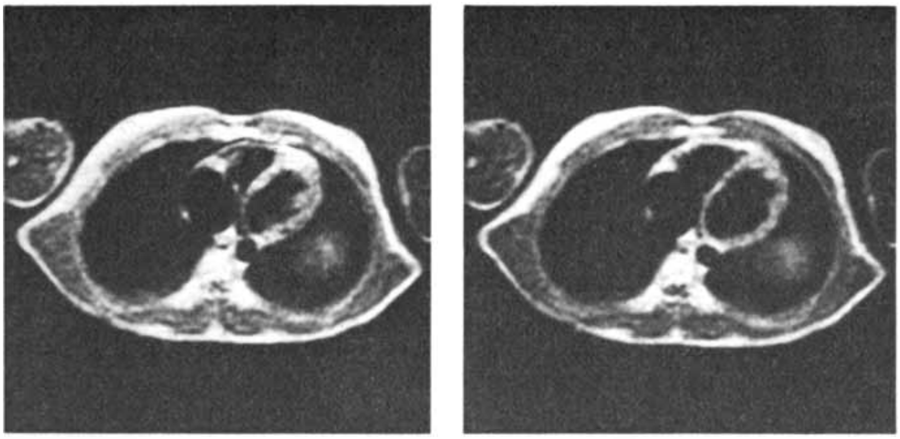
The first images produced by Lauterbur borrowed image-reconstruction computer algorithms, used also in CT scanning. If a sample of water is placed in a homogeneous magnetic field, the NMR frequency spectrum of the hydrogen nuclei in the water molecules is a single narrow line. If the magnetic field



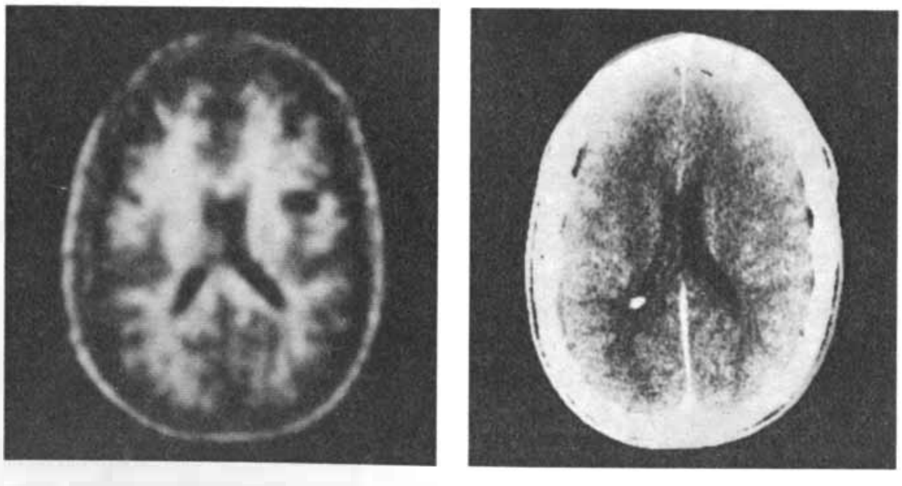

**IN FOURIER ZEUGMATOGRAPHY another magnetic field gradient is applied for a short time just before the usual gradient, *G*, is turned on. This new gradient, termed a phase-encoding gradient, is applied at right angles to the original one, and its magnitude is raised from zero in a series of many small steps. The effect of the two gradients on the NMR signal is such that after Fourier transformation a series of increasingly "phase-distorted" projections are generated. Corresponding points from each of the projections (*broken lines in color*) are then subjected to a second Fourier transformation to generate the final image. The term zeugmatography is from the Greek *zeugma,* that which joins together. It alludes here to the way the sample couples the radio-frequency field and the magnetic-gradient field.**



**LUMBAR REGION OF THE HUMAN SPINE is depicted in two reconstructions from three-dimensional NMR data set. The image at the left shows the spinal cord within the spinal canal. In the image at the right the plane of the image has been moved a few centimeters to reveal the intervertebral disks. It is difficult to obtain clear pictures of the spine over such a large area with a CT scan or any other imaging method. The images were provided by Technicare.**



**"GATED" IMAGES OF THE HEART are shown in NMR cross sections of the human chest. (The two smaller shapes at each side of the chest are the arms.) The image at the left shows the heart at the end of systole, when the chambers are being emptied. The image at the right shows the end of diastole, when the chambers are filled. The images, supplied by Technicare, were made by gating, or synchronizing, the recording of data to match the stage of the cardiac cycle.**



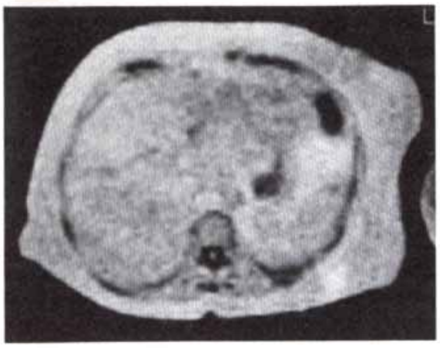
**NMR IMAGE AND X-RAY CT-SCAN IMAGE can be compared in these two transverse views of the brain of a patient with a brain tumor. In the NMR image (*left*), reconstructed from three-dimensional data, the tumor is a dark circular area at the right side of the brain. In the CT scan (*right*) the tumor is all but invisible. The pictures, made at the Massachusetts General Hospital, are reproduced with permission of *Journal of Computerized Axial Tomography.***

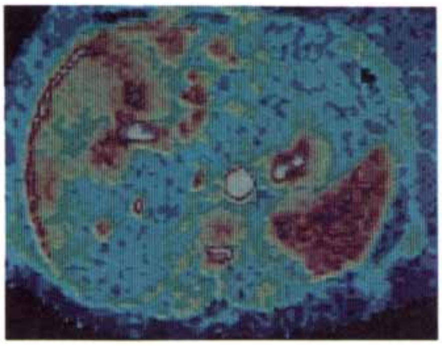
is perfectly uniform, the shape of the line is independent of the geometry of the sample. If a linear magnetic field gradient is now superimposed, resonant nuclei at one side of the sample will feel a weaker total magnetic field than those at the other side. There will thus be a linear distribution of Larmor frequencies across the sample. Then the free induction decay signal can be subjected to Fourier transformation, a mathematical procedure that transforms the data from a curve representing signal strength v. time into one representing signal strength v. frequency.

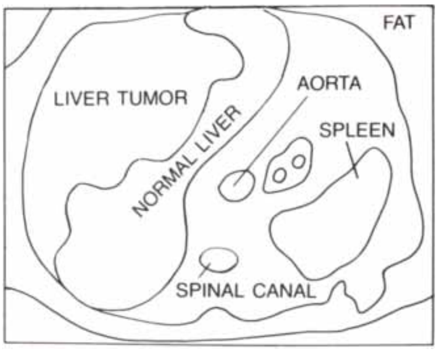
The result is a spectrum that is broadened to a shape corresponding to the one-dimensional projection of the strength of the NMR signal onto the frequency axis. By rotating the magnetic field gradient electronically one can get a projection from a slightly different angle. Computer analysis of many such projections reconstructs the sample's geometry. In the two-dimensional application of the technique the direction of the gradient is rotated within a single plane. In the three-dimensional extension of the method the gradient is rotated in three-dimensional space through at least half a sphere.

Two-dimensional Fourier-transformation imaging, or Fourier zeugmatography, was first demonstrated in 1975 by A. Kumar, D. Welti and R. R. Ernst of the Swiss Federal Institute of Technology. Their method owes more to NMR spectroscopy than to CT reconstruction algorithms, because both amplitude information and phase information are acquired to spatially encode the signal. The first step again calls for generating a one-dimensional projection, but here a phase-encoding gradient is applied just before the original gradient is turned on. The phase-encoding gradient is applied at right angles to the original gradient, and its duration (or amplitude) is successively increased from zero, rather than the gradient's being progressively rotated, as was done in the original form of zeugmatography. The phase-encoded projections are stacked in increasing order of magnitude of the phase-encoding gradient, and the corresponding points from each projection are Fourier-transformed a second time to generate the final image [*see illustration on preceding page*].

Nuclear magnetic resonance is inherently a three-dimensional phenomenon. Since NMR signals are normally obtained from the total volume of material enclosed within the transmitter and receiver coils, considerable ingenuity is required to reduce the volume from which signals come to defined points, lines or planes. In one method, called selective irradiation, a radio-frequency pulse is applied that is specially tailored








**MIDSECTION OF A CANCER PATIENT appears in two NMR images made by James M. S. Hutchison, Francis Smith and John R. Mallard and their co-workers in the Faculty of Medicine at the University of Aberdeen, using a .04-tesla imaging system. The black-and-white image is a simple record of proton density. The colored image is a computer synthesis of a series of images obtained by altering the pulse sequence to discriminate among tissues on the basis of their $T_1$ relaxation-time differences. Color coding was then introduced to enhance the discrimination. The patient had a cancer of the rectum that had spread to the liver. The normal liver tissue appears light blue. The cancerous regions are yellow-brown. The spinal cord and the spinal canal are more clearly seen in the proton-density image.**

to consist of only a very narrow band of frequencies. Only those nuclei lying within a single slice perpendicular to the direction of a plane-selection gradient will exhibit resonant frequencies corresponding to those in the radio-frequency pulse. Hence only a thin, isolated slab of material is irradiated. The thickness or position of the plane can be altered by changing the width or the frequency offset of the irradiation spectrum electronically.

A second method, devised by Waldo S. Hinshaw, who was then working at the University of Nottingham, imposes an oscillating magnetic field to select a particular plane. In this method the direction of the plane-selection field gradient is periodically reversed, often sinusoidally. There is then just one plane at the fulcrum of oscillation within which the magnetic field remains time-independent; signals from outside this plane vary in such a way that they do not contribute to the intensity of the image. These are just two of several possible methods for selecting planes.

Although whole-volume methods offer important advantages, they also have some technical drawbacks. The sheer mass of data acquired calls for a computer with a large data-handling and storage capacity. For example, generating a three-dimensional image that shows 256 data points in each dimension, with 256 levels of signal intensity (equal to eight bits) per data point, calls for a system with more than 134 million bits of memory ($256^3 \times 8$). Moreover, many sequential values of field-gradient magnitude or direction are needed to uniquely define all the points in such a three-dimensional data matrix. This lengthens the data-collection time, particularly if $T_1$ maps are desired. For such reasons it is sometimes preferable to generate a small number of selected two-dimensional images. Although it is true that a single two-dimensional image can usually be acquired in less time than a true three-dimensional image, the three-dimensional one can be "dissected" at leisure into a great many slices, so that the imaging time per plane is much reduced.

The spatial resolution of a three-dimensional set of data is usually isotropic, or equal in all three dimensions. Therefore two-dimensional slices of selected thickness, with any position or orientation, can be generated at any time after the primary data have been acquired. With three-dimensional data in hand, surfaces can be detected mathematically, enabling the clinician to determine the volume of organs or of pathological lesions.

In medical practice many factors must be considered when a particular imaging method is being chosen, particularly the time scale of involuntary movements of the tissue being studied. The head, for example, is particularly amenable to true three-dimensional imaging because it can be held still for the duration of the scan. The heart, on the other hand, which beats incessantly, requires either a high-speed imaging method or one that can "gate," or synchronize, the data collected over a series of cardiac cycles.

The apparatus needed for NMR imaging exploits the same basic technology developed for NMR spectroscopy. Indeed, many of the early imaging experiments were conducted with modified NMR spectrometers. The signal-to-noise ratio of an NMR image can be improved by increasing the strength of the static magnetic field of the apparatus. As the field strength is increased the Larmor-precession frequency of the nuclei in the sample of material being examined increases linearly, and a higher radio frequency is needed. The drawback is that both the transmitted and the emitted signals are more strongly absorbed as the frequency is increased. For imaging the whole human body such attenuation may become a limiting factor when the frequency exceeds about 15 megahertz, which for proton imaging corresponds to a field strength of .35 tesla (3,500 gauss). This field strength would be regarded as a low one in NMR spectroscopy, but in NMR imaging the large working volume in which uniformity of magnetic field is needed has called for novel approaches to magnet design.

At present the two commonest magnet designs are the four-coil air-core ambient-temperature magnet and the large-bore, helium-cooled superconducting magnet. NMR systems with conventional nonsuperconducting magnets are less costly and are entirely satisfactory for whole-body proton imaging at field strengths below about .2 tesla, for which the power consumption is about 50 kilowatts and cooling requirements are not prohibitive. Superconducting magnets, although they are initially much more expensive, may have lower operating costs and are able to generate magnetic fields that are both stronger and stabler than those attainable with nonsuperconducting magnets. Superconducting magnets are therefore preferred if it is intended to image nuclei other than hydrogen, which demand significantly stronger fields. For example, the observation of phosphorus 31 at 15 megahertz calls for a field strength of .87 tesla. One critical component of an NMR imaging apparatus not usually found in an NMR spectroscopy apparatus is the gradient coil system, which is needed to create linear field gradients whose magnitude must sometimes be altered very rapidly. The conflicting requirements of large-magnitude gradients and fast switching times have given rise to many original designs.

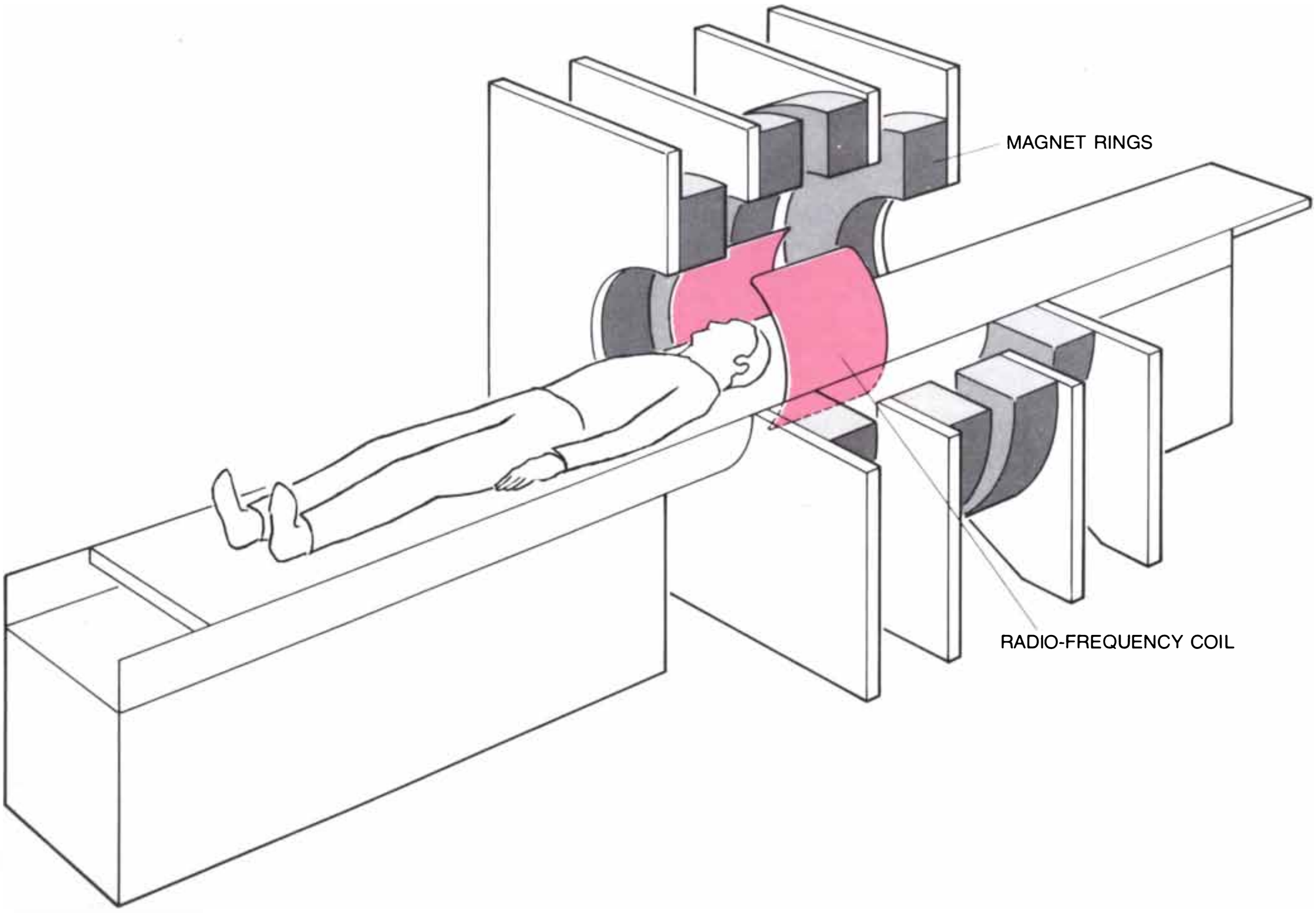
It should be emphasized that the spatial resolution of an NMR image is not dictated by the wavelength of radiation that yields the image, as it is in many imaging systems. (The wavelength of

15-megahertz radiation in free space is 20 meters.) In NMR imaging spatial resolution depends rather on the uniformity of the static magnetic field and on the strength of the gradient fields. The $T_2^*$ relaxation time of the free induction signal is usually the dominant decay process in current NMR imaging systems, and therefore $T_2^*$ sets a minimum line width of the NMR spectrum. The strength of the gradient fields must be high enough for the frequency increments from pixel to pixel in the final image to dominate the $T_2^*$ line width. Some spatial encoding methods, however, require that this condition be satisfied less rigorously than others.

NMR imaging in medicine is still at an early stage. A systematic body of data must be built up so that the efficacy of the technique for detecting pathological lesions in various parts of the body can be established. Early clinical results at the Massachusetts General Hospital with two imaging systems made by the Technicare Corporation suggest that NMR may be particularly good at detecting necrotic (dead) tissue, ischemia (local anemia caused by mechanical obstruction of the blood supply), malignancies and degenerative disease of various kinds. The soft-tissue contrast is inherently superior to that of X-ray techniques, and the sensitivity for the detection of lesions in general appears to be very high.

One promising idea is to administer NMR "tracers," or contrast materials, orally or by injection. For example, paramagnetic ions such as those of manganese ($Mn^{++}$) have unpaired electrons, so that they have a magnetic moment and tend to align themselves in the static magnetic field of an NMR system. They increase the local magnetic field slightly and substantially modify the relaxation time of NMR-sensitive nuclei. The resulting change in image intensity would allow the distribution of the paramagnetic species to be mapped. Such tracers could play a role analogous to that of the radioactive tracers administered in nuclear medicine. Several laboratories are investigating the feasibility of using NMR to measure the velocity at which blood flows through the major vessels, and recent experiments show that the action of the heart in various parts of its cycle can be "frozen" by high-speed or stroboscopic NMR imaging.

Perhaps the greatest potential of all lies in the imaging of nuclei other than hydrogen, particularly the phosphorus nucleus. Phosphorus is a major constituent of the high-energy molecules adenosine triphosphate (ATP) and phosphocreatine, which mediate the transfer of energy in the living cell. NMR spectroscopy of phosphorus in localized volumes within the body clearly reveals several chemically shifted resonance peaks whose heights correspond to the concentration of the individual phosphorus compounds. From knowledge of such concentrations it is possible to infer the metabolic status of internal organs, and it may eventually be possible to add this capability to an imaging instrument. The next few years will undoubtedly see both an improvement in the quality of NMR images and a growing diversity of applications for nuclear magnetic resonance in clinical practice.




**NMR IMAGING MAGNET is depicted schematically. The four rings hold windings that generate a static magnetic field ($B_0$). The radio-frequency pulse is generated by a coil in the form of two curved panels. The NMR signals evoked by the radio-frequency pulses are detected by the same coil and forwarded to a computer. Costlier systems with higher magnetic fields employ superconducting magnets.**


It's difficult to ask a blood chemistry analyzer to say "Aah." Or have a microfilmer take a deep breath and hold it. Or get the pulse of a color print processor.

Difficult. But not impossible.

Rapid, precise diagnosis is essential to the productivity of capital equipment. And it's the foundation of our service philosophy.

The field maintenance processor shown here is just one example of the extra dimension of serviceability that you can expect in Kodak equipment. Linked into the controls of a customer's machine, this portable, state-of-the-art microcomputer swiftly checks internal functions (software drive, electrical and/or mechanical), allowing the Kodak equipment service representative to isolate and correct the problem, quickly.

Another is a unique telephonic link-up that enables the equipment to "call in" its vital signs to a sophisticated diagnostic computer that responds instantly to the ESR on site.

And once diagnosis is made, the treatment is equally rapid because Kodak designs serviceability into its equipment from the very first blueprint... with drop-in subassemblies and built-in ease of access.

Of course, all this leading-edge technology is backed by extensive training of our technicians (and, if desired, of your people). Backed by a computerized global network of parts stocks and a variety of service and maintenance options. And, most important, backed by the Kodak commitment to respond to the market's needs for productivity.

We know that you evaluate capital expenditure by ROI. Equipment ROI depends largely on maximum uptime. And that uptime often hinges on service. Timely, thorough service.

If you'd like to know more about the ways we can get a machine to stick out its tongue, send for a copy of our Brochure A-3000, "Performance Plus." Write to Eastman Kodak Company, Dept. GB-SA-1, 343 State Street, Rochester, NY 14650.

Kodak

# Before you look under the ho

At Honda, we believe the car that goes together better stays together better. So we're very precise when we assemble Honda automobiles.

Hood, doors, trunk and hatch fit the body with close tolerances all around. Body panels and moldings align.

Run your hand along the edges and down the seams. The whole car feels smooth to your touch. That's the result of painstaking preparation throughout.

Lift the hood and you'll see the pains our engineers have taken to make servicing less troublesome. The engine compartment is well organized.

Open the door and look inside. There's not an unfinished area or rough edge in sight. Carpets fit snugly. No wrinkles or loose threads in the headliner. The upholstery seams are neat.

# od, look at how the hood fits.

Sit behind the wheel and feel the controls. They work the way they were designed to work. Try the smooth action of the front seat adjustment. It's all done right. It meets Honda quality.

Honda quality is an important reason why Honda owners are so loyal. Why they continue to buy more Hondas. Honda is number one in owner loyalty among all leading import cars.*

And Honda cars have traditionally retained a high resale value, among the best in the automotive industry.

To understand why, you only have to look at how the hood fits.

*Source: R. L. Polk and Co. 1981 model year *New Car Buyer Analysis.*

HONDA

We make it simple.

# SCIENCE AND THE CITIZEN

## *Gigatons*

Discussions of nuclear weaponry tend to focus not on the bombs and warheads themselves but on the launch vehicles: the aircraft and missiles that would deliver the weapons. One reason for this emphasis is that public information about the nuclear explosive devices is meager. Nevertheless, an unofficial inventory of the American nuclear arsenal has recently been compiled by the Center for Defense Information and the Natural Resources Defense Council. The inventory, which is based on published information, shows that the U.S. stockpile includes some 30,000 nuclear weapons, of which 23,000 are operational. An analysis of the Reagan Administration's budget requests suggests there are plans for building 17,000 more weapons in the next 10 years.

Some of the new weapons will replace existing ones, which will then be dismantled, but an estimated 10,500 bombs and warheads will be added to the stockpile. Most of the additional devices will be incorporated into strategic (rather than tactical) weapons systems. According to estimates by the Center for Defense Information, the U.S. now has about 12,000 nuclear weapons that could be delivered onto the territory of the U.S.S.R.; after the scheduled buildup the total will approach 20,000. Although the rate at which weapons have been manufactured in the past is not known with certainty, a plan to build 17,000 weapons in 10 years is thought to represent a significant acceleration. Indeed, there is some question whether the two existing plants for the production of nuclear-weapons materials can meet the increased demand.

Several of the new nuclear munitions are about to go into production. They include warheads for two types of cruise missile and a high-yield nuclear bomb. Among weapons in an earlier stage of development are warheads for the MX land-based strategic missile and the Trident II submarine-launched missile.

Calculations based on the unofficial inventory indicate that the total explosive power of the operational nuclear weapons is at least two gigatons, or the equivalent of two billion tons of TNT. This is a minimum estimate; the actual total is probably several times greater. The ratio of the explosive yield to the Russian population is on the order of 10 tons per capita.

## *Quantum Immunodynamics*

The immune response begins when a foreign antigen binds to receptors on the surface of a cell called a *B* lymphocyte and thereby stimulates the cell to proliferate and manufacture antibodies against the antigen. The receptors are specimens of the particular kind of antibody the cell can make; their specificity is the end result of a complex process of gene shuffling in the developing *B* cell (see "The Genetics of Antibody Diversity," by Philip Leder, page 102).

A number of puzzling observations about the triggering of the immune response have long gone without a satisfactory explanation. One such puzzle is the bell-like shape of the dose-response curve: *B* cells make more antibody as the amount of an antigen increases to a certain maximum, after which antibody production falls off with further increases in the dose of the antigen. Another observation is that an antigen is most strongly immunogenic when it is multivalent: when it presents to the *B* cells a large array of identical sites, each of which serves as an epitope, or antigenic determinant, that fits the antibody's combining sites. The surface of a bacterium, a virus or a pollen grain is studded with such epitopes; smaller "nonself" particles, including most proteins, are immunogenic only when they have somehow been aggregated. Small molecules called haptens bind readily to antibodies, but they stimulate an initial immune response only when they are arrayed on a large carrier protein.

What kind of triggering mechanism might account for these observations? Three investigators at the Johns Hopkins University School of Medicine have concluded that the triggering event is quantized. They propose that a minimum number of receptors on a single *B* cell must be bound to epitopes on a single antigen; the receptors are thereby drawn together to form a cluster called an immunon. The formation of enough immunons triggers the developmental process that leads to the manufacture of antibodies in quantity. The immunon theory has been incorporated into a mathematical model that makes quantitative predictions, which have been tested experimentally. Howard M. Dintzis, Renee Z. Dintzis and Bert Vogelstein describe the model, list the predictions and present their experimental confirmation in two recent papers in *Proceedings of the National Academy of Sciences.*

Dintzis, Dintzis and Vogelstein based their theory on general observations of the kind noted above and on experiments they did with artificial antigens: linear polymers to which a number of haptens are attached, so that each hapten serves as an individual epitope. (The polymer employed was polyacrilamide and the haptens were dinitrophenyl groups.) They found that a polymer molecule was immunogenic only when it had between 14 and 21 effective (appropriately spaced) haptens. Smaller molecules, with from seven to nine effective haptens, did not stimulate antibody synthesis; in fact, they inhibited triggering of the immune response by the larger molecules.

The immunon theory holds that a *B* cell capable of responding to a hapten-bearing molecule has a large number of hapten-specific receptors on its surface. Sequential binding of the receptors to haptens on a single immunogenic molecule leads to the clustering of the receptors (which are free to move about the surface of the cell membrane). Once the cluster includes a critical number of polymer-linked receptors, it undergoes a slow structural transformation to become an immunon. The cell receives a stimulus when enough immunons have formed. The strength of the primary immune response is therefore related to the rate of immunon formation in the population of cells having receptors for a particular hapten. An immunogenic molecule is defined as one able to bind at least the critical number of receptors. The presence of too many immunogenic molecules decreases the average size of the clusters because receptors bind to haptens on different molecules rather than binding to multiple haptens on a single molecule and thus being drawn into a cluster. This effect could explain the falloff observed in dose-response curves. Polymers that are not strongly immunogenic (because they have too few haptens) similarly compete nonproductively for receptor sites, perhaps explaining the inhibitory effect of small molecules.

The mathematical model developed by Dintzis, Dintzis and Vogelstein is a set of equations describing the formation of immunons in the presence of immunogenic and nonimmunogenic molecules. From the equations they plotted curves predicting the response to various doses of an immunogenic antigen and the inhibitory effect of various doses of a nonimmunogenic one. They tested their predictions by injecting mice with hapten-bearing polymers and also by incubating immune-system cells (from the mouse spleen) with the polymers. The data on antibody formation from both kinds of experiment are in remarkable agreement with the calculated curves.

## *Beyond the Black Horizon*

In the standard big-bang model of cosmology the observed expansion of the universe is extrapolated backward in time to a singularity: a hypothetical


80 PROOF IMPORTED BY © W.A. TAYLOR & CO. MIAMI, FLORIDA 1981
PAINTING: THE CALVIN BULLOCK COLLECTION
Courvoisier
VSOP
Fine Champagne
COURVOISIER
The Brandy of Napoleon
COGNAC
Produced and Bottled by
Courvoisier, Jarnac, France
COURVOISIER. THE COGNAC OF NAPOLEON


Sheaffer introduces Slim Profile Plus Dual Writing System—as slim and sleek as a pen ought to be. It comes with interchangeable micro tip marker and rolling ball, in surfaces from 23-karat gold electroplate to black matte to stainless steel. Each pen, a doubly remarkable gift.
SHEAFFER®
JEWELRY FOR WRITING
SHEAFFER EATON TEXTRON

state of infinitely dense, infinitely hot matter in the four-dimensional geometry of space-time. The singularity exploded to create the primeval fireball out of which galaxies and stars eventually coalesced. The most compelling evidence in favor of this view, apart from the universal expansion itself, is the cosmic microwave background radiation. Most cosmologists hold that this faint, all-pervasive radio noise is a remnant of the high-energy thermal radiation that dominated the universe in the immediate aftermath of the big bang.

There is a problem with this interpretation, however. It arises from the fact that the cosmic microwave background is almost perfectly isotropic, that is, the intensity and the "temperature" of the radiation are essentially uniform in all directions. The isotropy implies that different parts of the early universe were in thermal equilibrium with one another. If the primeval fireball originated in a singularity, however, its initial expansion would have been too rapid for the different parts of the universe to have been causally connected, even by means of light waves. Hence the standard big-bang model, by failing to allow enough time for thermal equilibrium to be attained, appears to violate the principle of causality.

To deal with this problem several modified versions of the big-bang model have been proposed. In one class of models the universe is assumed to have had an early constant-density phase in which the density was very high but less than the infinite density of a singularity. The postulated state is characteristic of a particular kind of curved space-time called de Sitter space, after the Dutch astronomer and mathematician Willem de Sitter, who defined its properties mathematically in 1917. By allowing more time for the parts of the early universe to have communicated with one another such models avoid the causality problem.

A new model of this kind, formulated by J. Richard Gott III of Princeton University, goes a step further. In his model the origin of the universe is traced back to the spontaneous appearance of a low-density bubble in the primordial high-density de Sitter space. The advantages of this concept are enumerated by Gott in *Nature.*

In common with some other modified big-bang models Gott's model presumes no singularity and allows the various parts of the early universe to be causally connected. The notion of bubbles appearing spontaneously in de Sitter space is also not unique to Gott's model. Nevertheless, Gott's bubble universe has a number of novel features.

For one thing, according to Gott, outside the bubble in his model universe there is an event horizon. Light from some parts of the primordial de Sitter space beyond the bubble wall can never reach an observer inside the bubble.

The concept of an event horizon is more familiar in the context of a black hole: the site of any exceedingly compact mass, such as a collapsed star, whose gravitational field is strong enough to prevent light from escaping. In 1974 the British astrophysicist Stephen Hawking made a remarkable discovery about such an event horizon: it can appear to emit radiation. According to quantum mechanics, pairs of particles and antiparticles are continually created everywhere in space-time. Ordinarily the particle and the antiparticle move apart only briefly before coming together again and annihilating each other. In the vicinity of a black hole, Hawking showed, one member of a pair may fall inside the black hole's event horizon, leaving the other member free to escape. The apparent thermal radiation emitted near the event horizon is called Hawking radiation.

The event horizon surrounding a black hole formed by a collapsed star would produce too little Hawking radiation for it to be detected. In Gott's model, however, the amount of Hawking radiation generated by the cosmological event horizon associated with the hypothetical bubble universe would be enormous. In the first $10^{-42}$ second or so after the appearance of the bubble the radiation would cool enough to form all the matter (and antimatter) in the universe. Thereafter the expansion of the universe would proceed just as it does in the standard big-bang model. The cosmic microwave background, however, would be the remnant of the primeval Hawking radiation, and the observed isotropy of the cosmic background would be attributable to the uniformity of the early cosmological event horizon in the constant-density de Sitter space.

The assumption that the early universe was filled with intense Hawking radiation, Gott notes, is consistent with recent grand unified theories of elementary-particle interactions. The theories address the question of how the preponderance of matter over antimatter observed in the universe today could have evolved from an earlier state of pure thermal radiation with equal amounts of matter and antimatter. The answer suggested is that the asymmetry arose from particle-decay processes in the expansion phase of the standard model.

Gott's bubble universe is said to be open, that is, it is infinite in extent and will continue to expand forever. Nevertheless, it may coexist in space-time with other bubble universes, perhaps an infinite number of them. These separate open universes, he says, would all arise from the same hot, dense de Sitter space "like bubbles in a boiling liquid." Because they would lie beyond our cosmological event horizon, however, they would be causally disconnected from our universe (and from one another). As a result they would be forever undetectable.

## *Mortal Complications*

In April of 1935 a 27-year-old man named Leonard Thompson was admitted to Toronto General Hospital. He had been suffering from nausea and vomiting for six days before his admission. The diagnosis was severe ketoacidosis, an increase in the acidity of body fluids; it is a common complication of diabetes. In the hospital record it was noted that Thompson had had diabetes mellitus for 15 years and that he had been receiving insulin for 13 years. Although a strenuous effort was made to save his life, he was extremely weak and died on his third day in the hospital.

The death of Leonard Thompson would have been unremarkable except for two facts. He was the first person to be treated with insulin for diabetes. Furthermore, it has been widely accepted that he was free of the effects of diabetes after insulin treatment was begun and that he died in a motorcycle accident. In *The New England Journal of Medicine* Gerard N. Burrow, Barbara E. Hazlett and M. James Phillips of the University of Toronto Faculty of Medicine review Thompson's case history. They show that he died of pneumonia but at the time of his death had several significant complications of diabetes.

Thompson's death points up another fact that has become apparent in the 60 years since he received the first therapeutic injection of insulin. Contrary to early claims, insulin does not cure diabetes; it is only a means of controlling the major symptom of the disease, an elevated level of glucose in the blood. Of the two major forms of diabetes the one designated diabetes mellitus generally has the earlier onset and the more dangerous consequences. It results from the inability of the pancreas to manufacture a sufficient quantity of insulin, the protein hormone required for the breakdown of carbohydrates.

By early 1922 Frederick G. Banting, an orthopedic surgeon, Charles H. Best, a graduate student, J. J. R. Macleod, a physiologist, and J. D. Collip, a biochemist, working together in Toronto, had purified an extract of pancreatic tissues that lowered blood glucose in dogs. Banting and Best had demonstrated the safety of the extract by injecting themselves with it; they were ready to try it on a patient with diabetes. Thompson, who was 13 years old at the time, had been admitted to Toronto General Hospital in December, 1921. When the pancreatic extract was administered to him on January 11, 1922, his blood glucose decreased sharply. On January 23 a purer extract was administered and an even more dramatic decrease was observed. Thompson continued to receive injections of insulin for the rest of his life.

These events represented a major therapeutic advance; up to then the only therapy for diabetes had been a restricted diet. According to Burrow and his colleagues, soon after Thompson was first treated "reports abounded that insulin represented the cure for diabetes mellitus." The risks and benefits of insulin injection are actually more complex. An editorial accompanying the article by Burrow and his colleagues observes that "the case was the first example of the conversion of what is now known as insulin-dependent diabetes from an acute, fatal disease into a chronic disorder with degenerative complications."

The records of Thompson's autopsy indicate that he had several such complications, including generalized atherosclerosis, enlargement of the liver and pathologies of the central nervous system. It is now generally acknowledged that the current treatment regimen, which entails daily injections of insulin extracted from the tissues of pigs or cattle, cannot control these side effects.

Three approaches to improving the treatment of diabetes are now being tried. The first is the development of methods of releasing insulin into the bloodstream in response to changes in the level of blood glucose rather than on a regular schedule. Such a strategy for mimicking the body's own release of insulin might employ surgical implants containing the hormone. It might provide a more sensitive control of blood-glucose levels. The second approach is the manufacture of human insulin. It has been suggested that human insulin might control side effects better than animal insulins do. For obvious reasons natural human insulin is a rare commodity. New genetic methods, however, have made it possible to synthesize human proteins in bacterial cells. The third approach is the transplanting of cells or tissues from the pancreas of a healthy individual into the pancreas of the diabetic patient; the operation is now an experimental procedure.

## *Demon Deuteron*

According to the quark model of subnuclear particles, the proton and the neutron are not elementary; instead each is made up of three quarks. It follows that in a large atomic nucleus there are many quarks. In general each proton and neutron in the nucleus is thought to retain its identity as a distinct particle, but there remains the possibility that the quarks could be organized in other ways. Now Sverker Fredriksson and Magnus Jändel of the Royal Institute of Technology in Stockholm have suggested that there is evidence for the existence of a new kind of nuclear matter whose internal organization exhibits such a shuffling of quarks. The proposed new state of matter has the same quark composition as the deuteron, the nucleus of the heavy isotope of hydrogen made up of one proton and one neutron. Instead of condensing into two triplets, however, the six quarks in the hypothetical nucleus are organized in three groups of two quarks each. Fredriksson and Jändel call the new nucleus the demon deuteron. They discuss their findings in *Physical Review Letters.*

The main evidence on which Fredriksson and Jändel base their hypothesis was presented in the fall of 1980 by a group of experimenters working at the Lawrence Berkeley Laboratory of the University of California. The experimenters directed a high-energy beam of ionized oxygen or iron atoms into a photographic emulsion, which was then developed to reveal the tracks of electrically charged particles. Each collision of an ion in the beam with a nucleus in the emulsion gave rise to a shower of daughter particles; many of the daughter particles also collided, creating secondary showers, and so on. After the emulsion was developed the experimenters measured the distance between collisions for each daughter particle. The average distance between collisions, or the mean free path of a particle, is an estimate of the strength of its interactions with other particles. A shorter mean free path indicates that a particle is more likely to collide with a nucleus; in many cases the probability of collision is related to the size of the particle. The investigators found that about 6 percent of the particles created in the emulsion had an unusually short mean free path. No nuclear process known could account for the results.

The demon deuteron invoked by Fredriksson and Jändel to explain these results is not larger than an ordinary deuteron; its short mean free path is caused by its inherently strong interaction with other nuclei. The interaction is much like the one between chemically active atoms. When electrons in an atom are less tightly bound to the atomic nucleus, the electromagnetic force that binds the electrons to the nucleus "leaks" outside the neutral atom. The atom is therefore more likely to encounter other atoms and form chemical bonds with them. Similarly, in the demon deuteron the three pairs of quarks must occupy an excited energy state and so are less tightly bound to one another than the neutron and the proton are in an ordinary deuteron. Hence the force that binds them together, which is called the color force, can leak away from the immediate vicinity of the quark pairs and make itself felt over a much wider area.

Fredriksson and Jändel have proposed several additional ways the demon deuteron might be observed. Experimenters at the Lawrence Berkeley Laboratory and at the Joint Institute

for Nuclear Research at Dubna in the U.S.S.R. plan further searches.

### *Attributes of Excellence*

What makes a good physician? Three decades of research has led a group of investigators to conclude that excellent performance in medical school (as measured by grades and by scores on examinations) has no correlation with excellent performance in medical practice. It is a conclusion, they say, that surprises "practically everyone involved in medical education or in any other kind of education." Yet it fails to surprise practicing physicians, each of whom has been first a medical student and then "quite a different person afterward in professional practice."

The group of investigators, which currently includes Philip B. Price of the College of Medicine of the University of Utah, Calvin W. Taylor of the department of psychology at Utah and David E. Nelson of the Utah State Board of Education, began work in the 1950's on the relation between the performance of students in medical school and their earlier performance on the tests they had taken as applicants to medical school. The aim was to find some means of judging the process by which medical students are chosen and educated. They found that what they needed were criteria by which to evaluate the performance of the physicians who ultimately emerged from the educational process.

In a study done in the 1970's the investigators solicited the opinions of 372 physicians as well as the opinions of medical administrators and educators and some 1,600 people who could be taken to represent American society as recipients of medical care. The intent was to identify the attributes of good physicians. The result was a list of 87 desirable attributes and 29 undesirable ones. In a condensed version of the list that the investigators have recently prepared for the evaluation of surgeons, the desirable qualities are gathered under five headings. "Dependability and Commitment" includes items such as stamina, availability when needed and willingness to assume responsibility for patient care. "Problem Solving, Thought Processes and Clinical Skills" includes receptivity to new ideas and willingness to consult with others about complicated problems. "Qualities of a Superior Physician" includes emotional stability in critical situations, the eliciting of trust in patients and their families and frankness with patients. The other two headings are "Fund of Knowledge" and "Surgical Skills."

When the list of attributes had been compiled, the investigators undertook a pilot project in which 10 practicing physicians were evaluated. The evaluators included nurses, administrators, physicians and patients with whom the subject was associated as well as observer physicians, who typically spent a day with the subject. The subjects also evaluated themselves. A larger study followed. Here subjects were selected at random from among the physicians practicing in Utah. Only 8 percent of the physicians who were approached refused to participate. Ultimately 34 surgeons, 35 pediatricians, 34 internists, 31 rural general practitioners and 33 urban general practitioners were evaluated. Their rating as physicians showed no relation to the grades they got as medical students.

The investigators suspect that the attributes of excellence they have identified are evident (and can therefore be evaluated) in the later stages of the training of a physician: in residency, for example. At earlier stages, however, and particularly in premedical education, there is little sign of them. The investigators note that "only a very small amount of data can now emerge from premedical education . . . to enable admissions committees [in medical schools to select] those who will become excellent physicians rather than . . . those who will become excellent medical students." Nevertheless, the investigators suggest that the stages of a medical education could all be aimed not only "at the scholarly pursuit of scientific knowledge and acquisition of technical skills" but also at the "stimulation and cultivation" of the attributes of excellence in a physician. They further suggest that their findings "have implications for the selection and education of people in all other complex professions."

## *Token Impressions*

Evidence has been gathering lately that the earliest precursors of written language served the needs not of the poet or the historian but rather those of the accountant. Denise Schmandt-Besserat of the University of Texas at Austin has shown that fired clay "tokens" found at Neolithic and early Bronze Age sites in the Middle East constitute a system of symbols representing various quantities and commodities; furthermore, she has established a connection between the tokens and later Sumerian pictographs and ideographs. Now an intermediate stage in the transition from tokens to writing is becoming clear.

Since the turn of the century archaeologists working at 10 widely separated sites in Syria, Iraq and Iran have unearthed a number of ancient tablets bearing impressions of various depths and shapes. Of the 185 tablets discovered up to 1977 a few are made of plaster but most are made of unfired clay. They come from strata dating to about the end of the fourth millennium B.C. Until recently the impressions were interpreted as numerical notations, and so the curious objects came to be known as numerical tablets. They were shelved with little additional study.

In 1974 interest in the impressed tablets revived when the Russian epigrapher A. A. Vaiman proposed that the impressions, which are of 19 different shapes, are not numerical notations but records of measurements such as pecks and bushels of grain, rods and acres of land and flocks and herds of animals.

Vaiman's suggestion inspired Jöran Friberg of the Chalmers University of Technology and the University of Göteborg to undertake a mathematical analysis of the tablets. Friberg's study, published in 1978 and 1979, confirmed Vaiman's findings: Friberg proposed that six of the signs represent measures of grain and four others are land measurements. To two of the grain measures he assigned the historical Sumerian names "ban" and "bariga."

At the same time Schmandt-Besserat was puzzled. The impressions on the numerical tablets matched the shapes of several of the tokens she had been studying. She found, however, that when she applied the numerical values assigned to various written Sumerian signs by the German scholar Adam Falkenstein, the principal student of these early inscriptions, the tablet texts made no sense. She decided to substitute Vaiman's and Friberg's units of measurement for Falkenstein's numerals. The texts of many of the impressed tablets, she discovered, were thereby transformed into sensible accountancy records of a kind suited to temple administration.

Writing in the journal *Visible Language,* Schmandt-Besserat proposes that the impressed tablets are mainly reckonings of foodstuffs, grain and animals. The quantities recorded are strikingly small. For example, the amount of grain is most commonly between one bushel and four bushels and is seldom more than 10. Because the reckonings are similar in form and content to later inscribed temple tablets, Schmandt-Besserat suggests that the impressed tablets also record such transactions as the receipt of temple offerings, the distribution of food to temple workers and their dependents and the release of seed for planting in temple fields. With the new interpretation of the impressed tablets it is possible to perceive a sequence of events in the evolution of early Sumerian writing systems. First came the invention of clay tokens, which represented particular quantities of goods. Later the tokens were stored in clay "envelopes," whose contents were recorded by making an impression of each token in the outer surface. In the tablets the impression left by a token was made to stand for the token itself. Ultimately the tokens were eliminated and symbols were shaped in clay with a stylus.

# ARE JAPANESE EXECUTIVES BETTER PEOPLE MANAGERS THAN AMERICAN EXECUTIVES?

*Are employee exercise programs, company songs and lifetime employment the best ways to increase productivity?*

# The Genetics of Antibody Diversity

*Segments of DNA and RNA are shuffled and joined in various ways as cells of the immune system develop. The combinatorial process can generate information specifying billions of different antibodies*

by Philip Leder

The immune system of a vertebrate animal has a virtually unlimited capacity to generate different antibodies, which recognize and bind to many millions of potential antigens, or "nonself" molecules. This implies that many millions of species of antibody molecules are synthesized by the cells of the immune system. An antibody is an assembly of protein chains, and the structure of a protein chain is specified by a unit of genetic information: a gene. Hence it would seem there must be many millions of antibody genes. Yet the genome, or total genetic complement, of a mammal amounts to perhaps a million genes, and only a small fraction of them can specify antibodies. The paradox of a limited number of genes and an apparently limitless capacity to generate different antibodies has been a major puzzle for immunologists and geneticists for more than two decades. The solution of the puzzle is now emerging. It represents both a dramatic confluence of the two disciplines and an early triumph of some remarkable new techniques of molecular genetics.

In essence the answer is that the genes ultimately specifying the structure of each antibody are not present as such in germ cells (the male sperm and the female egg) or in the cells of the early embryo. Rather than harboring a set of complete and active antibody genes, these cells contain bits and pieces of the genes: a kit of components. The components are shuffled in the cells of the immune system called *B* lymphocytes as those cells develop and mature. The shuffling can lead to a different result in each of millions of lines of cells. Individual mutations amplify the diversity. The result is that in the mature descendants of each line a unique gene is assembled, whose information is expressed in the form of a unique antibody.

## Proteins and Genes

Antibodies, like other proteins, are made up of the subunits called amino acids. There are 20 kinds of amino acid, which can be linked together in any combination to form a protein chain. The amino acid composition of a chain and the sequence in which the amino acids are arrayed along the chain determine how the chain folds in three dimensions and perhaps combines with other chains. Substituting one amino acid for another or altering the sequence changes the protein's properties.

The amino acid composition and sequence of a protein chain are prescribed by a gene, which is a segment of another kind of chain: a molecule of the nucleic acid DNA, the hereditary material of the cell. The subunits of DNA are four nucleotides, each of which is characterized by one of four chemical bases: adenine (A), guanine (G), thymine (T) and cytosine (C). The nucleotides are assembled to form a strand of DNA, and the sequence of their bases along the strand defines the information carried by the gene. The information is deciphered by means of the genetic code, whose code words (called codons) are triplets of bases: CTG, say, or AGC. In general each codon specifies an amino acid, and a long series of codons supplies instructions for assembling an entire protein chain consisting of hundreds of amino acids. The information in DNA is not translated directly into a protein, however. First it is transcribed into a single strand of the similar nucleic acid RNA. The new informational molecule, which is called messenger RNA, is translated into protein.

In higher organisms the genetic information for most proteins is not arranged in a continuous sequence of DNA codons. Instead most genes are split: patches of coding sequences are separated by noncoding intervening sequences. A split gene is first transcribed, noncoding segments and all, into RNA. Then the primary transcript is spliced: the intervening sequences are eliminated and the coding sequences are joined to form a coherent messenger RNA.

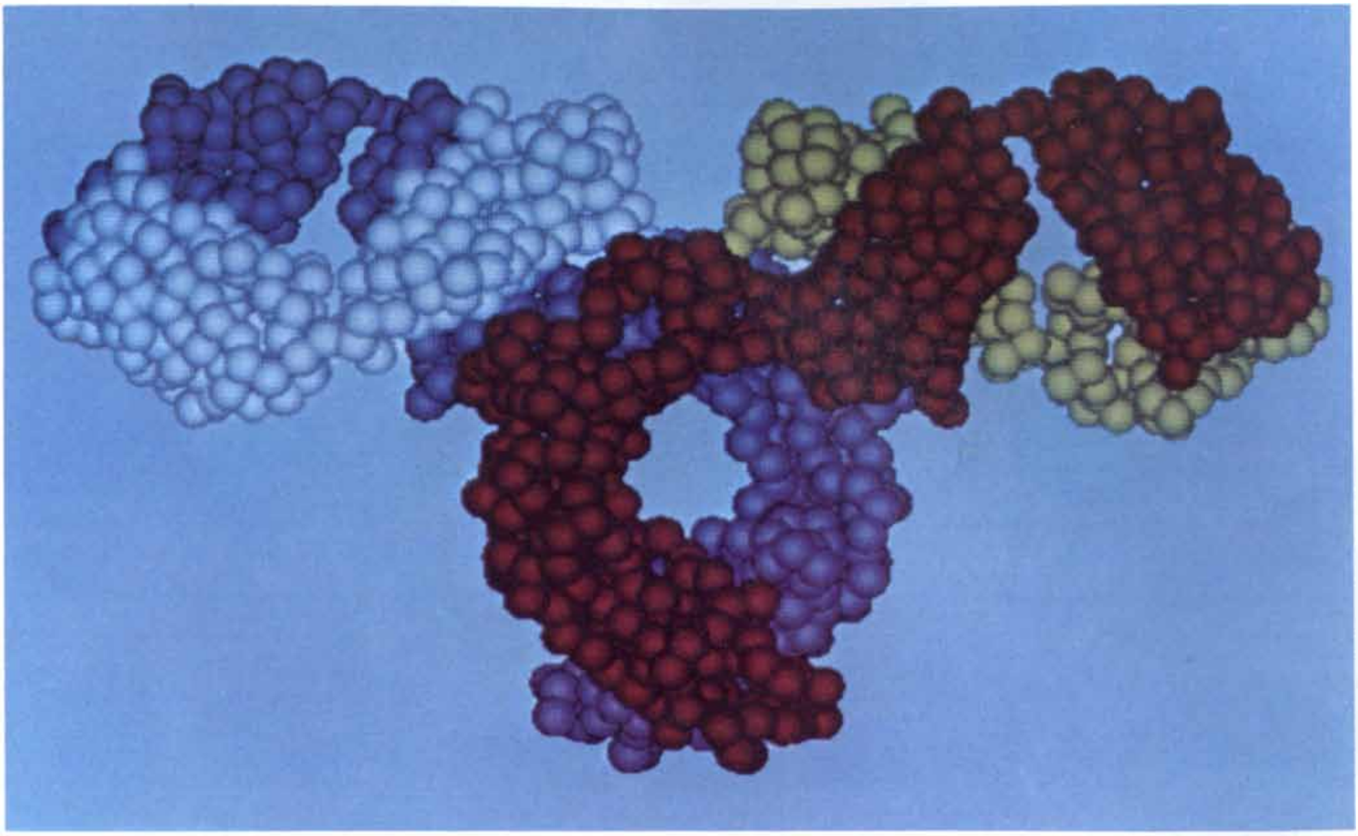
Immunoglobulins, the antibody molecules, have structural features that reflect their function. An antibody molecule is made up of two kinds of related protein chains, designated light and heavy. When the amino acid sequences of light chains from various antibodies were compared almost two decades ago by Norbert Hilschmann, who was then working at the Rockefeller Institute for Medical Research, he found that the chains have a peculiar property. The sequences of chains from different antibodies are different from one another, but the differences are confined to the first half of each chain. The remainder of the chain has essentially the same sequence in all antibodies of a given type.

## A Bifunctional Molecule

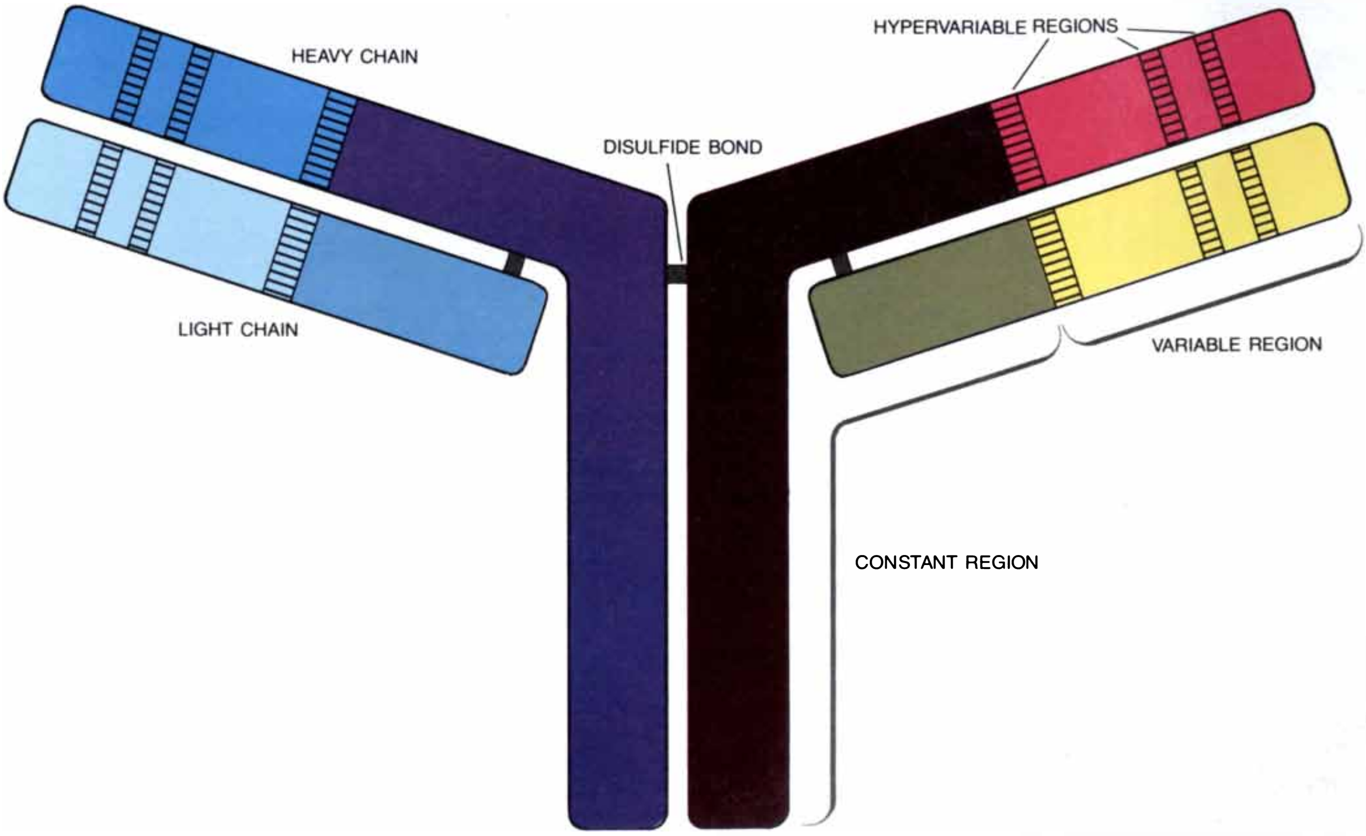
The presence of both variation and constancy in a single protein molecule turned out to have great functional significance. Indeed, an antibody is a bifunctional molecule. Each chain has a variable region (about half of a light chain and about one-fourth of a heavy chain) and a constant region. It is the variable regions of the chains that fold up in space to form an antibody-antigen combining site: the site that binds to the particular antigen against which the antibody is directed. Changing the amino acid sequence in the variable region changes the combining site's chemical structure and thereby changes the affinity of the antibody for an antigen, much as altering a notch on the serrated part of a key makes it fit a different lock.

The constant region of an antibody's light and heavy chains is analogous to the handle of a key, which is identical from one key to another of a given make and type and serves a function common to all keys. The constant region of an antibody molecule of a particular type serves the same function in every molecule of that type. For example, there are two types of light chain in most vertebrate animals, kappa and lambda; every antibody molecule must have light chains of one type or the other. In any species the constant region of each kappa or lambda light chain is identical with the constant region of other chains of the same type.

In addition each antibody molecule has one of five heavy-chain types: mu,



ANTIBODY MOLECULE is an assembly of four protein chains, which are folded and interconnected to form a split *T*. There are two identical heavy chains (*red and dark blue*) and two identical light chains (*yellow and light blue*). In this model, generated with the aid of a computer by Richard J. Feldmann of National Institutes of Health, the spheres represent amino acids, the subunits of the protein chains.




BIFUNCTIONAL NATURE of an antibody is reflected in its structure, as is seen in this schematic diagram of the molecule modeled at the top of the page. Each protein chain has a variable region and a constant region. In the variable regions the sequence of amino acids is different in each antibody; the constant regions are the same in every antibody of a given type. The variable regions recognize and bind to a specific antigen; the constant regions thereupon carry out some immunologic task. The chains are folded so that their hypervariable regions, where the amino acid sequence is particularly volatile, come together to form a highly specific antigen-combining site.


## 1982 Buick Century.

This could be just another year for you. Or, it could be the year of the Century. Especially once you consider the many pleasures our new 1982 Buick Century can bring you season after season.

To begin with, Century is the most aerodynamic Buick ever. In fact, compared to last year, Century uses less fuel to overcome air resistance caused by driving through the blustering winds of fall. And, in winter, its front-wheel drive makes it an ideal ally against the snow.

With a standard 4-cylinder engine—or an available V-6 gasoline or V-6 diesel—Century will splash through spring with impressive mileage. And, its remarkable packaging efficiency means lots of space with less bulk. Enough, in fact, to carry a family of five—with luggage—comfortably through summer.

Considering the pleasure a new Century could bring all year, isn't it worth a trip to your Buick dealer's today?

| | EST. HWY. | EPA EST. MPG |
|---|---|---|
| 2.5 Liter L-4 | 40 | 25 |
| 3.0 Liter V-6 | 33 | 21 |
| 4.3 Liter V-6 Diesel* | 42 | 26 |

Use estimated MPG for comparison. Your mileage may differ depending on speed, distance, weather. Actual highway mileage lower. Estimates lower in California. Some Buicks are equipped with engines produced by other GM Divisions, subsidiaries, or affiliated companies worldwide. See your Buick dealer for details.

*Buick projection of 1982 EPA V-6 Diesel estimates. See your dealer for actual EPA estimates.

BUICK

Wouldn't you really rather have a Buick?

delta, gamma, epsilon or alpha. The heavy-chain type defines the class of the immunoglobulin as IgM, IgD, IgG, IgE or IgA. In secreted antibodies of the IgM class, for example, all the heavy chains have the same mu constant-region sequence and all the light chains have the same kappa or lambda constant-region sequence. Their variable regions, on the other hand, differ from one antibody to the next, reflecting their different antigenic specificities.

The constant region of the heavy chains determines the effector function of an antibody, or how it carries out its immunologic task in the body. Consider an antibody whose variable region is specific for an antigen found on ragweed pollen. If the heavy chain is of the delta type, the IgD antibody it forms remains associated with the surface of the cell that makes it. If the gamma heavy chain is present, the resulting IgG antibody is likely to circulate in the blood. If the epsilon chain is present, the antibody (IgE) may bind to the surface of a specialized cell that releases histamine, giving rise to the symptoms of hay fever or asthma when the antibody interacts with the ragweed antigen. All the antibodies are specific for the same antigen, namely the ragweed one. Furthermore, the same effector functions are found in antibodies directed against other antigens; the effector function is independent of the variable region.

## The Dreyer-Bennett Hypothesis

It was the structural features of antibodies that offered the first clues to the genetic source of their diversity. William J. Dreyer and J. Claude Bennett, who were working at the California Institute of Technology in the mid-1960's, recognized that a million different antibodies could be generated by combining 1,000 different light chains with 1,000 different heavy chains. Nevertheless, Dreyer and Bennett wondered how the genetic information for this diversity of proteins could be organized. Above all, what organization could account for the strange variable-and-constant structure of each chain?

Consider just the problem of one type of light chain. There might conceivably be 1,000 genes, each gene specifying one of 1,000 light chains. In that case, however, some mechanism must have operated in the course of evolution to preserve unchanged the sequence of the constant-region half of each of the 1,000 genes while the other half of each gene (the half encoding the variable region) was allowed to mutate widely. Such a mechanism seemed unlikely because there is no obvious biological reason for any given constant-region amino acid sequence to have been conserved. Comparisons of immunoglobulins in different individuals and species indicate there is little evolutionary pressure to maintain absolute identity among constant-region sequences; a change of a few amino acids in the constant region seems to have no ill effect.

Dreyer and Bennett made a radical proposal. Instead of assuming that the genetic information for an antibody light chain is specified by a continuous array of codons, they proposed that the light chain is encoded in two discontinuous stretches of DNA, one for the variable region and the other for the constant region. Moreover, they proposed that there are several hundred or several thousand separately encoded variable-region genes in the DNA of germ cells but only one constant-region gene.

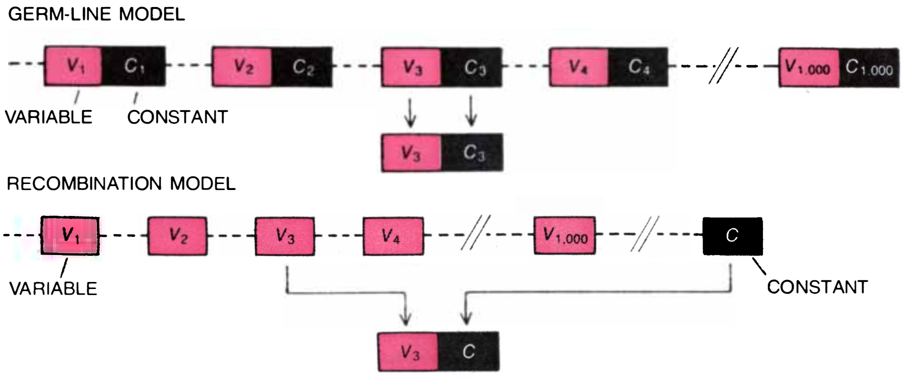
By postulating a single constant-region gene, the Dreyer-Bennett proposal showed how essentially the same sequence might be conserved to appear in every constant region of a given type in a given species. If there is only one constant-region gene, any mutation in it must immediately alter the amino acid sequence of every light chain. Also implicit in the proposal was the notion that the information in the separate genetic elements must somehow come together to form a contiguous and coherent genetic message and then a single protein chain. The proposal of Dreyer and Bennett initially attracted considerable criticism; it called for split genes and for mechanisms to join them, both of which were then quite without precedent. Yet the idea proved to be essentially correct.

## Early Experiments

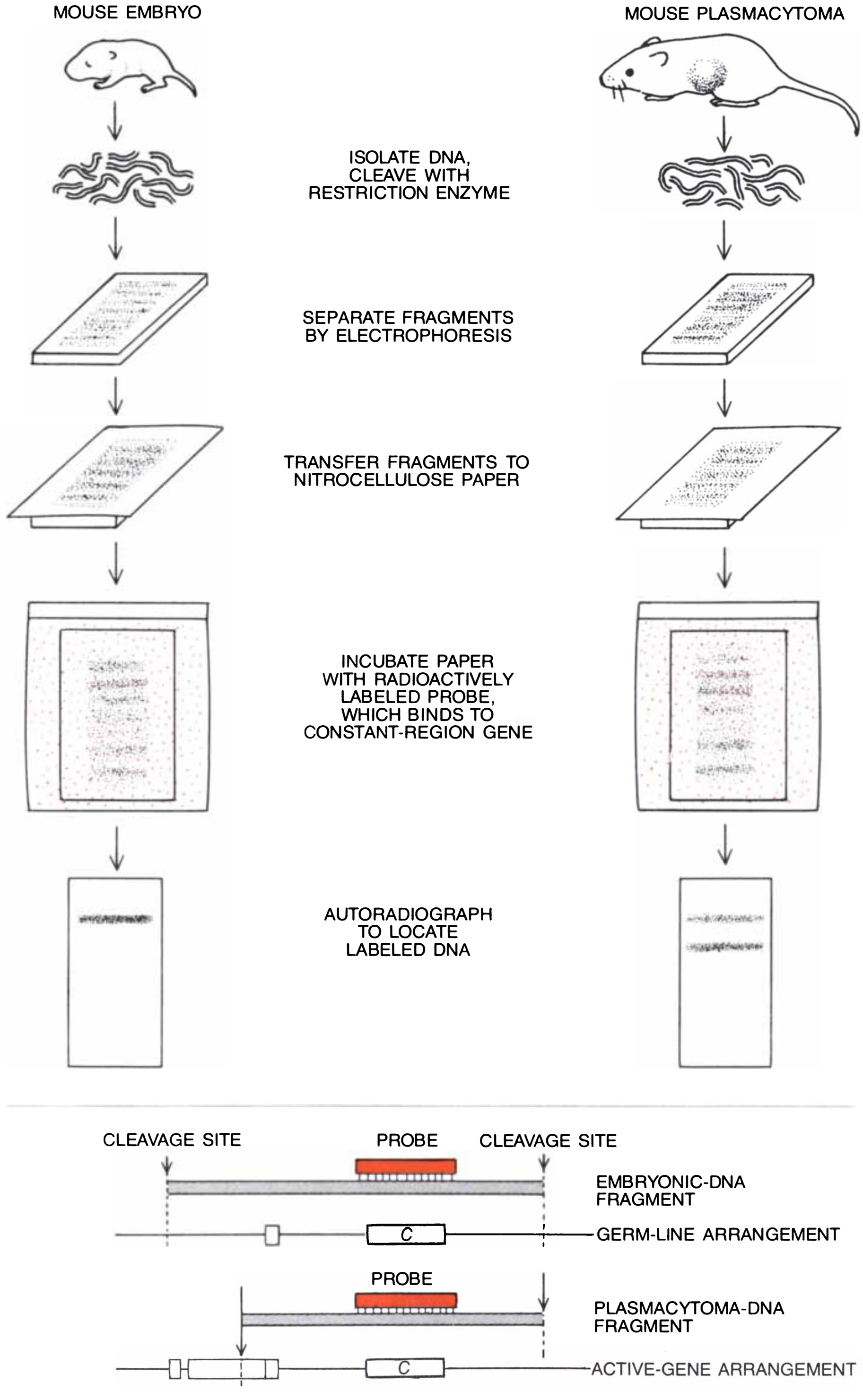
In 1971 I set to work with my colleague David C. Swan at the National Institute of Child Health and Human Development to test the Dreyer-Bennett hypothesis. Our strategy was to detect and isolate the initial product of an antibody gene: its messenger RNA. A strand of DNA artificially copied from that messenger (by means of the enzyme reverse transcriptase) would serve as a probe with which to detect the antibody genes in embryonic cells and estimate how many such genes are present. This would enable us to test the central prediction of the Dreyer-Bennett hypothesis: that there is only one constant-region gene, or at the most a few, as opposed to the many genes required by the more straightforward genetic models.

We used the technique of hybridization kinetics. Two stretches of nucleic acid with complementary nucleotide sequences are able to hybridize, or bind to each other. If one of the sequences has been labeled with atoms of a radioactive isotope, the hybrid molecules can be identified by their radioactivity. The speed with which a radioactive DNA probe finds and hybridizes with any complementary DNA molecules (the hybridization kinetics) is an indirect but effective measure of the number of such complementary molecules in the preparation. We measured the rate at which our probe (DNA copied from the RNA specifying the constant region of the mouse light chain) hybridized with DNA representing the entire genome of the mouse embryo. With the help of Tasuku Honjo, who had joined us from Kyoto University, we got results indicating clearly that there are very few copies of the light-chain constant-region gene, perhaps no more than two in each cell. Similar results were soon obtained in a number of other laboratories. Clearly the Dreyer-Bennett hypothesis had to be taken seriously.

If there are only a few copies of the constant-region gene and many sepa-




**ARRANGEMENT OF THE GENES** that encode an antibody protein chain has been described by two models. The germ-line model assumed that perhaps 1,000 chains are encoded in germ cells (eggs and sperm) by 1,000 genes, each gene having a DNA sequence specifying the variable region (*V*) of the chain and a sequence for the constant region (*C*). One of the two-region genes would be expressed in each antibody-producing lymphocyte. By combining 1,000 heavy-chain genes with 1,000 light-chain genes a million distinct antibody molecules could be generated. It is hard to understand, however, how the 1,000 identical constant-region sequences could have been conserved in the course of evolution while the variable-region sequences were allowed to develop differences. In the recombination model put forward in 1965 by William J. Dreyer and J. Claude Bennett the same number of protein chains are encoded in germ cells by 1,000 alternative variable-region genes at some distance from a single gene for the constant region. One of the variable-region genes would be recombined with the single *C* gene and would be expressed; as a result every chain would necessarily have the same constant region.




**GENE SHUFFLING in the course of development is demonstrated by comparing the DNA fragments on which a constant-region gene is found in the mouse embryo and in a plasmacytoma: a tumor composed of one line of antibody-producing cells. The DNA is digested with a restriction enzyme, which cleaves DNA at specific sites defined by a certain nucleotide sequence. The resulting fragments are separated according to size by electrophoresis on an agarose gel and then are "blotted" onto nitrocellulose paper. Fragments that incorporate the constant-region gene (*C*) are identified by incubating the paper with a radioactively labeled probe (*color*), which in current versions of the experiment would be a cloned DNA copy of the messenger RNA specifying the light chain's constant region. The probe hybridizes with (binds to) any embryonic or plasmacytoma DNA having a largely complementary nucleotide sequence. The hybrid DNA is identified by autoradiography, in which the radiation from the labeled molecules exposes a photographic film. The embryonic DNA yields one radioactive band, which corresponds to a single fragment carrying the *C* gene in its germ-line configuration. The plasmacytoma DNA yields two bands. One band represents the germ-line configuration, which is maintained by one of the two alleles, or versions of the gene, in the antibody-producing cell. The other band corresponds to a fragment that moves farther on the gel and is therefore shorter. It represents the other allele, which has been rearranged and carries the *C* sequence in the configuration of an active antibody gene. As is shown at the bottom of the illustration, the rearrangement eliminates one of original cleavage sites and brings in a new site, closer to *C* gene.**

rately encoded variable-region genes, some mechanism must operate to bring the information together in a coherent sequence. The most economical way to bring this about would be at the level of the gene, that is, to join two sequences of DNA that are separate in an embryonic cell to form a single active sequence in the nucleus of a mature, antibody-producing lymphocyte. Such a rearrangement of the DNA in the course of the differentiation and development of somatic, or body, cells is referred to as somatic recombination. The discovery of the enzymes called restriction endonucleases, which cleave DNA at specific sites, paved the way for an important experiment testing the notion of somatic recombination.

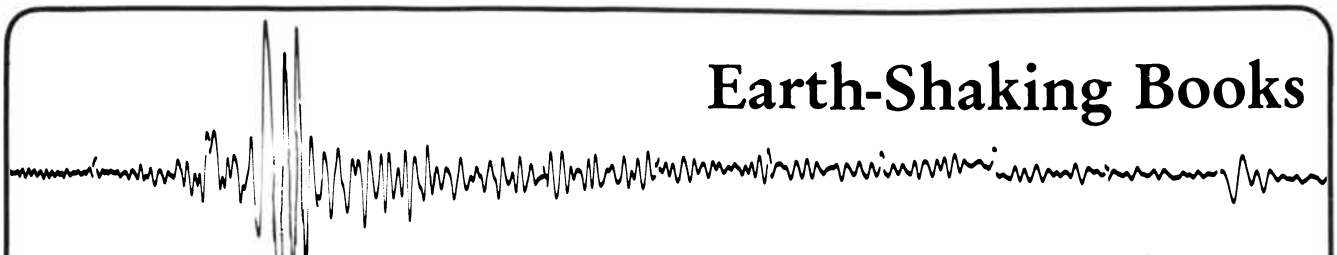
Susumu Tonegawa and Nobumichi Hozumi of the Basel Institute for Immunology compared DNA from the mouse embryo with DNA from a plasmacytoma: a tumor composed of plasma cells, or mature *B* lymphocytes, that produce antibody molecules of a single type and specificity. Cells from these tumors, most of which have been prepared by Michael Potter of the National Cancer Institute, serve to provide the investigator with a large quantity of a particular antibody and of the DNA and RNA that encode it; the cells have been invaluable in many lines of immunologic research.

Tonegawa and Hozumi reasoned that if the DNA surrounding the antibody genes had undergone somatic recombination in the tumor's antibody-producing cells, it should be arranged differently from the DNA of embryonic cells. The difference would be reflected in the spacing of the sites that are cut by a given restriction endonuclease. For example, if the constant-region gene in embryonic DNA is surrounded by two restriction sites (sites cleaved by a particular endonuclease) 5,000 nucleotides apart, after the cleavage the gene is found embedded in a fragment of DNA 5,000 nucleotides long. If somatic recombination eliminates one of the sites and brings in a new one, the plasmacytoma DNA may yield a constant-region fragment either longer or shorter than 5,000 nucleotides. Tonegawa and Hozumi were able to show that the arrangement of light-chain genes is different in embryonic cells and antibody-producing cells. The activation of the genes in the course of development is accompanied by their somatic recombination: the genes are shuffled.

## Cloning

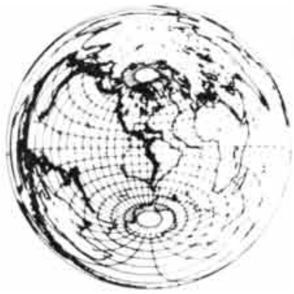
The mid-1970's was a period of extraordinary developments in molecular genetics, accompanied by some controversy and misunderstanding. A major advance, in 1973, was the first successful application of new recombinant-DNA techniques to insert foreign DNA into a bacterium or a bacterial virus and there-

# Earth-Shaking Books



*from* **William Kaufmann, Inc.**
**95 First Street, Los Altos, CA 94022**

**EARTHQUAKE INTERPRETATIONS**
**A Manual for Reading Seismograms**
Ruth B. Simon
This thorough primer for earthquake interpretation includes reproductions and interpretations of dozens of earthquake records. Clear instructions make this book an important tool for exploring the internal structure of the earth.
$9.95 (paper), 160 pp. A



**EARTH AND ITS INHABITANTS**
**Readings from *American Scientist***
Edited by Brian J. Skinner
Each of the inexpensive, colorful, paperbound books in the series *Earth and Its Inhabitants* contains full text and illustrations from *American Scientist* articles.
$8.95, *Climates Past and Present* B
$8.95, *Earth's Energy and Mineral Resources* C
$8.95, *Earth's History, Structure and Materials* D
$9.95, *Paleontology and Paleoenvironments* E
$11.95, *The Solar System and Its Strange Objects* F
$9.95, *Use and Misuse of Earth's Surface* G

**OUR MODERN STONE AGE**
Robert Bates and Julie Jackson
This book describes about two dozen rocks and minerals, explaining how they occur, how they are extracted from the ground, and what happens to them in the labyrinth of industry. Environmental problems associated with mineral production are also addressed.
$18.95 (cloth), 150 pp. H

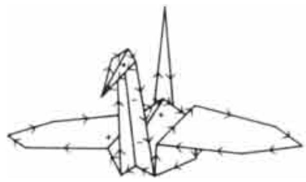
**THE ART IN THE GREAT TEMPLE**
**Mexico-Tenochtitlan**
Ruben Nuno and Fernando Robles
This handsome, well-designed book contains full-color illustrations of one of Mexico's most important archaeological finds.
$65.00 (cloth), 189 pp. I

**THE HUNTING OF THE SNARK**
Lewis Carroll. Illustrated by Henry Holiday. Annotated by Martin Gardner. Essay by Charles Mitchell. Bibliography by Selwyn H. Goodacre. Edited by James Tanis and John Dooley
A beautiful centennial production of Carroll's classic "nonsense poem," this literary classic is available in a magnificent morocco-bound and signed Subscriber's Edition with a separate portfolio of framable plates; a limited Collector's Edition; or a cloth trade edition.
$395.00 (Subscriber's Edition) J
$60.00 (Collector's Edition) K
$18.95 (cloth) L

**THE QUESTION OF ANIMAL AWARENESS**
**Second Edition**
Donald Griffin
Is it language that sets man apart from other animals? How do we know? If scientists accept biological evolution in animals, why do they shy away from the concept of continuity in animal mental experiences? Is that logical? Do animals have intentions or mental images? How can we find out?
$8.95 (paper), 221 pp. M

**READINGS IN ARTIFICIAL INTELLIGENCE**
Edited by Bonnie Lynn Webber and Nils J. Nilsson
Thirty-one classic papers on search, deduction, planning, expert systems, and other topics have been gathered in this convenient volume for class and reference use. Hard-to-find papers by AI experts are thus made readily accessible for present and future specialists.
$25.00 (paper), 557 pp. N

**PRINCIPLES OF ARTIFICIAL INTELLIGENCE**
Nils J. Nilsson
*Principles of Artificial Intelligence* lucidly explains the core ideas of artificial intelligence—the fundamental principles of this rapidly-expanding computer science. Its goal is to fill a gap between theory and practice for readers versed in the fundamentals of computer science.
$30.00 (cloth), 492 pp. O



**THE HANDBOOK OF ARTIFICIAL INTELLIGENCE**
Edited by Avron Barr, Paul R. Cohen, and Edward A. Feigenbaum
Copublished with HeurisTech Press
*Volumes II and III available Summer 1982*
*The Handbook of Artificial Intelligence,* written in plain English for people without a background in the field, contains over 200 short articles covering all of the important ideas and techniques developed in artificial intelligence over the last twenty-five years. *The Handbook* is a reference work, a textbook, a guide to programming techniques and to the extensive literature of the field, and a book for intellectual browsing.
$95.00 (Three-Volume Set) P
$30.00 (Volume I), 409 pp. Q
$35.00 (Volume II), 450 pp. R
$45.00 (Volume III), 650 pp. S

**WILLIAM KAUFMANN, INC.**
**Dept. S.A., 95 First Street, Los Altos, CA 94022**
Please send me the books whose letters are circled below. If not satisfied, I may return the order within 30 days for a complete refund.

☐ Enclosed is full payment plus $1.50 shipping charge. (California residents please add sales tax.)

☐ Please send me your free catalog.

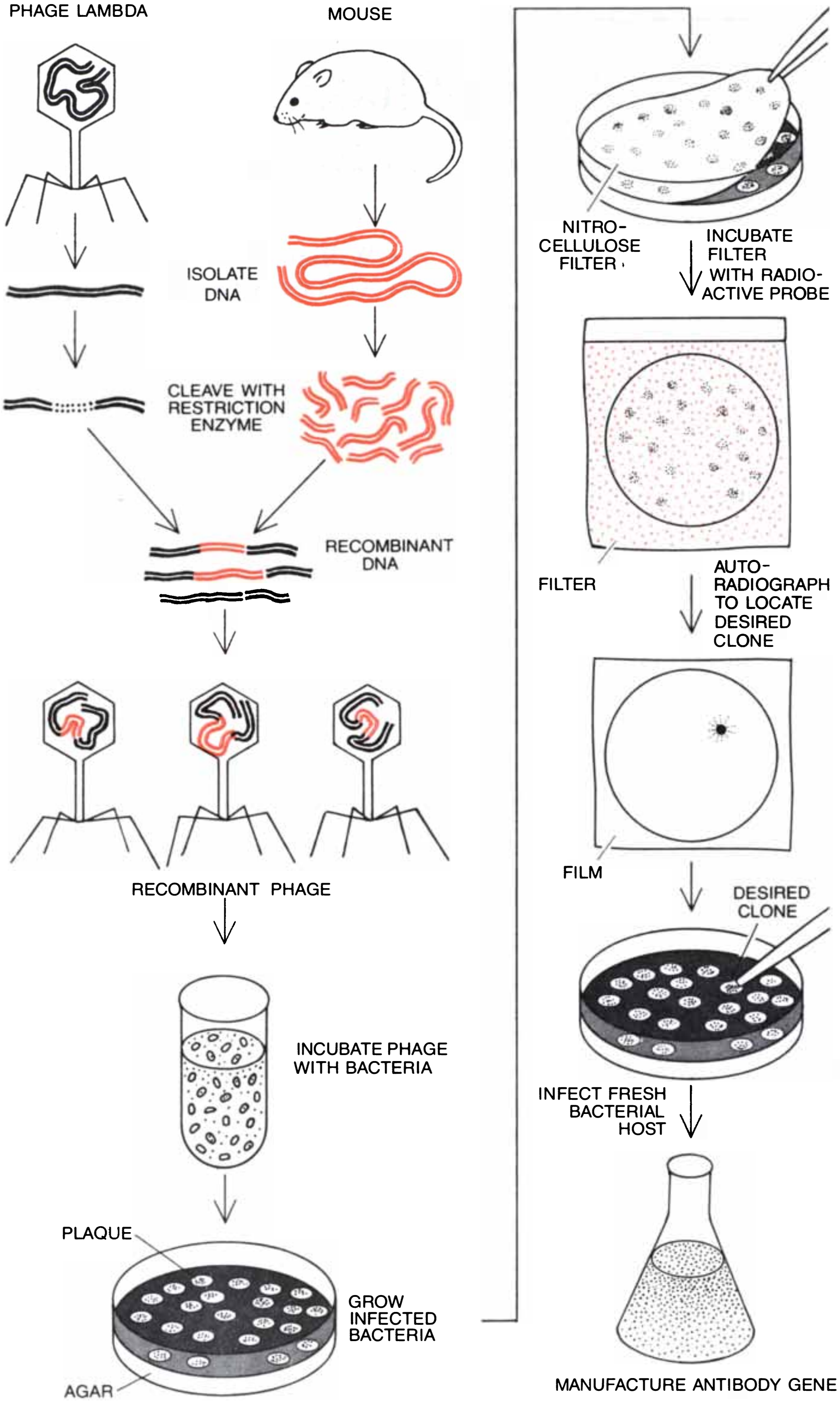
A B C D E F G H I J K L M N O P Q R S

Name ______________________

Address ______________________

City ______________________

State ______________ Zip ________



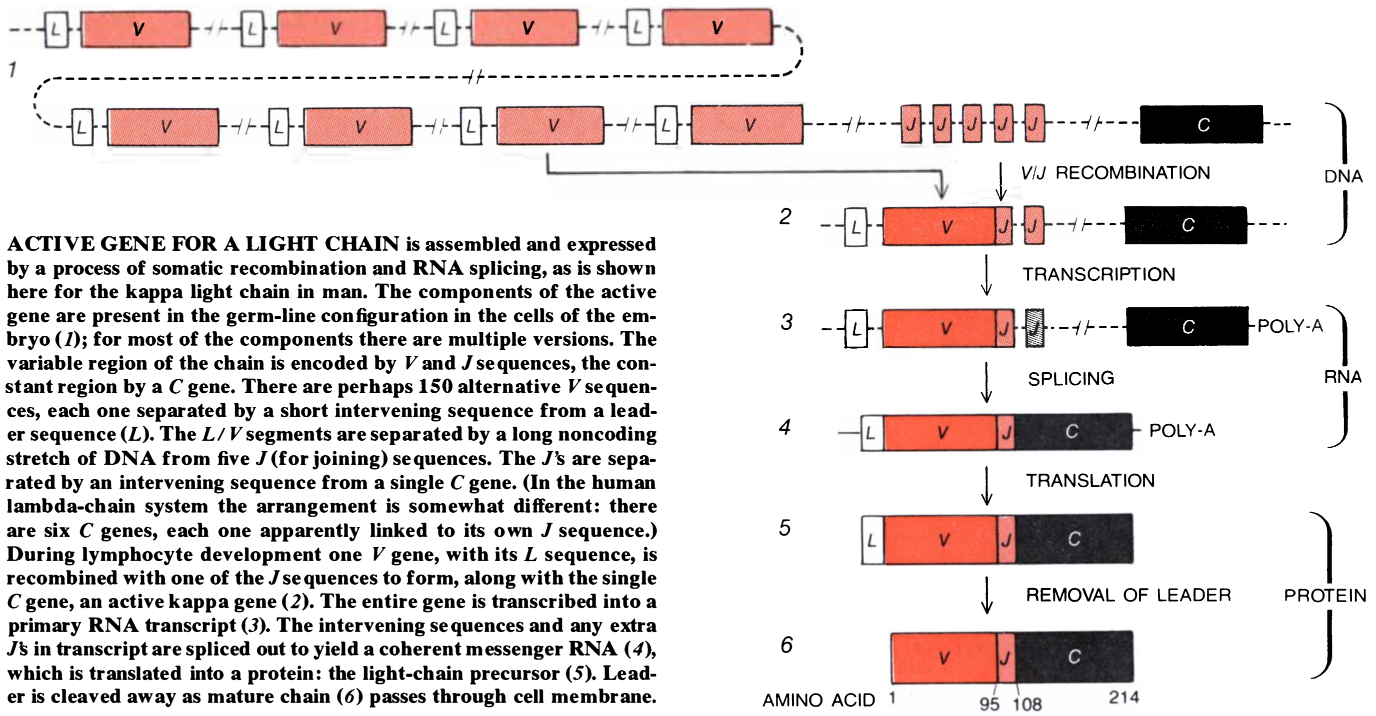

**ANTIBODY GENE IS CLONED by inserting fragments of mouse DNA into a bacterial virus (phage lambda), isolating a virus clone whose DNA includes the gene and growing that virus in quantity to produce large amounts of the single gene. First the phage DNA is cleaved with a restriction enzyme, which cuts the DNA into three fragments. The middle fragment is not needed for viral growth. It can be replaced with a fragment of mouse DNA that has been cleaved with the same restriction enzyme, so that the phage and the mouse fragments have complementary terminals ("sticky ends") and bind to one another. The recombinant DNA's are packaged in new phage particles, which are incubated with *Escherichia coli,* the bacteria they infect. The infected bacteria are layered onto an agar culture medium. Individual phages proliferate, killing their host and infecting neighboring cells, thus leaving clear areas called plaques in the bacterial layer; each plaque represents a phage clone. The plaque pattern is transferred to a nitrocellulose filter and the phage protein is dissolved, leaving the recombinant DNA. The filter is incubated with a radioactively labeled probe (*color*): a DNA copy of the messenger RNA representing the desired gene. The probe hybridizes with any recombinant DNA incorporating a matching mouse-DNA sequence, and the position of the clone having that DNA is revealed by autoradiography. Now desired clone can be selected from culture medium and transferred to fresh bacterial host, so that pure antibody gene can be manufactured.**

by clone a single gene in quantity. It was immediately clear to those of us working with more cumbersome genetic techniques that gene cloning would enable us to isolate antibody genes and determine their structure in a direct way. There was uncertainty, however, about possible hazards of the procedures; the safety of the methods has since been established, but at that time the National Institutes of Health promulgated cautious guidelines for recombinant-DNA experiments. Among the requirements were some that affected vector organisms: the bacteria or bacterial viruses in which foreign DNA is cloned. The vector had to be genetically crippled to reduce its chance of survival outside the laboratory by a factor of 100 million.

My colleagues David C. Tiemeier, Lynn Enquist, Nathan Sternberg, Robert A. Weisberg and I decided to adapt the bacteriophage, or bacterial virus, called phage lambda for this purpose. A version of the phage engineered by Ronald W. Davis of the Stanford University School of Medicine could be genetically rigged so that it would grow only if it had picked up a piece of foreign DNA. We introduced a series of crippling mutations into the phage vector, carried out appropriate survival tests and ultimately received the necessary approval. Soon Shirley M. Tilghman, Tiemeier and I were able to report the first successful cloning in a bacteriophage of a mammalian gene: a mouse gene for globin (the protein of the hemoglobin molecule).

In response to a request from Tonegawa's laboratory we sent a supply of our disabled phage to Switzerland, where the cloning strategy was exploited by Tonegawa and his colleagues to isolate the mouse genes for the lambda light chain in both embryonic and antibody-producing cells. They discovered that the constant-region gene and the variable-region genes are indeed encoded far apart in the DNA of cells that do not synthesize antibodies. In an appropriate antibody-producing plasmacytoma the two genes are much closer together. The rearrangement does not, however, bring the variable-region and the constant-region genes together to form a continuous sequence. Instead, the variable-region gene (*V*) is still about 1,500 nucleotides away from the constant-region gene (*C*); between them, and abutting the *V* gene, is a segment called the *J* (for joining) sequence, as I shall explain below.

Our own cloning efforts were directed toward the kappa light-chain genes of the mouse. We undertook to examine this system because more than 95 percent of mouse antibodies incorporate the kappa chain, making it the major source of mouse light-chain diversity. My colleagues Jonathan G. Seidman and Edward E. Max cloned both the embryonic and the active forms of the kappa variable-region and constant-region




**ACTIVE GENE FOR A LIGHT CHAIN is assembled and expressed by a process of somatic recombination and RNA splicing, as is shown here for the kappa light chain in man. The components of the active gene are present in the germ-line configuration in the cells of the embryo (*1*); for most of the components there are multiple versions. The variable region of the chain is encoded by *V* and *J* sequences, the constant region by a *C* gene. There are perhaps 150 alternative *V* sequences, each one separated by a short intervening sequence from a leader sequence (*L*). The *L*/*V* segments are separated by a long noncoding stretch of DNA from five *J* (for joining) sequences. The *J*'s are separated by an intervening sequence from a single *C* gene. (In the human lambda-chain system the arrangement is somewhat different: there are six *C* genes, each one apparently linked to its own *J* sequence.) During lymphocyte development one *V* gene, with its *L* sequence, is recombined with one of the *J* sequences to form, along with the single *C* gene, an active kappa gene (*2*). The entire gene is transcribed into a primary RNA transcript (*3*). The intervening sequences and any extra *J*'s in transcript are spliced out to yield a coherent messenger RNA (*4*), which is translated into a protein: the light-chain precursor (*5*). Leader is cleaved away as mature chain (*6*) passes through cell membrane.**

genes. Then, by means of the rapid DNA-sequencing technique developed by Alan Maxam and Walter Gilbert of Harvard University, they determined the nucleotide sequences of the genes.

## The Kappa Chain

A large number of *V* genes were identified in the embryonic DNA. They seem to fall into families, each family being made up of genes whose nucleotide sequences are closely related. Our own studies and those of Robert P. Perry of the Institute for Cancer Research in Fox Chase, Pa., indicate that as many as several hundred variable-region genes may be present in mouse-embryo DNA. Each of the *V* genes, whether kappa or lambda, retains certain structural features that appear to be of considerable significance. For example, each gene is divided into two discrete coding segments separated by a short intervening sequence. The first coding sequence specifies a hydrophobic (water-repellent) "leader," 17 to 20 amino acids long, that is thought to be important for the transport of the antibody molecule through the cell membrane. The leader is a part of the original protein product of the active light-chain gene but is cleaved away as the nascent antibody passes through the membrane.

The other coding region of the *V* gene specifies most of the variable region, but not all of it. The nonleader part of the *V* gene encodes only 95 of the 108 amino acids of the kappa chain's variable region. As Tonegawa's earlier findings had suggested was the case in the lambda system, we found that in the kappa DNA the remaining portion of the variable region is encoded by a sequence well "downstream" from the *V* gene, near the single constant-region gene, exactly at the site to which the *V* gene is joined to make an active immunoglobulin gene. This short sequence, the *J* gene, is repeated with slight but significant variations five times at intervals of about 300 nucleotides. (In the mouse lambda system further studies by Tonegawa, David Baltimore of the Massachusetts Institute of Technology and Ursula Storb of the University of Washington have shown that the arrangement is somewhat different. Instead of one constant-region gene there are four *C* genes, each with its own *J* gene. In the human lambda system, Philip A. Hieter, Gregory F. Hollis and I have found, there are six *C* genes.)

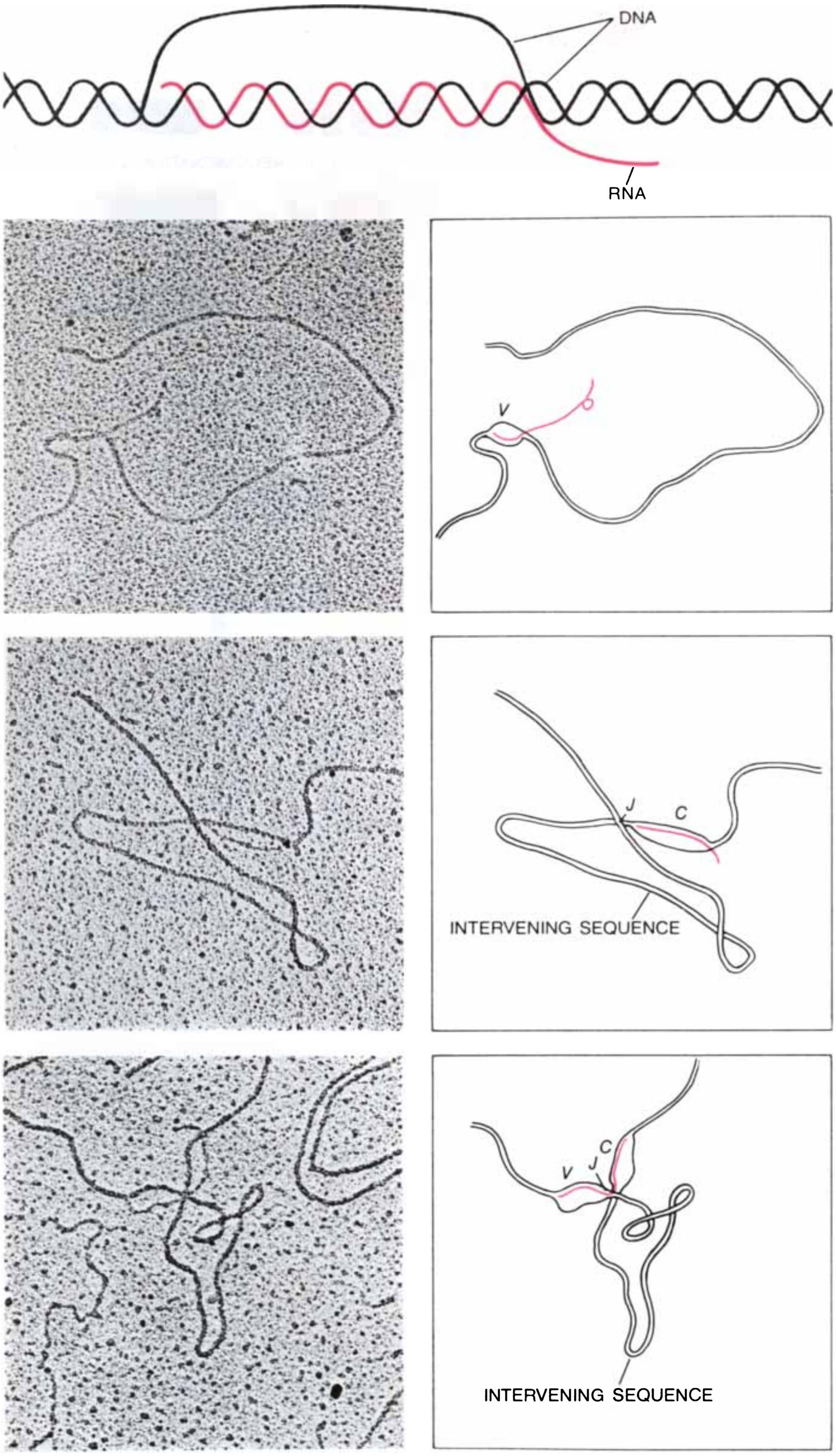
Now the potential of the kappa system to develop diversity begins to come clear. The joining of one of several hundred *V* genes—say 150 to be conservative—to one of five *J* genes can generate 150 × 5, or 750, different active genes for a light-chain variable region. Evidence from a number of laboratories indicates that this is exactly the way the sequences are shuffled. One of the *V* genes is joined to one of the *J* genes; the extra *V*'s and *J*'s (and the long noncoding spacer) between them are deleted. The finished active gene is encoded in three separate coding sequences: a leader gene, a *V*/*J* gene and a *C* gene. The sequences are assembled by RNA splicing to form a coherent light-chain messenger RNA.

The 750-fold diversity I have accounted for is multiplied by another source of variation. Careful comparisons of the amino acid sequences of light chains revealed a particularly high degree of diversity in a region close to the site of *V*/*J* joining. The amino acids around position 96 form one of the three regions of the light chains that Elvin A. Kabat of the College of Physicians and Surgeons of Columbia University had earlier designated as "hypervariable." The light chains fold up in such a way that the hypervariable regions form the antibody-antigen combining site.

At least some part of the variation in this region can now be explained by the fact that the *V*/*J* recombination site is not precisely defined. A *V* gene and a *J* gene can apparently be joined at different crossover points [*see illustration on page 111*]. As a result the codon for amino acid 96 (the nominal *V*/*J* junction) and the codons adjacent to it can change depending on what part of the sequence is supplied by the embryonic *V* region and what part by the *J* region. If one makes the assumption that alternative joining sites can increase the diversity tenfold, the total number of potential *V*/*J* combinations becomes 150 × 5 × 10, or 7,500.

## Signal Sequences

I mentioned above that certain features of the sequences of the light-chain genes have been conserved and seem likely to be of functional significance. In particular a pattern of signal-like sequences is found on the so-called 3′, or downstream, side of the *V* genes and on the 5′, or upstream, side of the *J* genes. Each such sequence has a stretch of about nine nucleotides (a nonamer) of which a large proportion are either A's or T's. The nonamer is followed, at an interval of either about 11 or about 22 nucleotides, by a seven-nucleotide sequence, or heptamer: CACTGTG or GTGACAC. The nonamer and the heptamer can be visualized as forming a "stem" structure in which the sequences




**COMPONENTS OF ANTIBODY GENES are visualized in electron micrographs of mouse DNA and RNA prepared by a special hybridization technique. Under certain conditions an RNA sequence complementary to one strand of a DNA double helix can be made to hybridize with that strand more readily than the complementary DNA strand does; the DNA strand is therefore displaced to form a loop, as is indicated in the diagram at the top. In the micrographs the RNA (which is shown in color in the accompanying drawings) is messenger RNA for the light chain; it is a contiguous *V/J/C* sequence. For the top micrograph the RNA was mixed with embryonic-cell, or germ-line, DNA. The *V* sequence of the RNA has hybridized with one strand of a germ-line *V* sequence to form a loop; the remainder of the RNA trails away. A similar loop in the middle micrograph defines the *C* gene, whose DNA is separated by a long, unhybridized intervening sequence from a *J* sequence. (One strand of the germ-line *J* has bound to a bit of messenger RNA that is not visible in the micrograph.) For the bottom micrograph the messenger RNA was mixed with DNA from an antibody-producing cell, where the *V* and *J* sequences have been joined. Now the messenger RNA hybridizes with all three genes, forming a *V/J* loop and a *C* loop; again the DNA's intervening sequence cannot bind to RNA.**

would be complementary according to the rules of base pairing (A pairs with T, G with C), bringing the *V* and the *J* genes together at the base of the stem [*see top illustration on page 112*]. It is then easy to see how the genes might be joined by some kind of DNA-recombination mechanism, with the signal sequences being deleted. The process is probably mediated by a specific (but as yet undiscovered) system of recombinatory enzymes.

The flexibility of the recombinational system, although powerful in its ability to generate diversity, does have its price. *V* and *J* genes are occasionally brought together aberrantly, yielding an inactive gene. This fact may in part explain the phenomenon called allelic exclusion. Each somatic cell has two sets of chromosomes, with one member of each chromosome pair supplied by the mother and the other by the father. The corresponding copies of a given gene on the two chromosomes are called alleles. In an antibody-producing cell it is usually the case that only the antibody genes on one copy of the chromosome carrying such genes undergo somatic recombination and are finally translated into protein; the alleles on the other chromosome are ordinarily excluded from the rearrangement process and are not expressed. Occasionally, however, an antibody-producing cell has two sets of rearranged antibody genes. My colleagues and I suspect that in such cases the genes on one chromosome were recombined aberrantly, forming an inactive antibody gene, and that the active gene was thereupon generated in a second try, with the "spare" chromosome serving as the backup.

It is also possible that the lambda light-chain genes serve as a fail-safe system for misjoined kappa genes. Studies done by Hieter, Hollis, Stanley J. Korsmeyer, Thomas A. Waldmann and me show that there is a clear order of events in the formation of a light-chain gene. First the kappa genes rearrange. If the kappa genes fail to form an active gene, the lambda system begins to rearrange.

Baltimore and his co-workers at M.I.T. have proposed the intriguing notion that the appearance of a functional antibody in a cell acts as a signal precluding further *V/J* joining; in the absence of the signal the cell keeps trying. In human beings, with six lambda *J/C* sequences on each chromosome No. 22, there are in effect 12 lambda genes per cell. Together with the two kappa genes (one on each chromosome No. 2), that would mean the cell has 14 opportunities to form an active light-chain gene.

## The Heavy Chain

In the heavy chain the formation of the variable region is governed by the same principles that apply in the light chain, but the potential for diversity is even greater: an extra piece of genetic information multiplies the combinatorial possibilities. When Leroy E. Hood and his colleagues at Cal Tech cloned an antibody-producing cell's active heavy-chain variable-region DNA and determined its structure, they found that a sequence of at least 13 nucleotides exactly at the *V/J* junction could not be accounted for by either the *V* or the *J* genes in the embryonic DNA. They reasoned that this segment must be supplied by a stretch of embryonic DNA they called the *D* (for diversity) gene. They noted that the *D* segment's location in the active gene corresponds to the major portion of the third hypervariable region of the heavy chain.

Philip W. Early and Hood also noted in the embryonic heavy-chain DNA the nonamer-heptamer sequences that seem to serve in light-chain DNA as signals for *V/J* joining. The arrangement of those signals in light-chain DNA had indicated that the spacing between the nonamer and the heptamer had to be different on the *V* side and the *J* side of the stem structure (about 11 nucleotides on one side and about 22 on the other) for recombination to take place. In the heavy chain, however, the *V* and the *J* signals were both found to have spacers about 22 nucleotides long. Early and Hood therefore predicted that when the *D* genes were identified in embryonic DNA, they would be found to be flanked by recombination signals having the 11-nucleotide spacing. In this way the joining of *V* to *D* and of *D* to *J* would conform to the 11/22 rule.

The putative *D* segments presented a problem in cloning. Their coding sequences were too short for them to serve as a hybridization probe for detecting their embryonic counterparts. My colleagues Ulrich Siebenlist and Jeffrey V. Ravetch solved the problem by taking advantage of an aberrant intermediate product of *V/D/J* recombination: a human *D* segment that had somehow been joined incorrectly to a *J* gene. The *D* gene had been processed by the recombinatory enzymes on only one side (the side next to the *J* gene), and so its opposite flank carried with it a large segment of embryonic DNA. The aberrant segment made an excellent probe for cloning embryonic *D* sequences.

When Siebenlist and Ravetch determined the structure of the DNA in these clones, the *D* sequences were found to be surrounded by recombination signals with 11-nucleotide spacers, fulfilling the predictions of the 11/22 rule. Then, with the cloned *D* segment as a probe, Siebenlist and Ravetch searched the human genome for sequences with which it would hybridize. They detected a large *D*-gene family consisting of at least five closely related sequences. There are indications that this is only one of several such families of *D* genes.

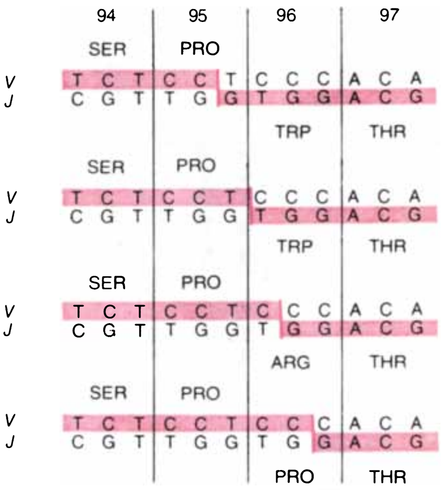
The formation of the active gene for the heavy-chain variable region can generate an extraordinarily large number of genetic possibilities. T. H. Rabbitts and his colleagues at the Medical Research Council Laboratory of Molecular Biology in Cambridge have estimated that there are as many as 80 embryonic heavy-chain *V* genes in man. Ravetch has found six active *J* genes within 8,000 nucleotides of the human mu *C* gene. Although one cannot confidently extrapolate the number of human *D* genes from what is known now, I shall assume that the *D* families have about 50 members. Somatic recombination can thus generate approximately $80 \times 6 \times 50$, or 24,000, genetic combinations. Another factor of about 100 (a very rough estimate) is contributed by recombinational flexibility: alternative codons at the two crossover points of *V/D* and *D/J* recombination. The total then comes to about 2.4 million possible different heavy chains.

## 18 Billion Antibodies

Taken together with the 7,500 combinatorial possibilities available to the human kappa light chain (150 for the complement of *V* genes, five for the *J* genes and 10 for recombinational flexibility), the 2.4 million heavy chains yield a total of some 18 billion (2.4 million multiplied by 7,500) possible antibodies. They can be generated from perhaps 300 separate genetic segments in the embryonic DNA.

The enormous diversity generated by means of recombination may be supplemented by yet another mechanism: solitary somatic mutation, which introduces sporadic single-nucleotide changes throughout the variable-region DNA in the course of somatic development. Immunoglobulin genes are highly unstable in antibody-producing cells. When Matthew D. Scharff and his colleagues at the Albert Einstein College of Medicine propagated clones of mature lymphocytes and screened successive generations for new antigen-binding specificity, they found the immunoglobulin genes underwent mutation at the remarkable rate of once per 10,000 cells per generation. In other experiments active genes have been isolated whose variable-region DNA differs by one nucleotide or two from the embryonic-DNA sequences that were its source.

These findings support the suggestion, made by Melvin Cohn and Martin Weigert of the Salk Institute for Biological Studies about 15 years ago, that some proportion of antibody diversity is the result of single-nucleotide mutation in the variable-region DNA of developing lymphocytes. Patricia J. Gearhart of the Carnegie Institution of Washington and Hood and Baltimore and their associates have evidence to suggest that the mutations accumulate as the lymphocyte passes through progressive stages in




**RECOMBINATIONAL FLEXIBILITY amplifies antibody diversity. The crossover point (*dark colored lines*) at which the *V* and the *J* sequences recombine can vary over a range of several nucleotides, giving rise to different nucleotide sequences (*colored bands*) in the active kappa-chain gene. The result is that the codon for amino acid 96 of the chain can vary: TGG codes for the amino acid tryptophan, CGG for arginine and CCG for proline. The variation is within the third hypervariable region of the kappa chain, and so it can have a large effect on the antigen-combining site.**

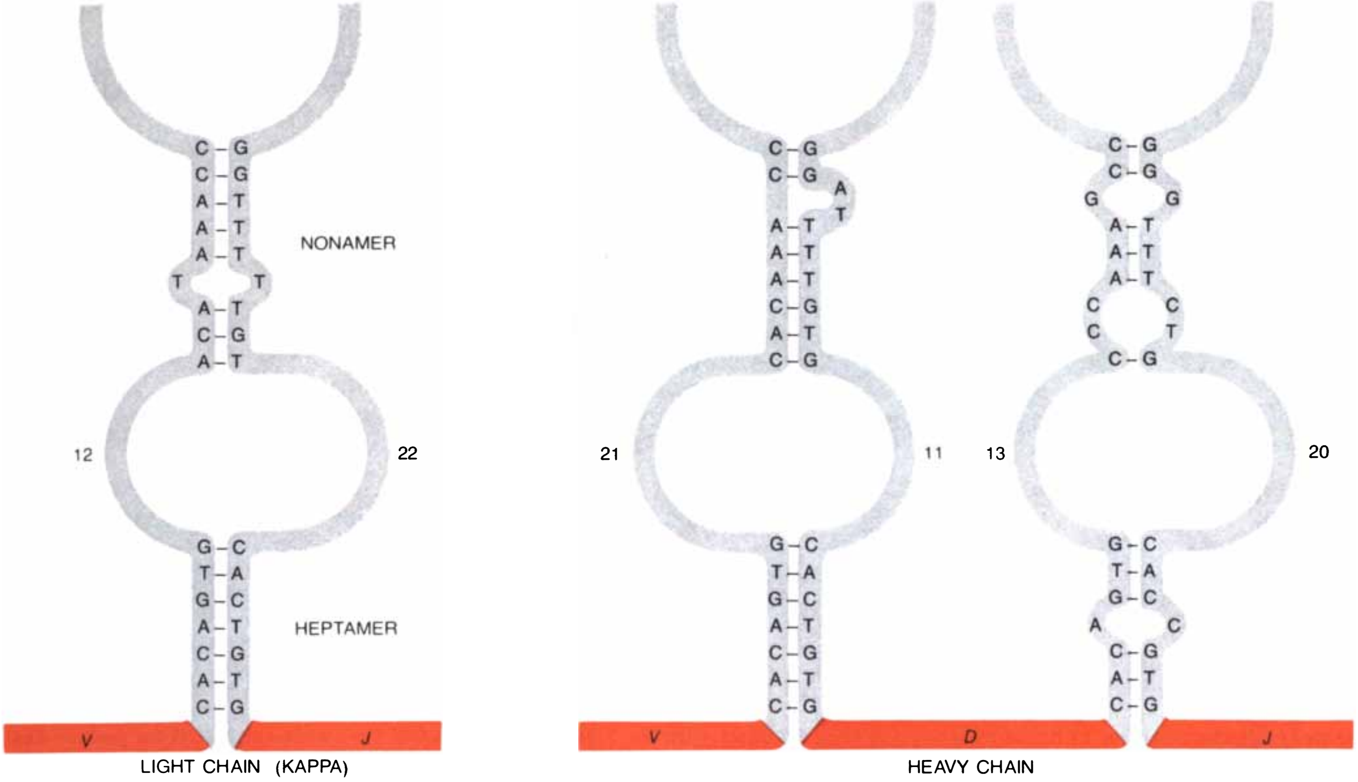
its development. The mechanism that leads to such mutations is not known.

The gene shuffling that can generate billions of variable-region genes is matched by two additional processes that explain how a single variable region can be joined successively to a series of heavy-chain constant regions. A precursor of the antibody-producing cells, the pre-*B* lymphocyte, makes a mu heavy-chain constant region linked to a specific variable region (a product of *V/D/J* recombination). This heavy chain at first remains inside the pre-*B* cell. Then, after the onset of light-chain and delta heavy-chain synthesis, both the mu and the delta heavy chains combine with the light chains to form complete IgM and IgD molecules. The next stage in development is distinguished by the concurrent appearance of both IgM and IgD on the cell surface. Both antibodies have the same variable regions and so both are directed against the same antigen.

## Clonal Selection

The subsequent steps in lymphocyte maturation are apparently antigen-driven. The primary event of the immune process is called clonal selection. An antigen binds to a receptor: the best-fitting antigen-combining site among millions or billions of surface immunoglobulins. By this interaction the cell displaying the selected immunoglobulin is driven farther along its developmental pathway. It proliferates to form a clone of antibody-producing *B* lymphocytes.

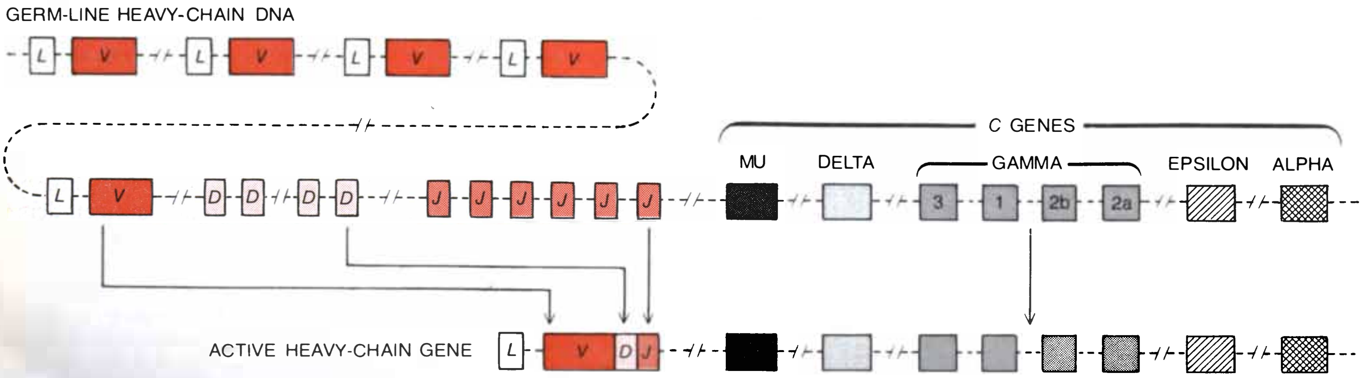
In the course of *B*-cell maturation the




**RECOMBINATION SIGNALS** in germ-line DNA apparently serve to bring *V* and *J* sequences for the light chain together prior to recombination. For the heavy chain a similar mechanism operates, but it includes signals for the additional sequence designated *D* (for diversity). Each signal segment has a seven-nucleotide sequence, almost always with an A or a T at the center, and a sequence about nine nucleotides long that is rich in A's or T's. The heptamer and nonamer are separated by a spacer roughly 11 or roughly 22 nucleotides long. Two signal segments whose heptamers and nonamers are largely complementary according to the rules for base pairing (A pairs with T and G pairs with C) might form a stem structure that brings the coding sequences adjoining them together for recombination. Apparently the recombination takes place only when one of the two complementary signals has a short spacer and the other signal has a long spacer.

IgD and IgM disappear from the cell surface and either IgM, IgG, IgE or IgA is instead secreted by the cell. Each of these classes of immunoglobulins has a different heavy-chain constant region, but antibodies of any class that are synthesized in a given cell have the same variable regions, namely the ones that were assembled in the precursor cell and that formed the combining site selected by the antigen. Because each heavy chain gives the antibody a different effector function, the same combining site can take part in different immune reactions. The process by which the same variable region appears in association with different heavy-chain constant regions is called heavy-chain class switching. The lymphocyte depends on two mechanisms to carry out class switching. One mechanism is based on differential RNA transcription and splicing, the other on a version of DNA recombination.

The arrangement of the heavy-chain genes in mouse embryonic DNA, extending over more than 100,000 nucleotides, has been deciphered by several groups of investigators in the past three years. Hood and Tonegawa and their co-workers contributed principally to understanding of the structure of the mu, gamma and alpha genes. Frederick R. Blattner and his associates at the University of Wisconsin at Madison established the location of the delta gene. Tasuku Honjo and his associates in the Osaka University Faculty of Medicine were able to clone and map the gamma




**ACTIVE HEAVY-CHAIN GENE** is assembled from four sets of sequences in the germ-line configuration (*top*): *L/V, D, J* and *C*. Somatic recombination brings together one of the *L/V* sequences, a *D* and a *J* to code for the variable region of the chain. As in the case of the light-chain gene, the constant-region DNA (*C*) is downstream, separated by a noncoding sequence. In the heavy chain, however, there are eight separate *C* sequences, each one coding for a different constant region. (Each *C* sequence is divided into from three to five domains, but sequences are diagrammed in simplified form.) Final assembly of coding sequences is accomplished by RNA processing.

genes (there are four IgG subclasses) and link them to one another and in turn to the epsilon and alpha genes. The arrangement of these genes (reading from upstream to downstream, or from the 5′ end to the 3′ end of the sequence) is mu, delta, gamma 3, gamma 1, gamma 2*b*, gamma 2*a*, epsilon and alpha.

## Bound and Secreted IgM

How is it that IgM can appear successively in two forms, one bound to the lymphocyte membrane and the other secreted? Jonathan W. Uhr of the University of Texas Health Science Center at Dallas first noted a structural difference between the membrane-bound mu chain and the secreted one. More detailed studies indicate that the membrane-bound form ends in a short sequence of hydrophobic amino acids, which evidently anchor the antibody in the cell's membrane through their affinity for the hydrophobic lipids of the membrane. The secreted mu chain lacks the hydrophobic sequence. Hood, with Randall Wall of the University of California at Los Angeles, and Baltimore and his co-workers at M.I.T. showed that two forms of messenger RNA are synthesized from the mu gene. In one form the message stops just short of two small coding segments that specify the anchor sequence; in the other form the segments coding for the anchor sequence are included.

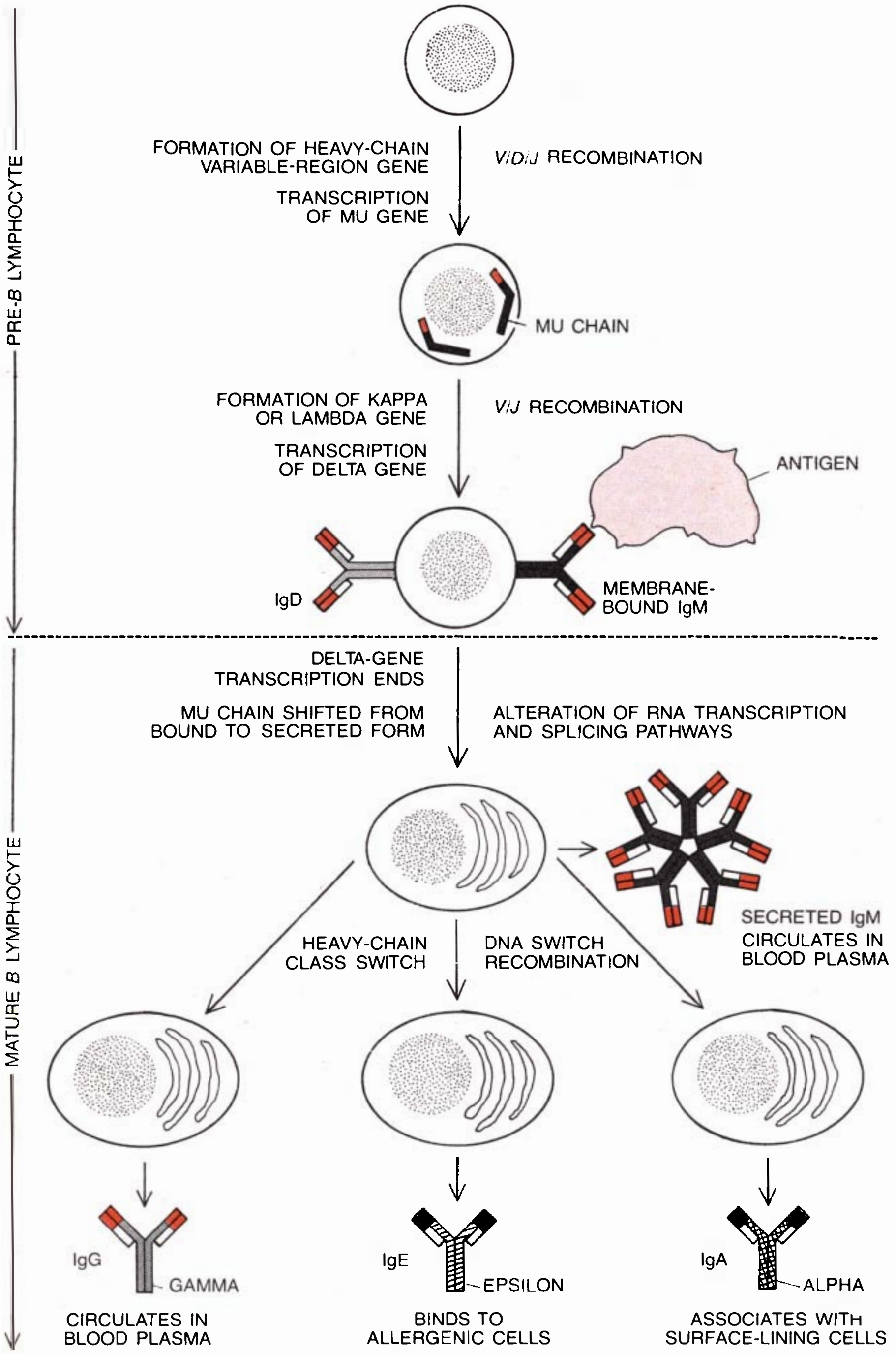
Each heavy-chain constant region is encoded in from three to six separate coding domains separated by short intervening sequences of noninformational DNA. As in the case of most other split genes, the primary transcript of the heavy-chain genes includes both the coding sequences and the intervening sequences. RNA processing thereupon splices the coding sequences to one another, eliminating the noninformational segments. The primary transcript of the mu gene sometimes includes the coding sequence for the hydrophobic anchor; if it is present, the enzymes that process the RNA splice it to the end of the main mu messenger so as to exclude a "stop" codon that would otherwise halt translation at the end of the main message. If the primary RNA transcript lacks the anchor sequence, no splicing takes place; the stop codon preceding the anchor sequence is recognized and the secretable form of the mu chain is synthesized [*see upper illustration on page 115*]. There is increasing evidence that many of the other heavy chains, and perhaps all of them, have a similar arrangement allowing them either to be anchored to the membrane or to be secreted.

The simultaneous appearance of mu and delta chains is likely to be due to similar splicing alternatives. Transcription proceeds through the variable-region DNA and then through the several domains of the mu gene. A certain fraction of the primary transcripts end in a way that yields mu chains, as described above. Another fraction of the transcripts continue for a few thousand nucleotides and therefore include not only the *V/D/J* and mu sequences but also the delta sequence. Among the many options for splicing such transcripts, one is to splice the *V/D/J* DNA directly to the beginning of the delta-gene transcript. In this way two messenger RNA's are formed simultaneously; one specifies the mu heavy chain and one the delta chain, but both encode the same variable region. Translation of the two RNA's into protein leads to the simultaneous display on the cell surface of IgM and IgD that have the same antigen specificity.

A second mechanism that shuffles heavy-chain genes depends on the rearrangement of DNA sequences to accomplish the remaining steps of the heavy-chain class switch. In contrast to the rather precise nature of the *V/J* and *V/J/D* joining events, the recombination process that leads to the expression of a particular heavy-chain class has a much greater degree of freedom.

Consider the switch from the mu chain to the alpha chain (from IgM to IgA). The active heavy-chain gene is arranged so that the variable-region coding sequence is a long way—about 8,000 nucleotides—from the first constant-region coding sequence (namely the first mu sequence). Ordinarily this noncoding spacer region is excised from the primary RNA transcript for the mu chain. Although the spacer region has no known coding function, it does have a stretch of about 2,000 nucleotides that, both Hood and Tonegawa and their co-workers showed, includes a series of repeated nucleotide sequences of various sizes.

In our laboratory Ravetch and Ilan R. Kirsch were able to show that the 2,000-nucleotide segment with the repeated blocks is similar in sequence to a stretch of DNA far downstream, adjacent to the alpha gene. Switching from IgM to IgA must involve recombination of these distant similar sequences; they apparently serve as switch signals to join the *V/D/J* sequence to the alpha coding sequence, deleting mu and the other constant-region genes. Since there are no coding sequences within the switch regions, the exact crossover points in the recombination event can apparently vary widely within the region. Both Tonegawa and Honjo have found analogous switch regions adjacent to the gamma genes. They have suggested that these regions constitute a signal for the switch from mu to gamma.



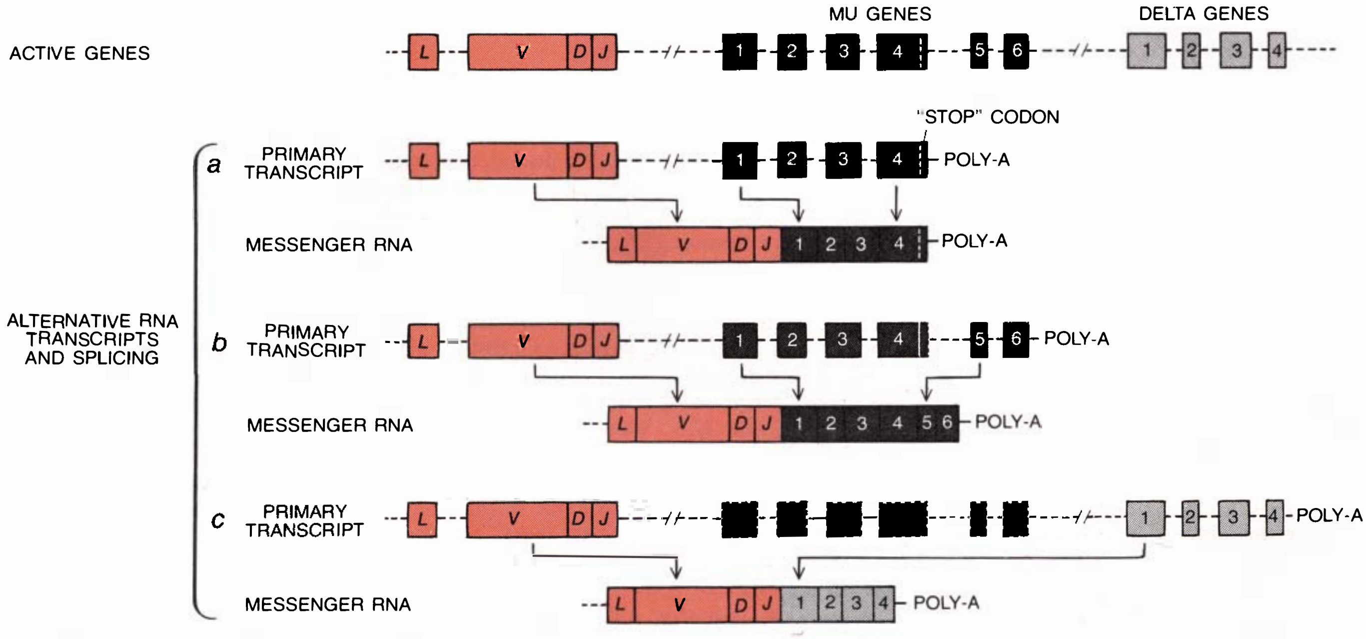

**DIFFERENTIATION of antibody-producing cells is traced beginning with the pre-*B* lymphocyte. At each stage the major genetic events are listed (*left*), along with the pertinent gene-shuffling mechanisms (*right*). The heavy chain is manufactured first, then the light chain (kappa or lambda). IgM and IgD are displayed on the cell surface. When a specific antigen is recognized and bound by a surface antibody, the cell is driven further in development: it proliferates to form a clone of mature *B* lymphocytes, which are specialized to synthesize large amounts of protein. Now the expression and the arrangement of the heavy-chain constant-region gene may change, so that different types of antibody are produced. In most cases the variable regions remain the same and the antibody continues to be directed against the same antigen, but point mutations can accumulate to change the variable regions in the course of maturation.**

## Broad Implications

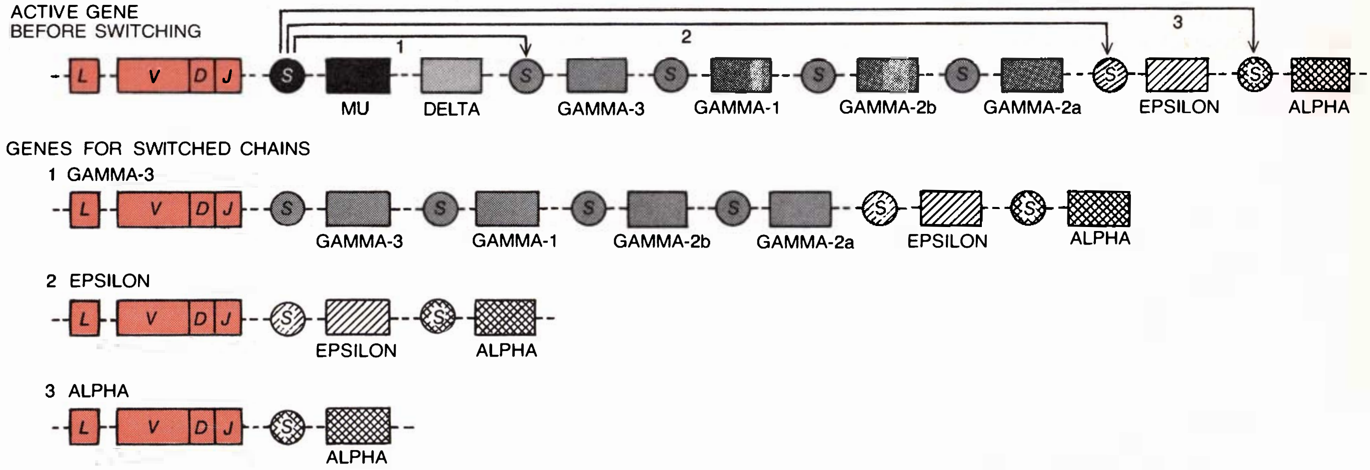
Not many years ago geneticists and molecular biologists had come to accept as axiomatic the principles that every protein chain is encoded by a particular germ-line gene, that the total genome present in embryonic cells remains unchanged as somatic cells differentiate to form various tissues and that the distinctive shapes and functions of the differentiated cells lie in the differential expression of the same genes. The discovery that the immune system generates many millions or billions of proteins by shuffling a few hundred germ-line genes shows that the principles do not always hold true. Proteins other than antibodies may also require more information than can readily be provided by an unaltered linear genome, and so it seems likely that somatic DNA recombination and selective RNA transcription and splicing may also operate to provide diversity in cells other than lymphocytes.

One indication that such mechanisms may be applied outside the immune system comes from an experiment done recently in my laboratory by Kirsch, Ravetch, Max and Robert L. Ney. They used the mu-alpha switch as a probe with which to search for any similar stretches of DNA in the mouse and the human genomes. Each probe identified from 10 to 15 fragments of DNA able to hybridize with the recombination signals. This seemed far too many to be accounted for by the need for heavy-chain class switching. The probe was then applied to the DNA of lymphocytes in which all heavy-chain genes between the *V* genes and the last constant-region gene, on both chromosomes, had been deleted. Again the probe recognized some 10 or 15 fragments, demonstrating that the "extra" signals are encoded outside the constant-region sequences, that is, elsewhere in the organism's genetic apparatus.

The implication is that such switching signals mediate the somatic recombination of other genes, for other proteins requiring a high degree of diversity. For example, recombination might serve to generate the diverse protein "addresses" that appear on the surface of embryonic cells and seem to guide them to a specific place and destiny in a complex organism. Or it might help to generate the diverse population of antigen receptors on the surface of *T* cells, a class of regulatory lymphocytes. Such possibilities remain to be investigated. What is clear is that the immune system has demonstrated the enormous potential of gene shuffling, which greatly increases the genetic information available for making antibodies and is likely to be central to the development of diversity in other gene systems as well.




**RNA TRANSCRIPTION AND SPLICING account for the successive appearance of membrane-bound and secreted IgM and for the simultaneous appearance of IgM and IgD. The top line shows the configuration of the heavy-chain genes for the variable region and for the mu and delta constant regions in an antibody-producing cell. If transcription ends at the fourth coding domain of the mu gene (*a*), the RNA is spliced to yield a coherent messenger RNA for a chain lacking a short amino acid sequence that would anchor the chain to the cell wall; IgM made from this messenger RNA is secreted. If the transcript includes mu domains 5 and 6 (*b*), splicing removes the "stop" codon at the end of domain 4 and brings the last two domains, encoding the anchor sequence, into the messenger RNA; the IgM is therefore membrane-bound. If the primary transcript includes the delta gene (*c*), splicing sometimes eliminates the mu gene and connects the delta domains directly to *J* in the variable-region sequence, making a messenger RNA that encodes a delta chain, and thus IgD.**




**HEAVY-CHAIN CLASS SWITCH is accomplished by DNA recombination. As long as a lymphocyte is synthesizing mu or delta chains the heavy-chain gene is arranged as is shown at the top. (The sequences encoding each constant region are actually divided into several domains and are separated by long noncoding segments.) Each constant-region gene is preceded by a switching signal (*S*) having some complementary relation to a similar signal between the variable-region sequences and the mu gene. The signals apparently mediate a recombination that joins a *V*/*D*/*J* sequence to one of the downstream constant-region sequences; three such possibilities are shown. The switched DNA is transcribed and the RNA is spliced to make a messenger RNA encoding the gamma-3, the epsilon or the alpha region.**

# The Asymmetry of Flounders

*Flatfishes have both eyes on one side of the head, and in most flatfish species the eyes are predominantly on the left side or the right. Is there some adaptive reason for this "handedness"?*

by David Policansky

Almost all vertebrate animals exhibit an approximate bilateral symmetry: structures to the left and the right of the midline of the body are mirror images. Among the most striking exceptions to this rule are the flatfishes, such as the flounder. In the adult flatfish there are many departures from the bilateral body plan, the most conspicuous asymmetry being that both eyes are on the same side of the head.

The asymmetry of a flatfish arises during the process of development. When the fish hatches from the egg, it has a symmetrical form, but the shape and position of the fins and other body structures are modified as the fish develops through its larval stages. Eventually one eye migrates across the top of the head, and the fish takes up the habit of lying on the bottom with both eyes facing up.

In addition to the asymmetry of the individual flatfish there is an equally intriguing asymmetry at the level of the species. In some species almost all adult fish lie on their right side and have both eyes on the left side of the head; they are said to be left-eyed or to exhibit left asymmetry. In other species virtually all the adults have the opposite asymmetry. This uniformity within most species (but not all of them) suggests that the asymmetry may be under genetic control and subject to the influence of natural selection. If it is, a troubling question arises: What selective advantage is there for the fish in lying on one side rather than on the other?

Depending on one's choice of classifications the order of flatfishes, the Pleuronectiformes, consists of six or seven families and includes about 500 species. Like most commercial fishes, the flatfishes have common names that can be wildly confusing. Some of the better-known species are the winter flounder, the summer flounder, the turbot, the English sole, the halibut, the starry flounder and the plaice. But whereas in England the sole really is a member of the sole family, the fish called the English sole in California is a member of the flounder family, and throughout the U.S. what most restaurants advertise as fillet of sole is more likely to be flounder. Here I shall use the name flounder to refer to the flatfishes in general.

The starry flounder, *Platichthys stellatus,* is particularly interesting to biologists: although its nearest relatives are mostly right-eyed, individuals of the species can be either left-eyed or right-eyed. The fish is found in Pacific coastal waters from southern California to Japan. Off the U.S. West Coast starry flounders are nearly half left-eyed and half right-eyed. Midway between the U.S. and Japan, off Alaska, about 70 percent of them are left-eyed. In Japanese waters nearly 100 percent are left-eyed.

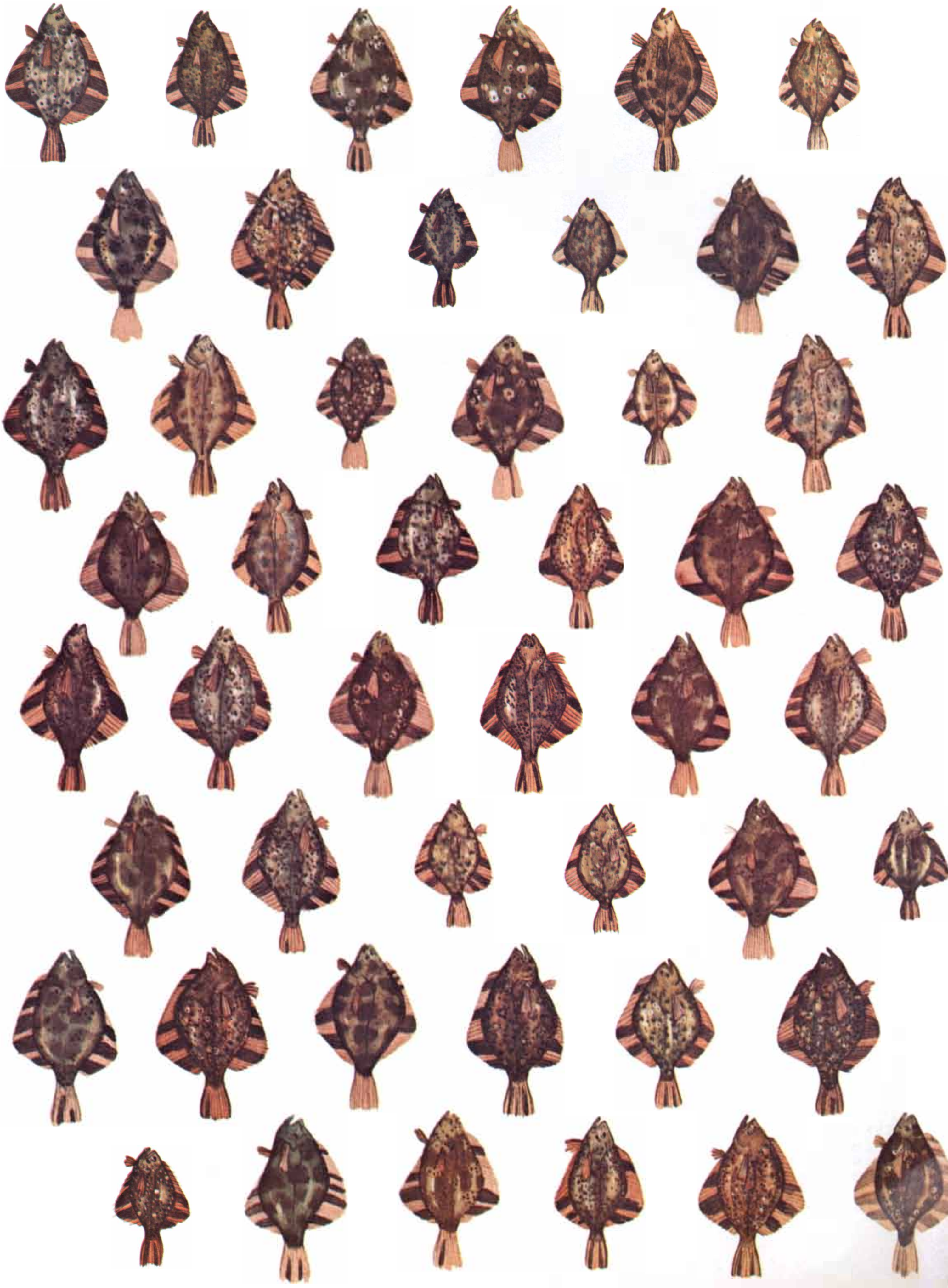
Biologists call this kind of gradual variation over a certain geographic range a cline. Clines are of interest because they usually are indications of natural selection, or differences in "adaptedness." For the starry flounder the difference is that between mirror images. The idea that such a geometric difference might be an adaptive one is intriguing. Thus the first question that arises is whether the cline is a historical accident or whether natural selection plays a role. In order even to think seriously about the question one needs to know whether the phenomenon is under genetic control.

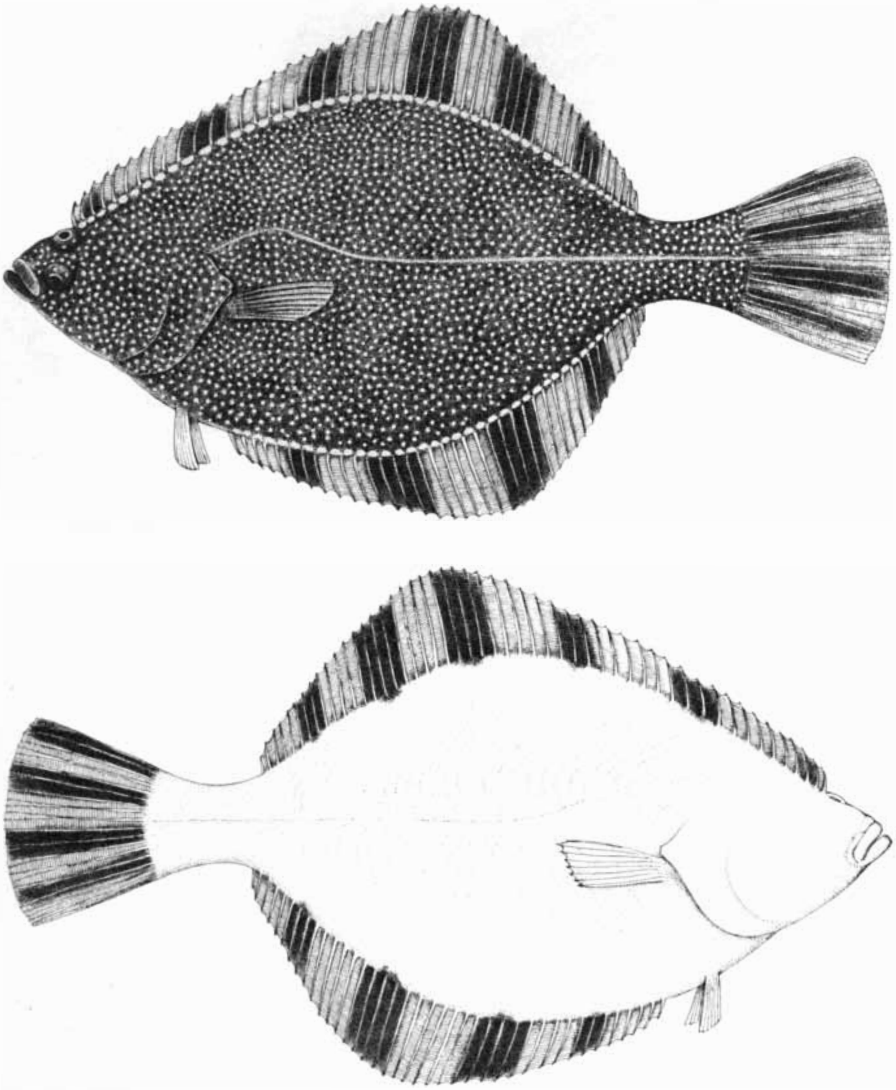
In spite of the current understanding of the way genes work, both in general and in many specific instances, no one can explain how genes code for asymmetries. As a look at the confusing literature on human handedness will show, it is sometimes even difficult to decide whether the directionality of asymmetries is under genetic control at all. Therefore it was of interest to me to study the left and right asymmetry of the starry flounder in the light of similar phenomena in other organisms and to compare the findings with the predictions of two proposed models of how asymmetries may be inherited. Here I shall first discuss the two models and then describe my study of Japanese and American starry flounders.

The clearest instances of genetic control of left and right asymmetry are found in three species of snails: *Lymnaea peregra,* which has a shell that is usually coiled to the right, *Laciniaria biplicata,* which has a shell that is usually coiled to the left, and *Partula suturalis,* which has a shell that coils either to the right or to the left. Other examples of asymmetries that are under genetic control are provided by an insect (a mutant laboratory strain of the fruit fly *Drosophila*), a protozoan (a mutant strain of *Tetrahymena*) and two closely related plant species that belong in the genus *Medicago,* a group of legumes that includes common alfalfa.

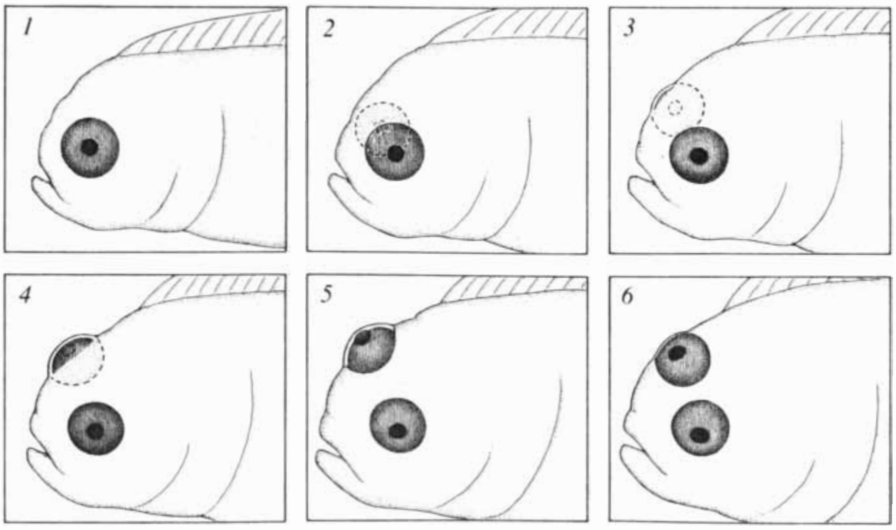
In all three of the snails the genetic constitution of the female determines the offspring's direction of coiling; coiling to the right is genetically dominant in *Lymnaea* and coiling to the left is dominant in the other two species. In the *Drosophila* mutant the abdomen is twisted either clockwise or counterclockwise; the different directions of twist are determined by variant genes at the same locus on one *Drosophila* chromosome.

**FORTY-EIGHT FISH rendered in watercolor on the opposite page are starry flounders, the species *Platichthys stellatus.* They inhabit Pacific coastal waters from California to Japan. The ones in this painting by the California artist Albert M. Plaza are members of the California population. Sixty percent of them have both eyes on the left side of the head; the remainder have both eyes on the right side of the head. Fewer than 1 percent of the starry flounders from Japanese waters are right-eyed. Experimental matings between the two populations provide a means of testing various hypotheses about the genetic and nongenetic control of asymmetry.**





**LEFT-EYED STARRY FLOUNDER is seen as it would appear on the sea floor to a viewer from above (*top*) and turned upside down (*bottom*). It is one of some 500 species of flatfishes.**




**EYE MIGRATION, which begins when starry flounder larvae are about seven millimeters long, changes the young fish from its originally symmetrical condition (*1*). The fish shown here is on the way to becoming left-eyed. The eye on the far (right) side begins to drift upward toward the top of the skull (*2, 3*). The fish is still transparent, and the shift of the far eye is visible when the fish is viewed from the near (left) side. The process continues (*4, 5*) over a period of some five days until the migrating eye has moved almost entirely to the near side (*6*).**

Such variant genes are called alleles. Hence at the "abdomen twisting" locus on the chromosome there might be a "left" allele, a "right" allele or perhaps a "neutral" allele.

Most animals and plants have paired chromosomes, and so at each chromosome locus they also have paired alleles, one allele coming from each parent. The alleles in a pair may be the same, in which case the organism is homozygous, or they may be different, in which case it is heterozygous. In the mutant protozoan the reversal of asymmetry is determined by an allele at a locus named *janus* (after the two-faced Roman protector of gateways). In the legumes different alleles found at the same locus determine the direction of pod coiling; the allele for "right-coiling" is the dominant one.

It is not surprising that the principal model concerned with the inheritance of asymmetries was devised by a student of the most intensively investigated and least understood of developmental asymmetries: human handedness. That is the British psychologist Marian Annett. She has proposed that whereas right-handedness is associated with a "right-determining" allele, there is no "left-determining" allele, only a neutral allele that does not determine handedness at all. Therefore when both parents have two of the neutral alleles, handedness is not genetically determined, and their offspring have just as much chance of being left-handed as they have of being right-handed.

Two other psychologists, M. J. Morgan of University College London and Michael C. Corballis of the University of Auckland, have proposed that all models of asymmetry inheritance should be like Annett's, that is, with one allele coding for directionality and the other allele neutral. They go on to suggest that even the coding alleles do not intrinsically code for directionality but instead work on a preexisting gradient of asymmetry in the fertilized egg.

The model alternative to Annett's suggests in effect that the coding alleles "know the difference" between left and right. This is to say that when the allele at the appropriate locus is right-coding, a right asymmetry results, and that when the allele is left-coding, a left asymmetry results. In either model more than one locus could be involved. In neither has a detailed mechanism of operation been suggested.

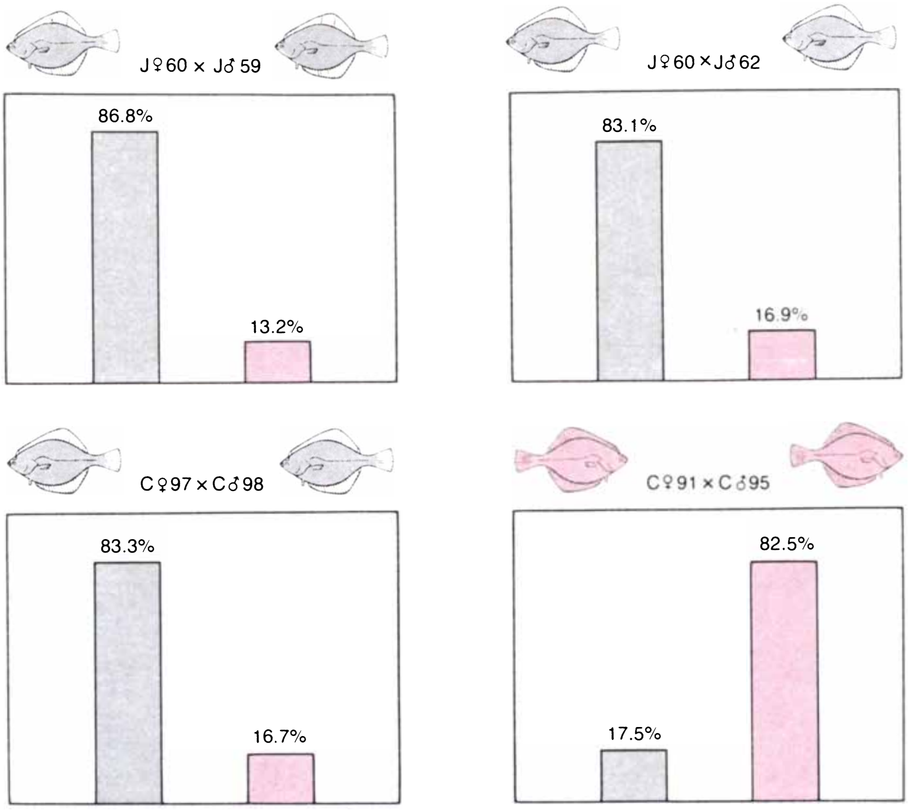
The flounder is asymmetric only as an adult; the newly hatched larval flounder is symmetrical. In *The Life Story of the Fish: His Manners and Morals* the biologist Brian Curtis quotes an anonymous journalist on the subject of this metamorphosis. The somewhat fanciful account is as follows: "The flounder is the ichthyological acme of lassitude. He begins life swimming in an upright position like any normal fish. Before he is many weeks old, however, he begins to tire in the cosmic struggle for existence. He sinks to the bottom, stretches out on his side and refuses to get up again. In this position he finds himself with one eye staring in a futile fashion into the mud. The eye, apparently tiring in its efforts to pierce the primordial ooze, behaves in a manner still unexplained by science. It moves around and joins the other optic, fortunate enough to be on top. This results in the flounder being one of the silliest looking of all fishes, but it also enables him to achieve his aim. In piscine indolence, he lolls on the bottom with his misplaced orb and its fellow peering upward for any food that may drift down to him. Even this occasionally wears on the flounder, and when it does he buries himself in the mud where he doesn't even have to look."

In order to study the genetics of asymmetry among starry flounders I had to establish a breeding program and then rear the offspring to the stage where their eyes began to migrate. The species' eggs are slightly less than a millimeter in diameter, and as the embryo develops it has to curl tightly to fit in such a small sphere. The larva hatches by breaking the egg membrane and wriggling free; at this stage it is little more than a yolk sac with two eyes and a tail. The sac is absorbed in a few days. In that time the larva has grown from an overall length of 2.2 millimeters to 3.5 millimeters. Its mouth has developed and it is able to feed.

This is a critical stage of growth, because with the supply of yolk gone the larva will starve if it does not soon begin to eat. It is necessary to supply food that is both nourishing and small enough for the tiny organism to ingest. Only in the past 20 years or so has such a food been available to fish breeders. It is a small marine organism, a rotifer of the genus *Brachionus,* that is easily raised in large numbers.

Once the fish have started to feed the greatest difficulty in raising them is over. They grow longer and deeper as their fins, bones and pigmentation develop. The interval before the onset of metamorphosis is surprisingly variable, ranging from 27 to 104 days after the fertilization of the egg. Evidently size is more important than elapsed time as a controlling factor: regardless of age the fish are all about seven millimeters long at the onset of metamorphosis.

The process begins while the bones are still incompletely ossified and calls for a rotation of the neurocranium. This rotation involves the brain and the orbits (sockets) of the eyes but does not directly involve the rest of the head.



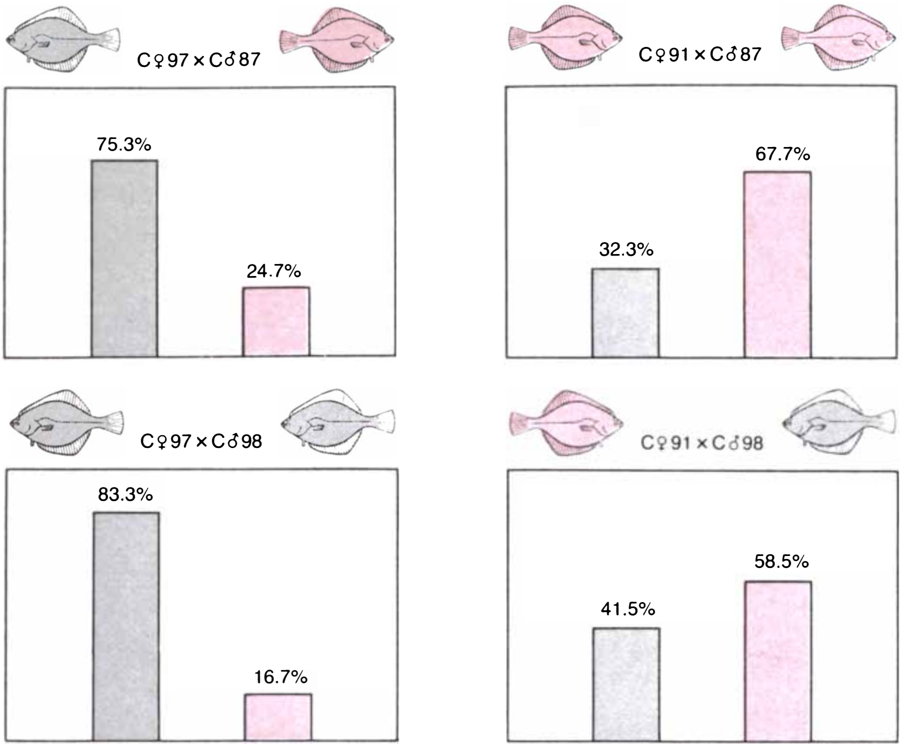

**EXPERIMENTAL MATINGS disposed of two hypothetical models proposed to explain the difference in the proportion of left-eyed and right-eyed fish in the Japanese and California populations of starry flounders. When a left-eyed Japanese female flounder (No. 60) was mated with two of the left-eyed Japanese male flounders (Nos. 59 and 62), the offspring of both matings averaged about 85 percent left-eyed. This is not the typical proportion in the natural population, but it is high enough to rule out the nongenetic hypothesis. When two left-eyed (Nos. 97 and 98) and two right-eyed (Nos. 91 and 95) California fish were mated, the offspring were respectively 83 percent and 18 percent left-eyed. This result also rules out the nongenetic hypothesis and contradicts the prediction of one genetic hypothesis: that the offspring of all matings between California starry flounders would be half left-eyed and half right-eyed.**

There is some resorption of the soft bone that would otherwise impede the travel of the eye, but the process is mainly one of migration rather than tunneling. In most of the species examined, the starry flounder included, the dorsal fin remains behind the head while the eye migrates. The fin then grows forward almost to the tip of the snout. In a few species it grows forward before the eye has completed its journey; the eye must then move through the skin.

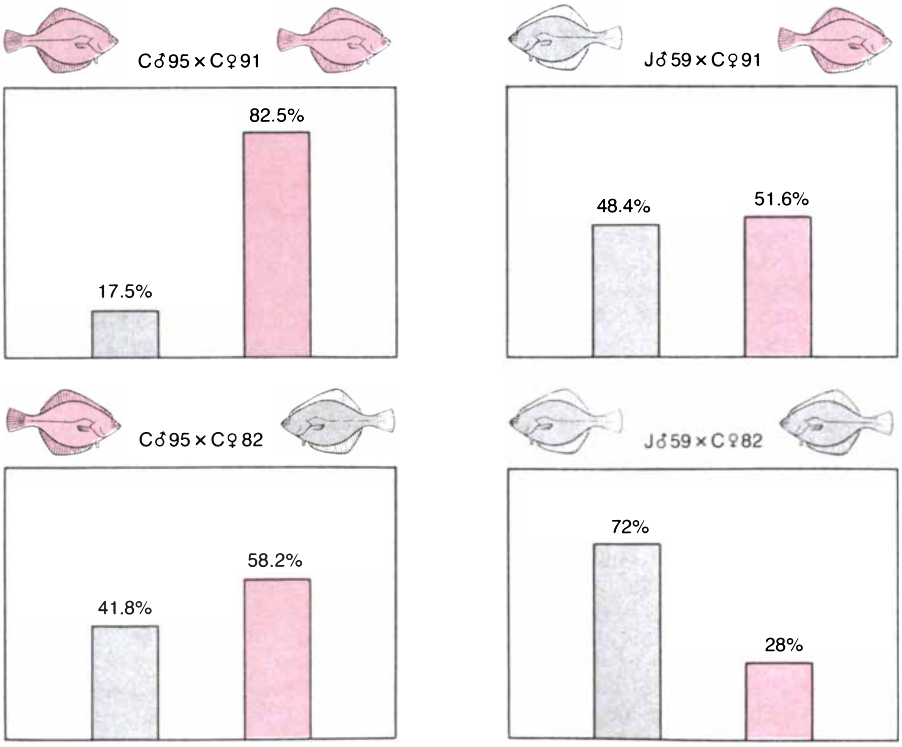
Some scholars have attributed the migration of the wrong-side eye to the tendency of the young fish to lean to one side or the other or to its resting on the sea floor on one side or the other. This does not appear to be the case. Among the thousands of flounders that I have reared through metamorphosis there seemed to be no relation between the side they first chose to lie on and the side on which the eyes ended up. Indeed, in some instances the metamorphosing fish continued to swim in an upright position near the top of the tank after the migrating eye had moved all the way to the top of its head.

Starry flounders are the best of all flatfishes to use in a study of the inheritance of asymmetries. There are three reasons for this. First, the difference in the percentage of left-eyed fish is so large (from 50 percent off California to nearly 100 percent off Japan) that it is statistically easy to decide whether the proportion of left-eyed fish in a cross is different from the proportion in the parental population. Second, the very existence of this cline seems to imply the action of some mechanism of natural selection. Third, in the absence of experimental evidence the fish give the appearance of conforming to the genetic model I shall call "M & C." I have so named it, including the quotation marks, because although it is like the version of Annett's model proposed by Morgan and Corballis, that model is a general one and was not specifically applied to flounders.

In terms of the "M & C" model the expectation is that the Japanese starry flounders are homozygous for left-eyed alleles: they all receive an allele for left-eyedness from both parents. With the California starry flounders, however,




**INFLUENCE OF THE FEMALE on the proportion of left-eyed offspring is demonstrated by two California females. When one left-eyed female (No. 97) was mated with a right-eyed California male (No. 87), it produced 75 percent left-eyed offspring. When a right-eyed female (No. 91) was mated with the same male, only 32 percent of the offspring were left-eyed. When the females were mated with a left-eyed California male (No. 98), the left-eyed female produced 83 percent left-eyed offspring and the right-eyed female fewer than 42 percent.**




**SIMILAR INFLUENCE of the male on asymmetries in offspring was demonstrated by two male starry flounders, one from California and one from Japan. When the right-eyed California male flounder (No. 95) was mated with a right-eyed California female (No. 91), it produced fewer than 18 percent left-eyed offspring. When the left-eyed male from Japan (No. 59) was mated with the same female, it produced more than 48 percent left-eyed offspring. When the males were mated with a left-eyed California female (No. 82), the California male produced 42 percent left-eyed offspring and the Japanese male produced 72 percent left-eyed offspring.**

the expectation is that they are homozygous for neutral alleles. The prediction that arises from this model is that any cross of two Japanese fish will produce mainly left-eyed offspring whereas any cross of two California fish will produce equal numbers of left-eyed and right-eyed offspring.

Alternative hypotheses and the predictions that arise from them are as follows. Assume, first, that in the Japanese population there are only left-eyed alleles whereas in the California population both left-eyed and right-eyed alleles are present. Some of the California fish would thus be homozygous for right-eyed alleles, others would be homozygous for left-eyed alleles and still others would be heterozygous. The Japanese population, however, would be homozygous for left-eyed alleles. From this condition one may predict, as before, that any cross of two Japanese fish would produce mainly left-eyed offspring. Because different California fish would have different pairs of alleles one may predict that two crosses between two different pairs of California fish might produce different percentages of left-eyed offspring. (This hypothesis includes subhypotheses relating to modes of inheritance and eyedness-allele frequencies.)

A further hypothesis is that there is no genetic difference between left-eyed and right-eyed starry flounders, either the left-eyed and right-eyed fish in California (this part of the hypothesis agrees with "M & C") or the left-eyed and right-eyed fish in both California and Japan (this part is not in agreement with "M & C"). Instead the difference between Japan and California in the proportion of left-eyed fish in the population is due to some environmental factor. One would predict from this nongenetic hypothesis that if the offspring were all reared in identical environments, all possible crosses would yield the same percentage of left-eyed fish.

In order to test these predictions I made various crosses between a total of 13 starry flounders: four females (one from Japan and three from California) and nine males (three from Japan and six from California). One of the California females and three of the California males were right-eyed. I reared all the offspring in small tanks at the New England Aquarium in Boston. The tanks were enclosed in water baths all of which were at the same temperature and had a common supply of seawater. All the larvae were fed the same food. These measures ensured that all the offspring were reared in identical environments.

The breeding results supported the first of the two hypotheses alternative to the "M & C" model and contradicted the predictions of both the "M & C" model and the nongenetic hypothesis, as follows:

Crosses between the only Japanese female and two of the Japanese males produced two sets of offspring that averaged about 85 percent left-eyed. Two separate California crosses (one between a right-eyed female and a right-eyed male and the other between a left-eyed female and a left-eyed male) produced offspring that were respectively only 18 percent left-eyed and 83 percent left-eyed. These California results run counter to the "M & C" prediction that all offspring of California fish will be 50 percent left-eyed. Both the Japanese and the California crossing results, of course, contradict the nongenetic prediction.

The results of other crosses identified certain of the fish as having tendencies toward the production of either left-eyed or right-eyed offspring. Hence one left-eyed California female, even when it was matched with a right-eyed California male, produced more than 75 percent left-eyed offspring. (This same female produced nearly 85 percent left-eyed offspring when, as was mentioned above, it was mated with a left-eyed California male.) In contrast, the only right-eyed California female, when it was bred to the same two California males mentioned above, produced 32 percent left-eyed offspring by the right-eyed male and 41 percent by the left-eyed male. Similarly, one of the left-eyed Japanese males, when it was mated with a left-eyed California female, sired 72 percent left-eyed offspring, and when it was mated with the right-eyed California female, still sired 48 percent left-eyed offspring.

It also appears likely that the proportion of left-eyed and right-eyed offspring was under some degree of environmental influence. The two crosses between Japanese parents respectively produced 13 and 17 percent right-eyed offspring. In Japan right-eyed starry flounders make up less than 1 percent of the population. It is possible that the single Japanese female was genetically aberrant and that this is why so many right-eyed offspring were produced. I nonetheless favor an environmental explanation. Many of the flounder larvae showed developmental anomalies; indeed, in a few instances they showed no eye migration. Thus it seems likely that the artificial environment of small tanks in an aquarium was not as conducive as the natural environment to normal developmental processes, the determination of asymmetry included.

These results also highlight a difference between snail and flounder asymmetries. In starry-flounder crosses both male and female parents influenced the proportion of left-eyed and right-eyed offspring. In snails only the female's genotype (its combination of alleles) influences which way the offspring coil. Furthermore, all the offspring from a single female snail coil the same way. The fertilized eggs of most mollusks develop by a process called spiral cleavage; the fertilized eggs of vertebrates develop by radial cleavage. "Spiral" and "radial" refer to the early cell divisions of the fertilized eggs. In mollusks the planes of these early divisions are rotated with respect to each other, resulting in a spiral development of the early embryo. In vertebrates the cleavage patterns vary, but they are never spiral.

According to experiments done by C. P. Raven of the University of Utrecht, who worked with the right-coiling snail *Lymnaea,* both the direction of cleavage and the direction of coiling are determined by an asymmetry in the cytoplasm of the egg. Raven found that the cleavage pattern in embryos that would develop into left-coiling snails was the mirror image of the cleavage pattern in right-coiling ones. These asymmetries, he determined, were due to asymmetries present in the egg before it was fertilized. This means that the asymmetry of the egg structure is laid down in the egg while it is still part of the female's body. That is why the female's genotype and not the male's determines the direction of coiling in the offspring. Flounder eggs, of course, do not have spiral cleavage; neither is control over the direction of their asymmetries purely maternal.

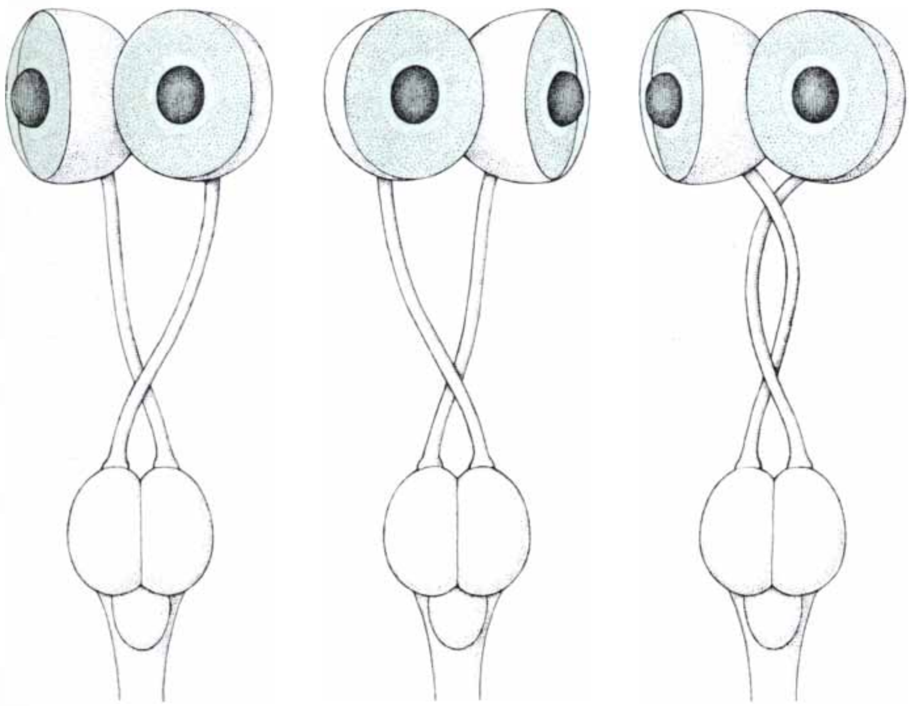
One cannot dismiss the difference between the California and the Japanese starry flounders' left-eyed and right-eyed proportions as being due to different developmental responses to different environments. The difference between right-eyed and left-eyed flatfishes is clearly an inherited one. This raises the question of whether left-eyedness (or right-eyedness) has some adaptive significance. It seems intuitively likely that it is advantageous to an individual flounder to have the same laterality, either left or right, as all the other members of its population. One source of this intuitive feeling is the fact that so few flatfish species exhibit a 50–50 laterality. Another is the fishes' mating behavior. If the male lies on top of the female, as it does in at least some species, imagine the difficulties when a right-eyed fish attempts to mate with a left-eyed one! Any such difficulty in mating would of course lead the species to evolve a single kind of laterality.

It is easier to imagine a disadvantage for left-eyedness in the starry flounder. Flatfishes, the starry flounder included, have a type of internal asymmetry. This involves the optic nerves. As in all bony fishes, the optic nerves cross, so that the right optic nerve is joined to the left side of the brain and vice versa. In mammals there is some interchange of nerve fibers at the chiasma (the crossing of the fibers) but in fishes there is almost none; therefore one can unambiguously identify a fish's right and left optic nerve.

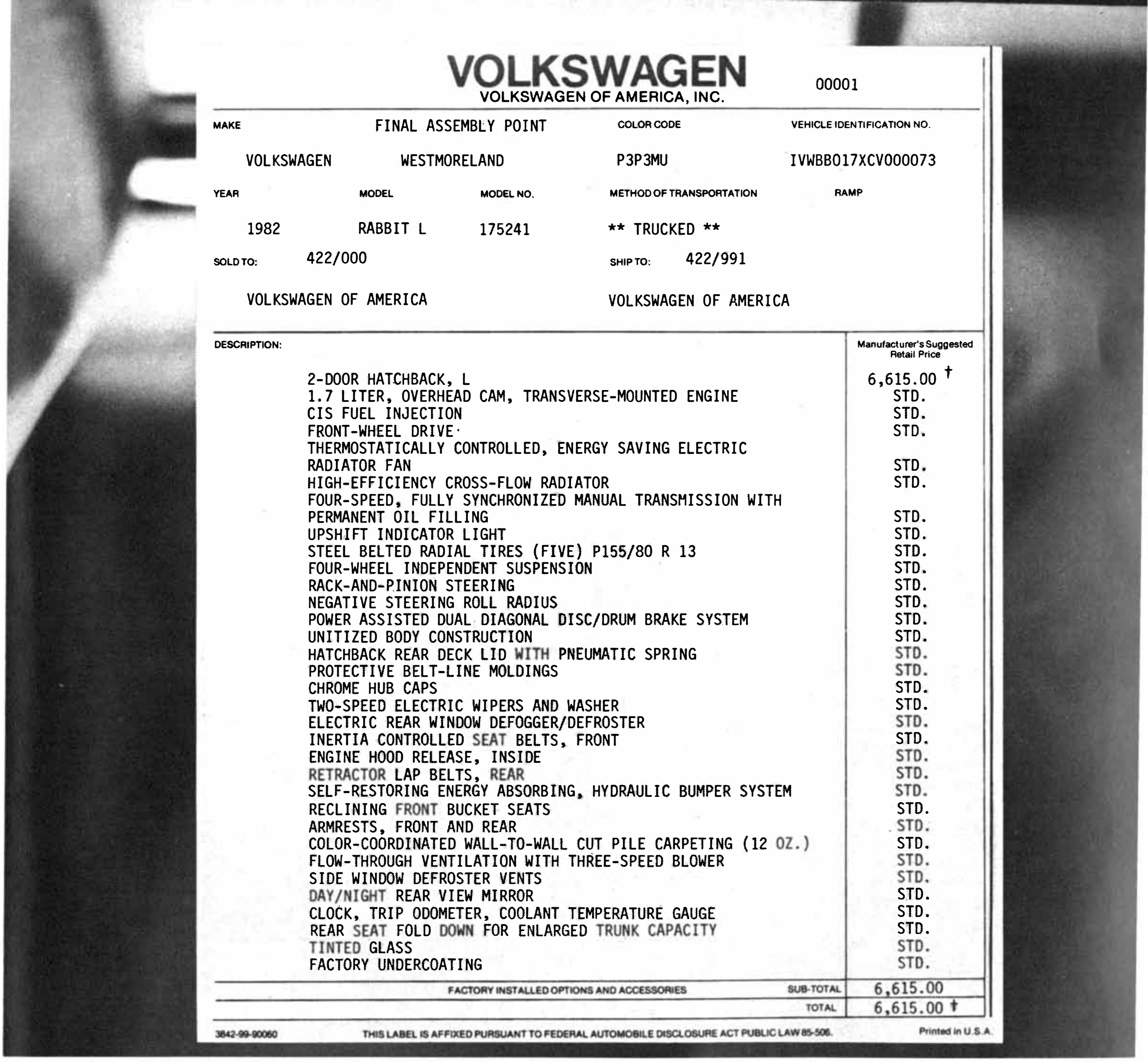
At the beginning of this century G. H. Parker of Harvard University published a study of the chiasmas of bony fishes. He found that in symmetrical fishes such as the cod the left optic nerve crossed above the right optic nerve in about half of the population and the right optic nerve crossed above the left in the other half. Some flatfishes were like the cod in this respect but others were different: the nerve of the eye that normally migrates was uppermost. The starry flounder belongs to the second group. Parker considered the arrangement to be one that facilitated eye migration. He reasoned that if the left eye migrated when the right optic nerve was on top, there would be a twisting of the nerves, whereas if the left optic nerve was in the higher position, there would be no twisting. He suggested that the twisting would be mechanically disadvantageous and proposed this as the reason for the optic-nerve arrangement in flatfish having evolved as it did.

If that is the case, then all the left-eyed Japanese starry flounders, with their left optic nerve uppermost and their right eye destined to migrate, are at a disadvantage. It is hard to see how this state of affairs can be the result of natural selection. It is also difficult to conjure up any kind of environmental asymmetry that would give left-eyedness an adaptive advantage over right-eyedness. For one thing, it is hard to imagine environmental asymmetries that would not be reversed simply by the fish's turning around. For another, even if one postulates some undefined environmental asymmetry as the reason for the predominant left-eyedness of the starry-flounder population in Japanese waters, what is the explanation for the presence of many right-eyed flatfish species in the same waters?

The only reasonable conclusion appears to be that there is no adaptive difference between left-eyedness and right-eyedness but that the two characteristics are genetically associated with some other characteristic, not yet recognized, that does have such significance. If this is the case, the situation is representative of a difficulty commonly encountered by biologists who study evolution: that of deciding which characteristics are adaptive and which are selectively neutral. As far as the difference between left-eyed and right-eyed flatfish species is concerned, this difference, however striking, appears to be an evolutionary red herring.



**OPTIC NERVES of flounders, like those of all bony fishes, cross so that the nerve from the left eye leads to the right lobe of the brain and the nerve from the right eye leads to the left lobe. Here three instances of such crossing are shown. At the left is the chiasma (the crossing of the nerve fibers) of a California halibut after its right eye has migrated to the left side of its head. In the middle is the chiasma of a right-eyed starry flounder after its left eye has migrated to the right side of its head. At the right is the chiasma of a left-eyed starry flounder after its right eye has migrated: a "twisting" of the optic nerves is apparent. This twisting may be disadvantageous and suggests that whatever the reason for the predominantly left-eyed population of Japanese starry flounders, it cannot be some kind of selective adaptation for left-eyedness.**



# VOLKSWAGEN

VOLKSWAGEN OF AMERICA, INC.

00001

| MAKE | FINAL ASSEMBLY POINT | COLOR CODE | VEHICLE IDENTIFICATION NO. |
|---|---|---|---|
| VOLKSWAGEN | WESTMORELAND | P3P3MU | IVWBB017XCV000073 |

| YEAR | MODEL | MODEL NO. | METHOD OF TRANSPORTATION | RAMP |
|---|---|---|---|---|
| 1982 | RABBIT L | 175241 | ** TRUCKED ** | |

SOLD TO: 422/000 — VOLKSWAGEN OF AMERICA

SHIP TO: 422/991 — VOLKSWAGEN OF AMERICA

| DESCRIPTION: | Manufacturer's Suggested Retail Price |
|---|---|
| 2-DOOR HATCHBACK, L | 6,615.00 † |
| 1.7 LITER, OVERHEAD CAM, TRANSVERSE-MOUNTED ENGINE | STD. |
| CIS FUEL INJECTION | STD. |
| FRONT-WHEEL DRIVE | STD. |
| THERMOSTATICALLY CONTROLLED, ENERGY SAVING ELECTRIC RADIATOR FAN | STD. |
| HIGH-EFFICIENCY CROSS-FLOW RADIATOR | STD. |
| FOUR-SPEED, FULLY SYNCHRONIZED MANUAL TRANSMISSION WITH PERMANENT OIL FILLING | STD. |
| UPSHIFT INDICATOR LIGHT | STD. |
| STEEL BELTED RADIAL TIRES (FIVE) P155/80 R 13 | STD. |
| FOUR-WHEEL INDEPENDENT SUSPENSION | STD. |
| RACK-AND-PINION STEERING | STD. |
| NEGATIVE STEERING ROLL RADIUS | STD. |
| POWER ASSISTED DUAL DIAGONAL DISC/DRUM BRAKE SYSTEM | STD. |
| UNITIZED BODY CONSTRUCTION | STD. |
| HATCHBACK REAR DECK LID WITH PNEUMATIC SPRING | STD. |
| PROTECTIVE BELT-LINE MOLDINGS | STD. |
| CHROME HUB CAPS | STD. |
| TWO-SPEED ELECTRIC WIPERS AND WASHER | STD. |
| ELECTRIC REAR WINDOW DEFOGGER/DEFROSTER | STD. |
| INERTIA CONTROLLED SEAT BELTS, FRONT | STD. |
| ENGINE HOOD RELEASE, INSIDE | STD. |
| RETRACTOR LAP BELTS, REAR | STD. |
| SELF-RESTORING ENERGY ABSORBING, HYDRAULIC BUMPER SYSTEM | STD. |
| RECLINING FRONT BUCKET SEATS | STD. |
| ARMRESTS, FRONT AND REAR | STD. |
| COLOR-COORDINATED WALL-TO-WALL CUT PILE CARPETING (12 OZ.) | STD. |
| FLOW-THROUGH VENTILATION WITH THREE-SPEED BLOWER | STD. |
| SIDE WINDOW DEFROSTER VENTS | STD. |
| DAY/NIGHT REAR VIEW MIRROR | STD. |
| CLOCK, TRIP ODOMETER, COOLANT TEMPERATURE GAUGE | STD. |
| REAR SEAT FOLD DOWN FOR ENLARGED TRUNK CAPACITY | STD. |
| TINTED GLASS | STD. |
| FACTORY UNDERCOATING | STD. |
| FACTORY INSTALLED OPTIONS AND ACCESSORIES — SUB-TOTAL | 6,615.00 |
| TOTAL | 6,615.00 † |

3842-99-90060 THIS LABEL IS AFFIXED PURSUANT TO FEDERAL AUTOMOBILE DISCLOSURE ACT PUBLIC LAW 85-506. Printed in U.S.A.

# You won't read it and weep.

Much to our consternation, a lot of people think Rabbits cost much more than they really do.

Regrettable, albeit really quite understandable.

Especially when you stop to consider that the Rabbit comes standard with fuel injection, front-wheel drive, and dual diagonal brakes, features usually found on much more expensive cars.

The Rabbit also comes standard with a gas-saving upshift light, not found on any other car regardless of price. Unless, of course, it just happens to be a Volkswagen.

Sure, if you're so inclined, you can indulge yourself with a long list of luxury options that can run up the price of even a Rabbit.

But the way we look at it, the basic Rabbit—reliable, economical* transportation at a reasonable price—may be the greatest luxury of them all.

**Nothing else is a Volkswagen.**

*EPA estimated 30 mpg, 42 highway estimate. (Use "estimated mpg" for comparison. Mileage varies with speed, trip length, weather. Actual highway mileage will probably be less.)
†Transportation, local taxes and dealer delivery charges additional.

# Cosmic Jets

*The violent activity at the center of many galaxies is manifested in the production of narrow, focused streams of ionized gas. Some are a few light-years long; others are a million times longer*

by Roger D. Blandford, Mitchell C. Begelman and Martin J. Rees

Astronomical observations made with radio telescopes have revealed that the center of many galaxies is a place of violent activity. The most recent discovery about this activity is that it is often manifested in the production of cosmic jets. Each such jet is a narrow stream of plasma (ionized gas) that appears to squirt out of the center of a galaxy, emitting radio waves as it does so. It can be more than a million light-years long. It terminates in an extended blob of radio emission well outside the optical image of the galaxy. The energy content of the blob can exceed $10^{60}$ ergs, an amount that would be produced by converting entirely into energy the mass of 10 million stars. Although a few jets have long been known from optical observations, it is only in recent years that new techniques in radio astronomy have shown how common jets are in the universe. Plainly jets must be produced under quite diverse conditions, and yet there is no consensus on precisely how they are produced.

Cosmic jets can take many different forms. One of the best-studied examples is the jet associated with the elliptical galaxy NGC 6251, which is 300 million light-years from our own galaxy. In 1977 Peter Waggett, Peter Warner and John Baldwin of the University of Cambridge discovered a long, straight jet emanating from the nucleus of NGC 6251. The jet has an angular width of only three degrees and yet it is more than 400,000 light-years long. Its narrow end coincides with the very center of the galaxy, and there the Cambridge observers found a small, pointlike source of radio emission. In 1978 Anthony C. S. Readhead and Marshall H. Cohen of the California Institute of Technology were able to map this pointlike source. They found that it too consists of a narrow jet with a point source at one end. It was as if a Russian doll had been taken apart only to reveal a smaller version inside. The small jet is colinear with the larger jet, but it is only three light-years long.

A somewhat different-looking jet is known as 3C 449 and is associated with an elliptical galaxy 100 million light-years distant. In 1979 Richard A. Perley and his colleagues, working with the Very Large Array (VLA) of radio telescopes near Socorro, N.M., found that two jets emerge from the center of this galaxy in opposite directions. Both jets show several sharp bends. The northern jet bends quite abruptly to the east about 100,000 light-years from the center; then it bends back and follows a northerly course until it appears to bend again. The southern jet also bends to the east; then it bends back toward the south. There are several jets that show this zigzag behavior.

The phenomenon of jets is not confined to other galaxies. A notable example is a mere 15,000 light-years from the sun in our own galaxy. It is called SS433. From an analysis of optical observations it is inferred that two jets emerge in opposite directions from a binary star system at a speed of 80,000 kilometers per second, or a little more than a fourth the speed of light. The same jets have now been detected at both radio and X-ray wavelengths. Evidently nature manages to produce jets within objects that are as light as a few suns or as heavy as a billion suns (the mass of the nucleus of a galaxy). The jets can be shorter than the distance from the earth to the sun or longer than 10 billion times that distance.

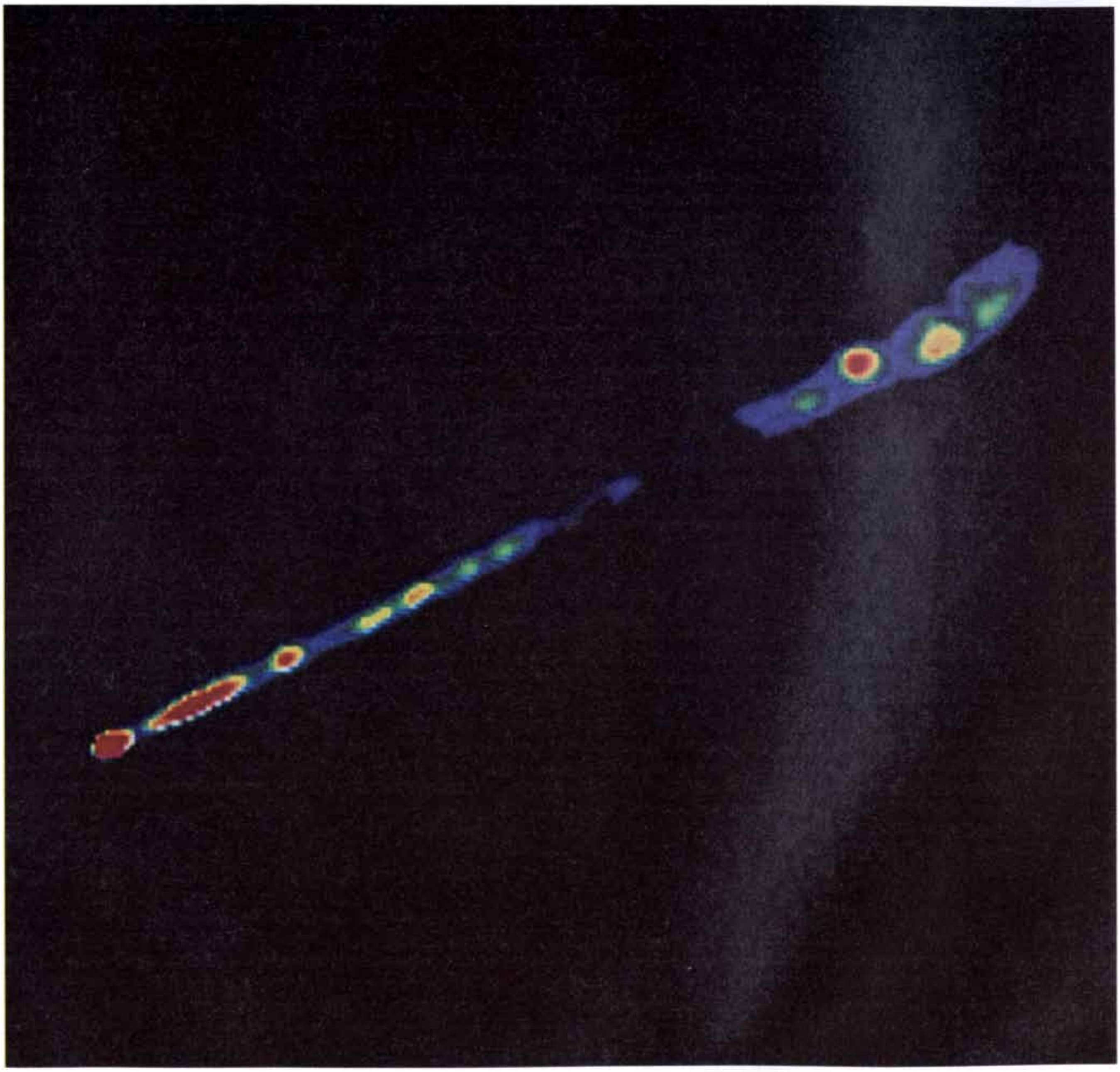
Jets are not new to astronomers. As long ago as 1917 Heber D. Curtis of the Lick Observatory discovered that an optical jet is associated with the large elliptical galaxy M87, which lies in a rich cluster of galaxies in the constellation Virgo. It was not until 1953, however, that R. C. Jennison and M. K. Das Gupta of the Nuffield Radio Astronomy Laboratories at Jodrell Bank near Manchester built the first radio interferometer: two radio telescopes electrically linked so that they can record features in a radio source too small to be distinguished by either telescope alone. Jennison and Das Gupta turned their interferometer toward the radio source Cygnus A, which had earlier been shown by Walter Baade and Rudolph Minkowski of the Mount Wilson and Palomar Observatories to be associated with a distant elliptical galaxy. Much to their surprise they discovered that the radio emission did not emanate from the galaxy itself. Instead it came from a diffuse patch on each side of the galaxy. Radio telescopes of increasing sophistication and power have now been employed to show that the majority of extragalactic sources detected at radio frequencies of less than one gigahertz (a billion cycles per second) have this basic double structure. Since 1953 one of the greatest challenges in extragalactic astronomy has been to uncover the reason. As we shall show, the discovery of radio jets brings the answer much closer.

Throughout the late 1960's and the 1970's a series of increasingly large interferometers were built, particularly at the University of Cambridge, at the Westerbork Observatory in the Netherlands and at the National Radio Astronomy Observatory (NRAO) in West Virginia. The most recent effort is the Very Large Array in New Mexico. It consists of 27 linked radio telescopes that are each 25 meters in diameter, and it can uncover features that subtend an angle in the sky as small as two-tenths of a second of arc. (That angle is subtended by a dime at a distance of 20 kilometers.) In 1971 George Miley and Campbell M. Wade used the NRAO interferometer to reveal the presence of hot spots in Cygnus A. Each such spot is a region emitting intense radiation at radio wavelengths. Typically it lies at the outer extremity of each lobe, or diffuse patch, in the brightest double radio sources (the ones like Cygnus A). Many of the brightest double radio sources show tails or bridges of low-intensity emission that extend backward from the hot spots toward the center of the source, where there is usually a compact focus of radio emission called a core. Most double radio sources are associated (like Cygnus A) with an elliptical galaxy or a quasar (a bright, pointlike object outside our galaxy); in such cases the core is invariably found at the source's optical center.

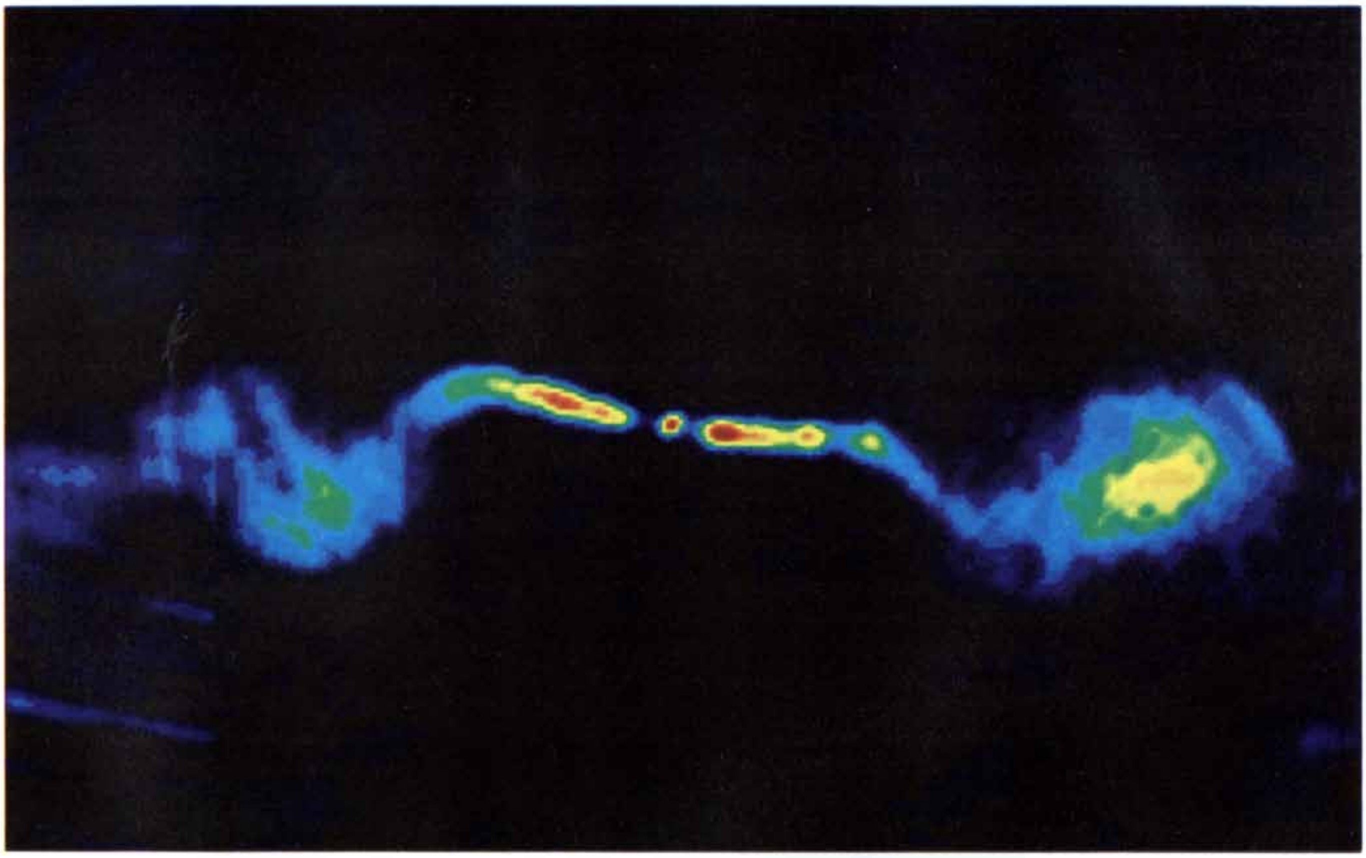
The resolution of maps made by radio interferometers is limited by the separation of the individual radio telescopes: the greater the separation, the greater the resolution. In the VLA the telescopes are separated by as much as 20 miles. In the technique called very-long-baseline interferometry (or VLBI), which was developed simultaneously with the construction of the modern generation of radio interferometers, the telescopes can be on different continents. The resolution that can then be achieved is correspondingly greater. Features as small as a thousandth of a second of arc can be distinguished. A thousandth of a second of arc is the angle subtended by a dime at a distance of 4,000 kilometers, by one light-year at the distance of the galaxy NGC 6251 and by 100 light-years at the distance of the farthest quasars.

This great resolution is achieved at a cost. The individual telescopes cannot be linked directly; hence the radio signals from the individual telescopes must be recorded on magnetic tape and compared well after the observations were made. This entails a loss of information, and so the maps that are made with very-long-baseline interferometry are neither as sensitive nor as detailed (on the scale of the observations) as those made with shorter-baseline interferometry. Still, astronomers have recently been successful at regularly using four or more radio telescopes in very-long-baseline work, and such combinations have gone a long way toward improving the results. The VLBI maps now being made are as good as the maps made with linked telescopes 10 years ago.

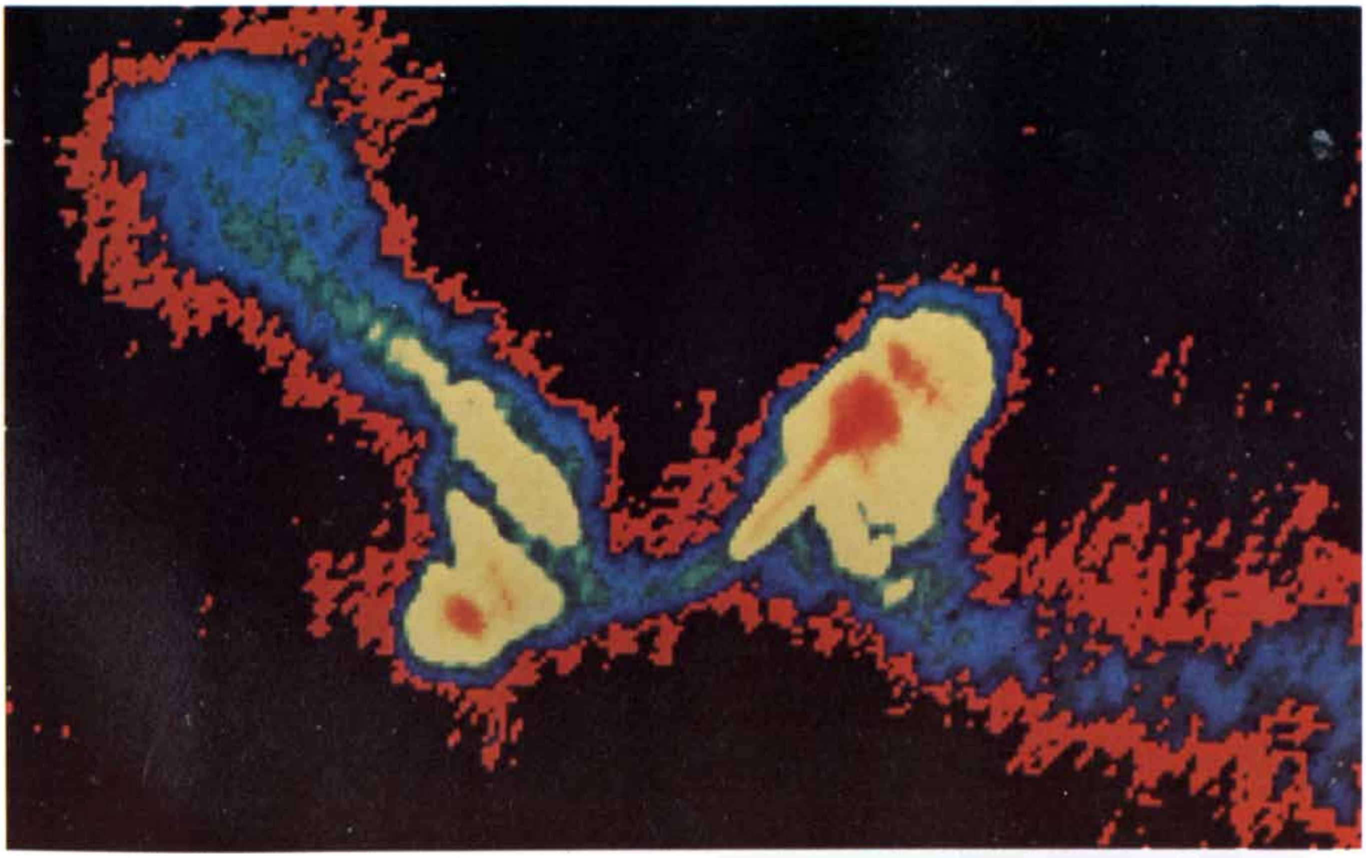
Very-long-baseline interferometry has been notably successful in probing the cores of radio sources. It has shown that they usually have features that remain unresolved. These features must



**ONE-SIDED JET some 400,000 light-years long is 300 million light-years away from our galaxy. It is the straightest known jet. Its origin (*lower left*) coincides with the center of the elliptical galaxy NGC 6251. Its origin also is known to coincide with a jet only three light-years long that is aligned with the longer jet. The radio emission from NGC 6251 and from the radio galaxies shown on the next page was mapped in arbitrary colors with the aid of the 27 radio telescopes that constitute the Very Large Array (VLA) near Socorro, N.M.**



MIRROR-SYMMETRICAL JETS in the radio source 3C 449 emerge in opposite directions from an elliptical galaxy. Each jet bends upward, then downward. Then each emerges into a large lobe of radio emission. It is thought the jets transfer power from the center of the galaxy to the lobes. The mirror-symmetrical bending of the jets may result from the orbiting of the galaxy around a companion.



INVERSION-SYMMETRICAL JETS in the radio source 4C 26.03 also emerge in opposite directions from an elliptical galaxy (NGC 326). Here one jet bends upward and the other bends downward. Again, however, the jets each terminate in an extended radio lobe. The inversion-symmetrical bending may result from the precession of the source of the jets, that is, the gyration of the source's spin axis.

therefore be smaller in angular size than a few thousandths of a second of arc. In contrast, the hot spots in the sources generally have no small-scale features. VLBI has also been successful in studying compact radio sources. Such sources are usually associated optically with distant quasars. In each of them the radio flux comes mostly from a small core, not from extended lobes. Nevertheless, radio emission is generally detected at a low level from a region (typically irregular in shape) surrounding the core. The compact radio sources are often found to vary in intensity on time scales ranging from months to years.

Jets have now been found both with linked interferometers and with very-long-baseline ones. They have also been found with a third technique, pioneered at Jodrell Bank, in which radio telescopes are linked by microwave transmissions. More than 70 extragalactic double sources are known to have large-scale jets emerging from the center of the associated galaxy, and the cores of at least six of these double sources show small-scale jets as well. Although a good sample of radio sources has not yet been properly surveyed, some suggestive trends have already emerged.

For one thing, large-scale jets are much more likely to be found in low-power double radio sources than in high-power double sources. Although this is partly due to the fact that the majority of powerful sources are at great distances, so that their jets would be hard to detect, distance cannot explain the trend entirely. For example, no large-scale jets have yet been found in Cygnus A, a powerful double source that is relatively close for such objects. (It is 450 million light-years away.) When a powerful double source does show a large-scale jet, the jet is generally one-sided and no counterjet is found. The jets associated with weaker radio sources are usually two-sided.

A second trend has emerged from the observation that the radio emission from extragalactic radio sources is invariably polarized, that is, the electric fields in the radio waves are directed preferentially along a line on the sky. The reason for this is known. The radio emission is taken to be generated by the synchrotron process, in which electrons moving with speeds close to the speed of light are accelerated by a magnetic field and therefore radiate electromagnetic waves. The electrons are accelerated in a direction perpendicular to the orientation of the field; thus the synchrotron radiation is polarized in that direction.

It follows that by mapping the pattern of polarization for a given radio source one can infer the orientation of the source's magnetic field. Moreover, by measuring the strength of the polarization one can estimate how well-ordered (or untangled) the magnetic field's geometry is. It turns out that in powerful jets the lines of magnetic field are strung out parallel to the direction of the jet. In weak jets the lines tend to be perpendicular. In jets of intermediate strength one often observes a region of transition from a parallel field in the jet near the center of the associated galaxy to a perpendicular field farther out along the jet.

Even before good examples of radio jets had been found in double radio sources a number of theorists (including us) had argued that the lobes of a double source must be continuously resupplied with energy. And since the radio emission from the core of the double source was clear evidence of continuing activity at the center of the galaxy with which the double source was associated, an energetic link between the center and the extended radio lobes seemed to be implied. The evidence therefore points strongly to the view that jets are streams of gas squirting out of the center of a galaxy. The jets would then supply the lobes not just with energy but with mass, momentum and magnetic flux as well.

After activity begins at the center the jets propagate outward through the galaxy rather like the jet of water that emerges from the nozzle of a hose. The cosmic jets pass first through the interstellar medium and then through the intergalactic medium. The density of the matter there ranges from roughly one hydrogen atom per cubic centimeter to one per million cubic centimeters. Still, the advancing jet must push the matter out of the way; hence the end of the jet moves slower than the gas flowing inside the jet. As a result energy accumulates at the end of the jet; this is the likely interpretation of a hot spot.

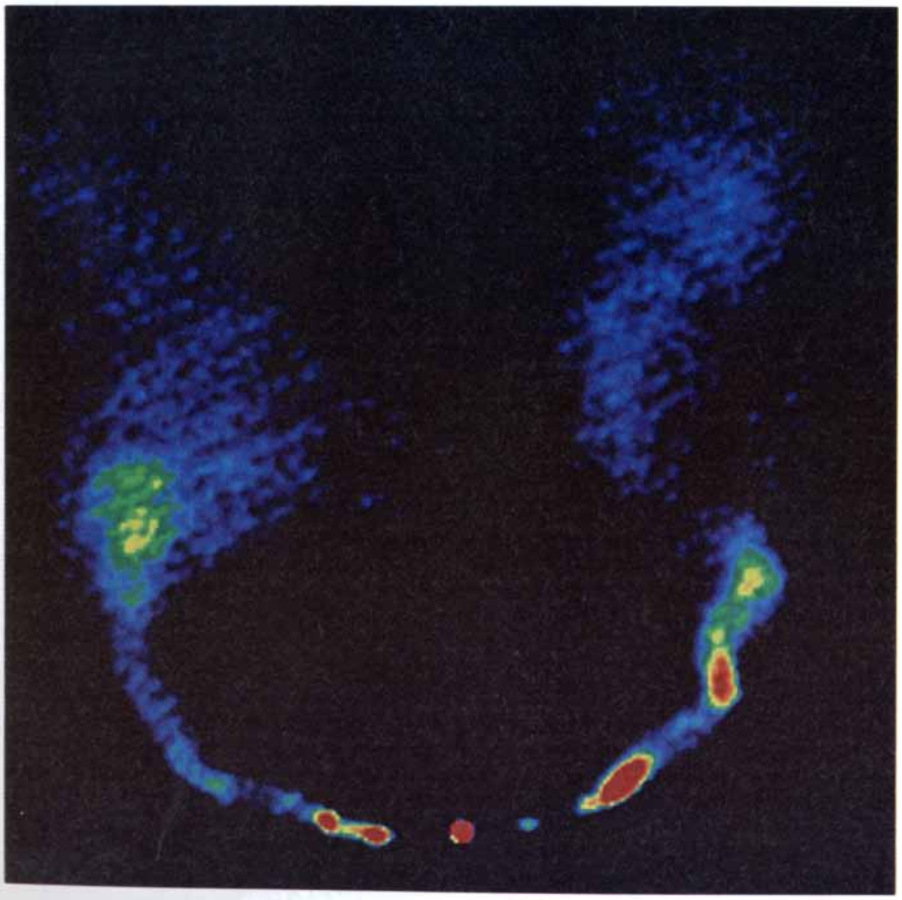
The flow of gas in the jet is thought to be supersonic, that is, the speed of the gas is faster than the speed of a sound wave in the gas. As the gas approaches the hot spot, however, it decelerates suddenly. This causes a shock wave to form across the jet. The effect of the shock is important. Before the jet reaches the shock wave most of its energy is in the form of ordered kinetic energy. Passage through the shock converts much of this ordered energy into two forms: the energy of relativistic electrons (electrons moving at speeds near the speed of light) and the energy of a magnetic field. It is entirely natural that the most intense radio emission in a double radio source should be generated where the jet is decelerated by the action of the surrounding gas.

After being stopped at the hot spots, the jet material flows back toward the galaxy. Thus it inflates the large lobes that can be seen in radio maps. The return flow carries with it the relativistic electrons and the lines of magnetic field. A given parcel of the jet spends a relatively short time (roughly 10,000 to a million years) in a hot spot. Most of the energy of the radio source is therefore contained in the lobes, which probably accumulate gas for 100 million years. The energy content of the lobes is prodigious. In a source such as Cygnus A it must be at least the amount that would be liberated if the mass of 100,000 stars were converted entirely into energy. In still larger sources it can exceed the mass equivalent of a million stars.

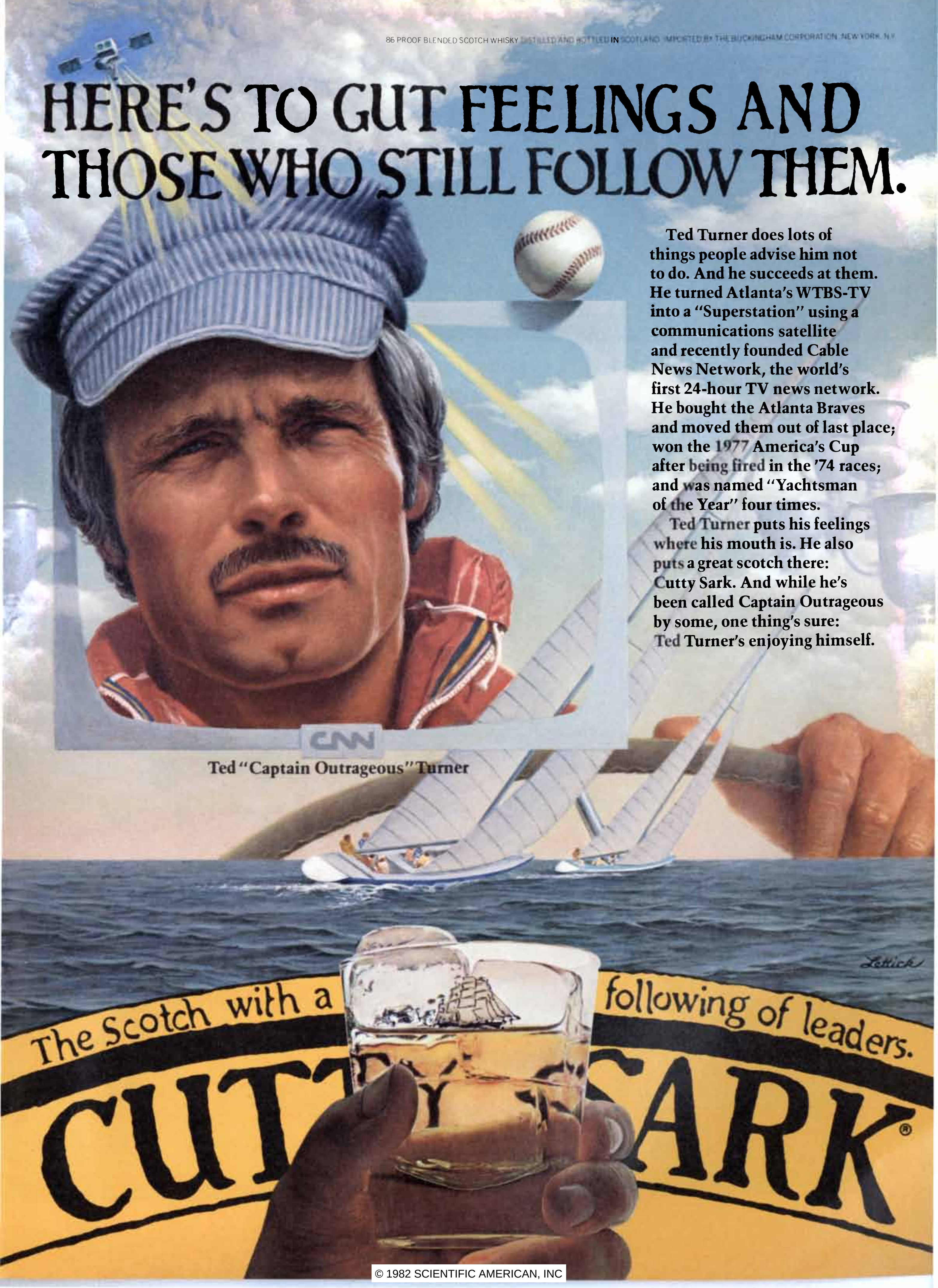
Many jets show large bends. In some cases this may be simply a manifestation of an instability that forms in the jet as it bores its way through the intergalactic medium. Anyone who has handled a flexible garden hose will be familiar with the phenomenon. If the hose is laid on the ground in a curve and the water is turned on, the hose will start to writhe. The flow of water in the hose is unstable because the water at the outside of a bend is traveling slightly faster to keep up with the water flowing at the inside of the bend. The pressure on the outside accordingly becomes less than the pressure on the inside, and the configuration of the hose changes in a way that increases its curvature. The physics is basically the same for cosmic jets, although the details are far more complicated if the flow is supersonic.

In a cosmic jet, however, not all bends are produced in a random manner. Consider the sources called radio trails. At low resolution the radio emission from such a source appears not to straddle an optical galaxy as it does in a typical double source but to extend in a long curve on one side of the galaxy. At high resolution the difference is explained. Two jets are seen emerging from the center of the galaxy, but instead of terminating at hot spots the jets are seen to bend steadily so that they both join up with the long, curving trail of radio emission.

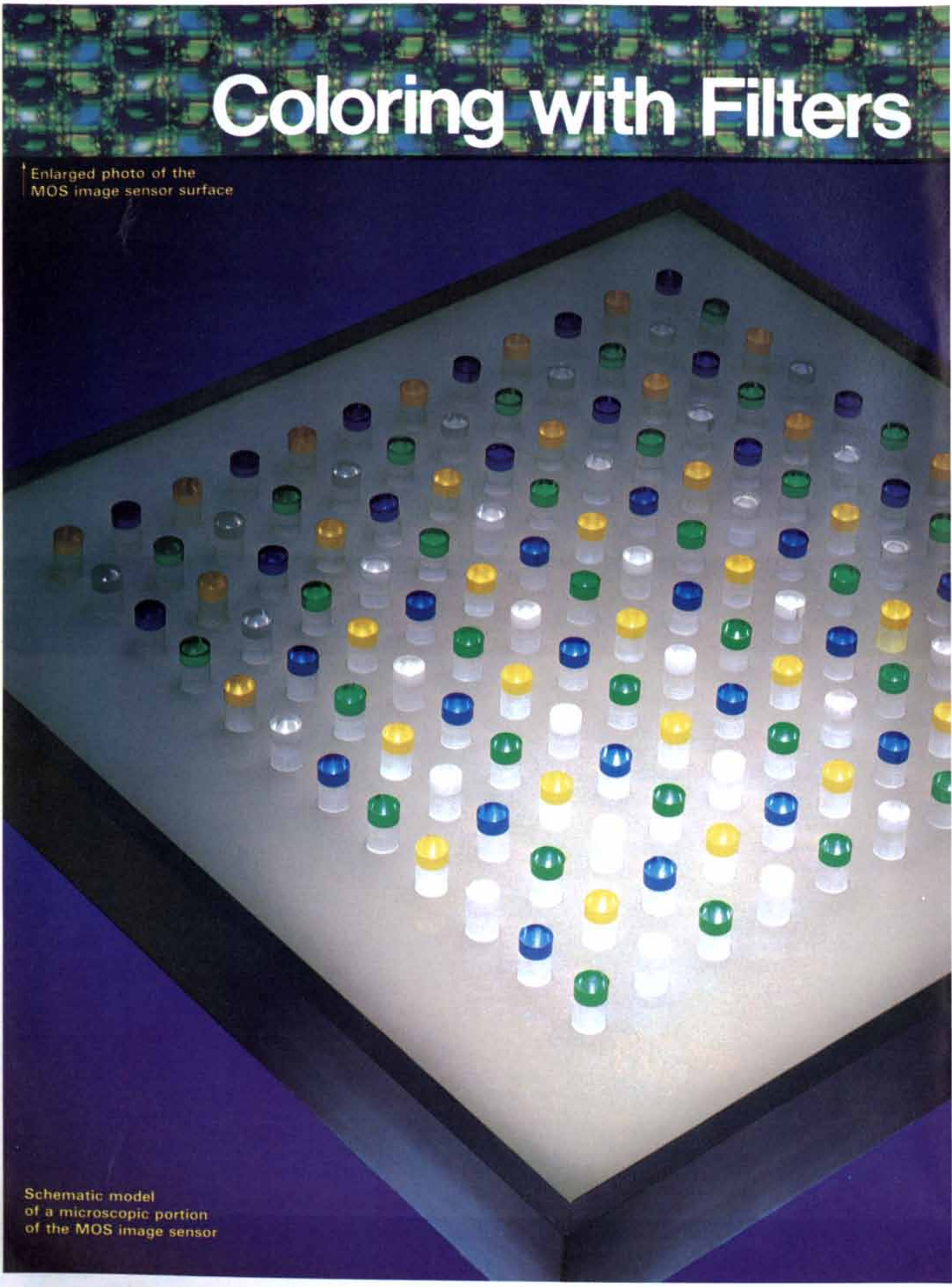
What apparently happens in radio trails is that two comparatively low-power jets are formed at the galaxy's center and are then blown sideways by the intergalactic medium as it rushes past the galaxy with speeds known to be several thousand kilometers per second. One is reminded of smoke that emerges from a chimney and then is caught in a high wind. The fact that radio trails are preferentially found in rich clusters of galaxies supports this interpretation of their morphology. (The space between galaxies in the rich clusters is thought to be filled with comparatively dense, hot, ionized gas.) What is most extraordinary about radio trails is that it appears their jets can be bent through almost a right angle by the intergalactic wind without losing their integrity. Thus it is clear that



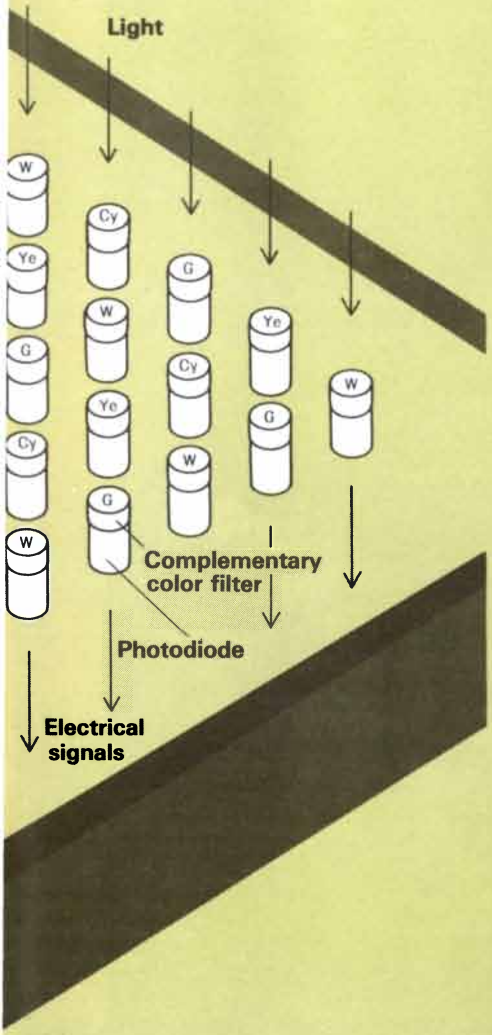
**RADIO TRAILS from NGC 1265, an elliptical galaxy in a rich cluster of galaxies, may represent jets blown into a long curve by intergalactic gas that passes the galaxy at a speed of several thousand kilometers per second. The map was made (in arbitrary colors) with the VLA.**



86 PROOF BLENDED SCOTCH WHISKY DISTILLED AND BOTTLED IN SCOTLAND. IMPORTED BY THE BUCKINGHAM CORPORATION, NEW YORK, N.Y.
HERE'S TO GUT FEELINGS AND THOSE WHO STILL FOLLOW THEM.
Ted Turner does lots of things people advise him not to do. And he succeeds at them. He turned Atlanta's WTBS-TV into a "Superstation" using a communications satellite and recently founded Cable News Network, the world's first 24-hour TV news network. He bought the Atlanta Braves and moved them out of last place; won the 1977 America's Cup after being fired in the '74 races; and was named "Yachtsman of the Year" four times.
Ted Turner puts his feelings where his mouth is. He also puts a great scotch there: Cutty Sark. And while he's been called Captain Outrageous by some, one thing's sure: Ted Turner's enjoying himself.
CNN
Ted "Captain Outrageous" Turner
Lettick
The Scotch with a following of leaders.
CUTTY SARK®


Coloring with Filters
Enlarged photo of the MOS image sensor surface
Schematic model of a microscopic portion of the MOS image sensor

# too Small to See




**Traditional color filters can't always do the job**

In color printing, a photographer uses red, green, blue-violet and pale yellow filters to separate a light image and produce negatives for the primary colors plus B&W. These negatives are in turn used to make four monochrome plates which re-create the original image when their colors are superimposed. But applied to video photography at the micron level, the conventional process of forming color filters on sheet glass for mounting on a photodiode matrix yields poor results. The photodiodes are so small that misalignment occurs and all of the colors are thrown off.

**Hitachi applies micrometric engineering skills**

When Hitachi developed its metal-oxide semiconductor (MOS) image sensor for video cameras, color separation filters had to be set in place for each of 186,240 photodiodes arranged on a silicon substrate less than 1 cm square. Since picture elements measuring just 20 μm each are too small to be seen with the naked eye, Hitachi engineers had to come up with micrometric techniques for forming the filters directly on the wafer-mounted photodiodes. A magnified picture of the MOS sensor surface reveals these tiny gem-like filters — yellow, white, green and cyan — arranged in a special mosaic configuration to provide improved resolution. The photodiodes read color signals indirectly from the filters as light passes through them, by employing an innovative system of color arithmetic whereby *white − cyan = red* and *cyan − green = blue*, etc.


**A better method of sensing colors**

Because of this tremendous technological breakthrough, Hitachi was able to perfect an MOS image sensor that not only provides life-like color but also eliminates the high-light lag and sticking that plagues conventional pickup tubes. What's more, the sensor is unaffected by electrical and magnetic fields, and image distortion is virtually non-existent. A few of the many applications to which this MOS device can be applied include equipment monitoring in factories, power plants and substations, dimensional measurement, and aircraft and space development, as well as videotaping.


**Technical excellence is embodied in all Hitachi products**

The development of color filters for the MOS image sensor is just one case demonstrating Hitachi's technological strength. You'll find other examples in all of Hitachi's products, from multicolor liquid crystal displays and cathode ray tubes to a whole host of high-quality electronic appliances. Our comprehensive technical expertise is your guarantee of convenience, easy operation and high reliability in every product that bears the Hitachi brand.

**HITACHI**

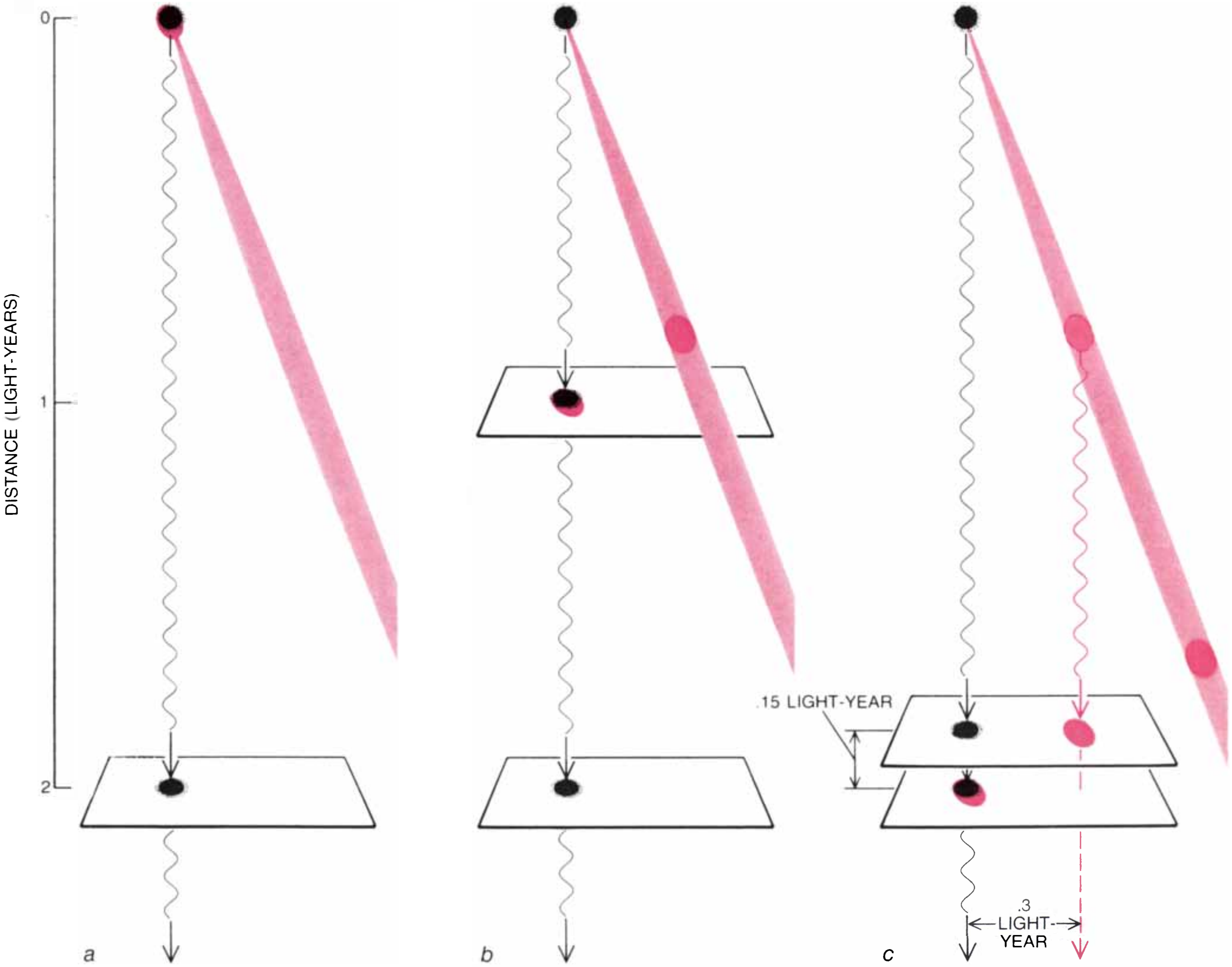
A World Leader in Technology

jets are not always unstable. After the jets merge with the trail their gas is presumably brought to a speed matching that of the intergalactic medum. Given, then, the velocity of the medium with respect to the galaxy and knowing the length of the trail, one can measure the age of the radio source. In a typical case the age turns out to be 300 million years.

A related explanation has been advanced for sources such as 3C 449, whose jets show several sharp bends but are basically linear. In general these sources are not found in rich clusters of galaxies. On the other hand, they usually have close companion galaxies. It is likely, therefore, that the radio source corresponds to a galaxy in orbit around its companions. If the velocity of the gas in its jets is not much greater than the orbital velocity, the motion of the galaxy will be "written" on the sky in the form of bends in the jets.

In a sense the bends are an illusion: the gas in the jets does not flow through them. Imagine a photograph of someone watering a garden by swinging a hose. The water hanging in the air has a zigzag pattern, but a given drop of water in the pattern is moving quite ballistically. The shape of the bends in a cosmic jet can be used to infer the three-dimensional positions of the galaxies from their two-dimensional projection on the sky. Still, the bends could be consistent with an orbiting galaxy only if the bends in one of the two jets are related to those in the other by a mirror reflection. If one jet bends to the right on the sky, the other jet must bend to the right at about the same distance from the center of the galaxy. The required mirror symmetry does in fact appear to be present in sources such as 3C 449.

A different class of radio sources exhibits inversion symmetry. In this case a bend to the right in one jet corresponds to a bend to the left in the other. The best example of inversion-symmetrical jets is SS433. Here the source of the jets is thought to be precessing like a top. Again the individual parcels of gas in the jets move in straight lines, but the pattern that should be observed at any one time is that of a corkscrew. Recent radio observations made with the Very Large Array by Robert M. Hjellming of the National Radio Astronomy Observatory and Kenneth J. Johnston of the U.S. Naval Research Laboratory appear to show just that. Perhaps the extragalactic jets that display inversion symmetry come from a source that is precessing just like SS433 but at the center of a galaxy. Such precession would




**SUPERLUMINAL EXPANSION (the apparent motion of components of a radio source at speeds exceeding the speed of light) is thought to be an illusion. Still, it suggests that those components can have speeds near the speed of light. In this example an observer two light-years from a radio source (*a*) is unaware that the source has emitted in its jet a bright knot of ionized gas (*color*). After a year (*b*) the observer still sees a single source; the light showing the emergence of the knot has a light-year to go to reach him. Meanwhile the knot has moved away from the source at an angle of 20 degrees to the line of sight. Its speed is .9 times the speed of light. After two years (*c*) the observer sees the knot emerge. The light the knot emitted in *b* is only .15 light-year behind. Hence in .15 year the observer will see that the knot has moved to .3 light-year from the source. He may mistakenly judge that the knot is moving at twice the speed of light.**

be an important clue to the conditions at the center that give rise to the jets.

We shall now turn our attention to the small-scale jets discovered by very-long-baseline interferometry. Within the past three years several compact radio sources and several of the compact cores associated with extended radio sources have been mapped with sufficient sensitivity to reveal a small-scale jet. In most such cases the jet turns out to have a characteristic morphology. It is a single jet emerging from an unresolved point of intense radio emission. In many instances the single jet points in the approximate direction of one of the source's radio lobes. If an undetected jet points in the opposite direction, it must be at least 10 times fainter than the one that is observed. Readhead, Cohen and their colleagues at Cal Tech have discovered an important correlation bearing on the small-scale jets' alignment. The ones associated with the compact cores of extended radio sources (that is, the sources whose emission comes mostly from double lobes) tend to be well aligned with the lobes. The misalignment is by no more than a few degrees. In contrast, the small-scale jets associated with compact radio sources are noticeably curved, so that they are misaligned with the faint extended structure of the source, typically by angles of 30 degrees or more.

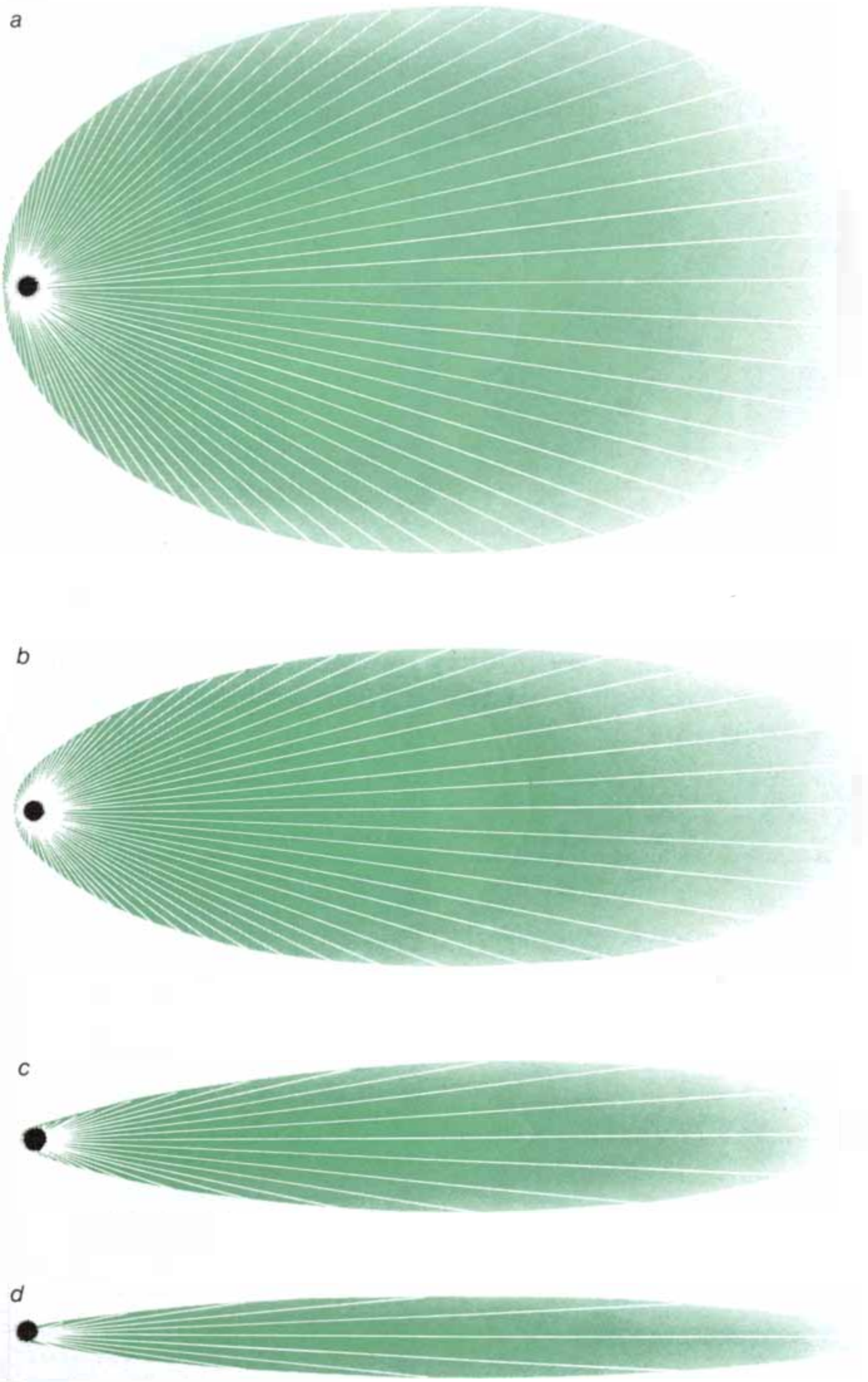
In attempting to interpret this correlation and offer an interpretation of radio sources quite generally, two phenomena are crucial. The first one, called superluminal expansion, has been suspected for some 10 years and has now been rather convincingly demonstrated on recent VLBI maps. Basically superluminal expansion means that features in certain VLBI jets have been observed to move on the sky with respect to the position of the unresolved core from which the jet emerges. The features move at a speed that seems to be in excess of the speed of light. The most convincing case of superluminal expansion has been documented by Timothy J. Pearson and his colleagues at Cal Tech. It involves the quasar 3C 273, which is famous among astronomers for having an optical jet some 60,000 light-years long. The core of 3C 273 emits a VLBI jet, which in accordance with Cohen and Readhead's generalization is bent and misaligned with the optical jet. Over the past four years a bright patch in the VLBI jet has increased its separation from the core by nearly 50 percent. On the hypothesis that 3C 273 is 1.5 billion light-years from our galaxy the bright patch appears to be moving with a speed five times the speed of light.

Astronomers do not contend that the patch is really moving faster than light. That would contravene the special theory of relativity, and there is no need to contemplate such a drastic revision of physics. Instead it is thought we are witnessing an illusion. The bright patch appears to be moving faster than light because the light from the patch and the light from the core of the quasar, which serves us as a benchmark, travel along paths of different length and therefore take slightly different amounts of time to reach us. The precise arrangement is still controversial, but the simplest possibility requires that the bright patch be moving at a small angle to the line of sight between the radio source and observers on the earth with a speed slightly less than the speed of light. It will then seem to move superluminally. The important point is that superluminal expansion suggests not speeds faster than the speed of light but speeds not much less than the speed of light for the components of radio sources.

The second phenomenon that must be appreciated before one can analyze radio sources is relativistic aberration. Suppose a hunter wants to shoot a duck when it is directly above him. He must point his shotgun upward and fire before the duck passes overhead. The shotgun pellets will travel vertically upward to the duck, which will have moved forward to meet them (the hunter hopes). Now consider all of this from the vantage of the duck. It sees the hunter moving toward it, and the pellets, instead of moving vertically, have a horizontal component in their motion. To put it another way, the pellets travel with a slight aberration that inclines their trajectory along the direction the hunter is moving with respect to the duck.

The same thing happens with photons, the particles of electromagnetic radiation. Hence a cloud of plasma that radiates photons equally in all directions will appear to be shining preferentially along its direction of motion. When the plasma is moving relativistically (that is, almost as fast as the photons), the effect is quite pronounced. Take the case of the plasma whose relativistic motion gives the illusion of superluminal expansion in a radio source. Here roughly half of the photons the plasma emits will be radiated within a cone facing into the direction of the plasma's motion with an opening angle of only five to 10 degrees. Furthermore, the photons within the cone will have been made more energetic by a Doppler-shift increase in their wavelength. The net result is striking. If an observer is in the cone, the source will look brighter than a stationary source, typically by a factor between 100 and 1,000. Conversely, if the observer is well outside the cone, the source will be effectively invisible.

We can now advance a powerful unifying interpretation of extragalactic radio sources. Suppose most sources consist of a pair of relativistic jets emerging in opposite directions from the heart of a galactic nucleus. If the jets are more or less aligned with the direction from the source to the earth, we detect the jet that is pointed more or less toward us. The other jet is invisible. The source appears bright and one-sided. The jets end in a pair of extended radio lobes, which radiate photons roughly equally in all directions. They produce the faint halo of extended emission that is seen to surround a compact source. One does not expect the jets to be precisely uniform and straight. Doubtless they exhibit some natural bending and inhomogeneity. In the compact radio sources the bending is exaggerated because our line of sight is almost colinear with the jet. If the inhomogeneities move along the jet at relativistic speeds, they will appear to be moving superluminally.



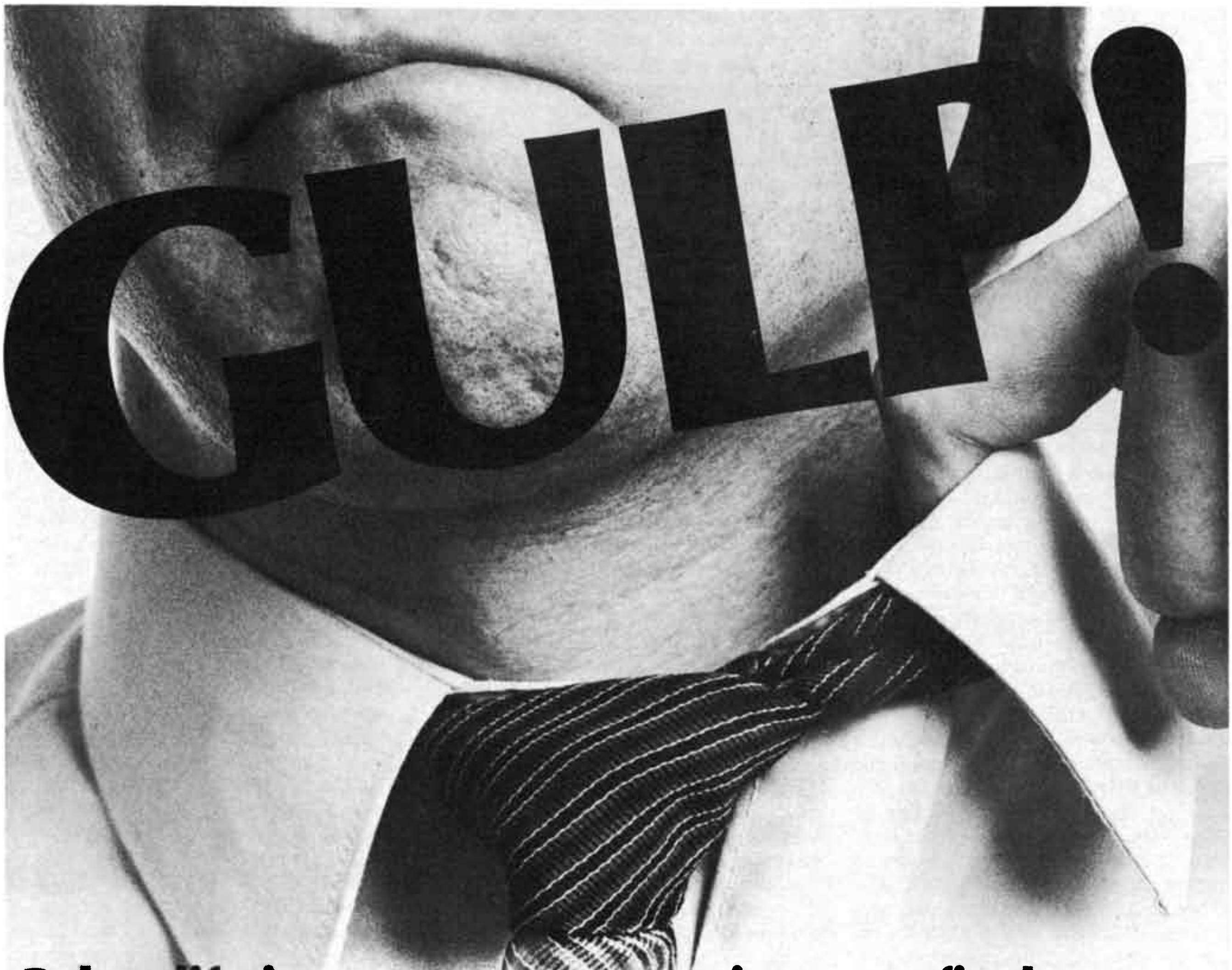

**RELATIVISTIC ABERRATION focuses the radiation emitted by an object moving at a speed approaching that of light, so that the object radiates intensely in the direction of its motion. In *a* the emitter (a cloud of ionized gas that intrinsically radiates equally in all directions) is moving toward the right at half the speed of light. In *b* it is moving at .75 the speed of light, in *c* at .94 the speed of light and in *d* at .98 the speed of light. The emitter is rendered invisible except from in front of it. The shape of each pattern shows only the way the intensity of the radiation varies with the angle of emission. Seen from directly in front, however, the emitter in *a* is seven times brighter than a stationary source radiating equally in all directions, the one in *b* is 30 times brighter, the one in *c* is 440 times brighter and the one in *d* is 3,100 times brighter.**

On this interpretation the extended radio sources are observed at large angles to the direction of their jets. For the most powerful of these sources, the ones such as Cygnus A, the jets are relativistic, so that they are effectively invisible. For sources somewhat less powerful, such as the radio source associated with the galaxy NGC 6251, the jets are mildly relativistic. Nevertheless, the aberration of their radio emission is sufficiently pronounced for only the jet on our side of the center of the radio galaxy to be detected. For weak sources such as 3C 449 the jets are subrelativistic, so that both jets are equally apparent.

The "Doppler beaming" interpretation of extragalactic radio sources accounts for so many diverse observations that it is unlikely to be totally wrong. It must be said, however, that the interpretation is difficult to reconcile with certain observations. For example, the density of the gas in some of the one-sided jets may turn out to be so high that if its speed were relativistic, the power carried by the jet would exceed by a factor of 10,000 the amount of power apparently dissipated by the jet and the source's radio lobes.

In some sources, therefore, it appears that one-sidedness is not the result of relativistic motion. An alternative hypothesis is that the visibility of a jet is related to its stability. In the strongest sources the jets would be stable, invisible conduits that efficiently transport energy from the core to the radio lobes. The jets, however, could be rendered unstable. The instability would readily channel the jets' kinetic energy into the acceleration of relativistic electrons and a greatly enhanced radio emissivity. One-sided jets would arise when only one of the jets is rendered unstable. Two-sided jets would be associated with low-power sources in which the transport of energy is inefficient. An observational test that will discriminate between the competitive explanations of one-sidedness in radio sources will be to determine in one-sided sources whether the radio lobe on the side of a one-sided jet is systematically different from the lobe on the side that seems to lack a jet. If this turns out to be the case, the Doppler-beaming explanation will be hard to maintain.

One difficulty plaguing the investigation of cosmic jets is that it is not known how much total power the jets are radiating. We know only how much they radiate at radio wavelengths. The problem is that radio telescopes are far more sensitive than optical telescopes. A few




# Other life insurance companies may find Northwestern Mutual's new Extra Ordinary Term policy a little hard to swallow.

Until now you had two choices. Whole life or term insurance.

Until now.

Recently, Northwestern Mutual Life has unveiled an entirely new life insurance plan.

We call it Extra Ordinary Term.

You might just call it the answer.

Unlike most term policies, our Extra Ordinary Term plan provides level premiums plus the cash-building advantages of whole life. Never before has Northwestern Mutual Life offered such an innovative insurance plan. But innovation such as this is nothing new for a company that for the past 125 years has been a leader in low-cost long-term life insurance. A record that is the envy of the industry.

And a record that has been maintained by constantly improving the products we offer you. Our new Extra Ordinary Term plan is just another reason why other life insurance companies have come to realize that we're a tough act to follow.

**Northwestern Mutual Life**
The Quiet Company

## A tough act to follow

jets have been detected at optical or X-ray wavelengths. This means their optical power must be at least 100 times greater than their radio power. We cannot rule out the possibility that the radio sources are generally most luminous at optical frequencies.

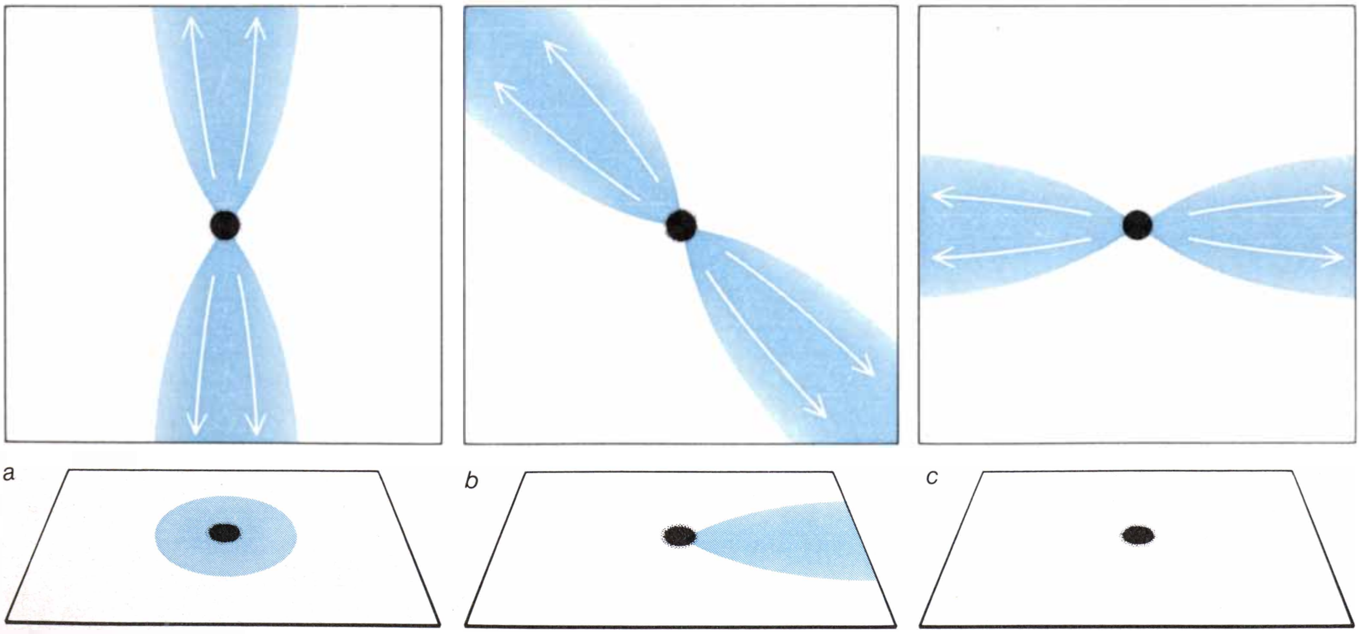
A further difficulty is that it is hard to tell whether a jet delineated by radio emission is identical with the region where ionized gas is flowing. If much of the jet's true cross section is invisible, such quantities as the power and thrust of the jet will tend to be underestimated. The estimates made today present an intriguing puzzle. Some of the most luminous jets appear to become collimated as they move away from the galaxy that emits them. What, then, squeezes the jet? Is it the pressure of the gas that surrounds the jet in the intergalactic medium? Then the heating and collisions in the gas should make the region around the jet a powerful X-ray source. Is it loops of magnetic field that are carried along with the jet? A property of curved magnetic field lines in the jet is that they exert a tension, so that a loop of magnetic field wrapped around a jet can act like a rubber band. One finds it worrisome that the strongest jets, the ones most in need of confinement, have magnetic fields that tend to be parallel to the axis of the jet rather than perpendicular to it. Still, this problem can be overcome if only the core of the jet is sufficiently luminous to be observable.

To a theoretician the most important question about the jets is what makes them. A rich variety of schemes have been suggested, but none of them commands universal support. One possibility is that jets are made by a nozzle. Imagine a copious supply of hot gas in the middle of a cooler large cloud. The hot gas may result from the presence of exotic astronomical objects such as black holes or pulsars that heat the gas when it falls in their intense gravitational field. Imagine that the cool cloud is spinning. The hot gas will be buoyant and will tend to escape along the directions of least resistance, which lie along the cloud's axis of rotation. The cloud will provide flexible walls for the outflow, and as the gas flows outward its pressure will decrease to match the external pressure. This can lead to an increasingly well-collimated jet.

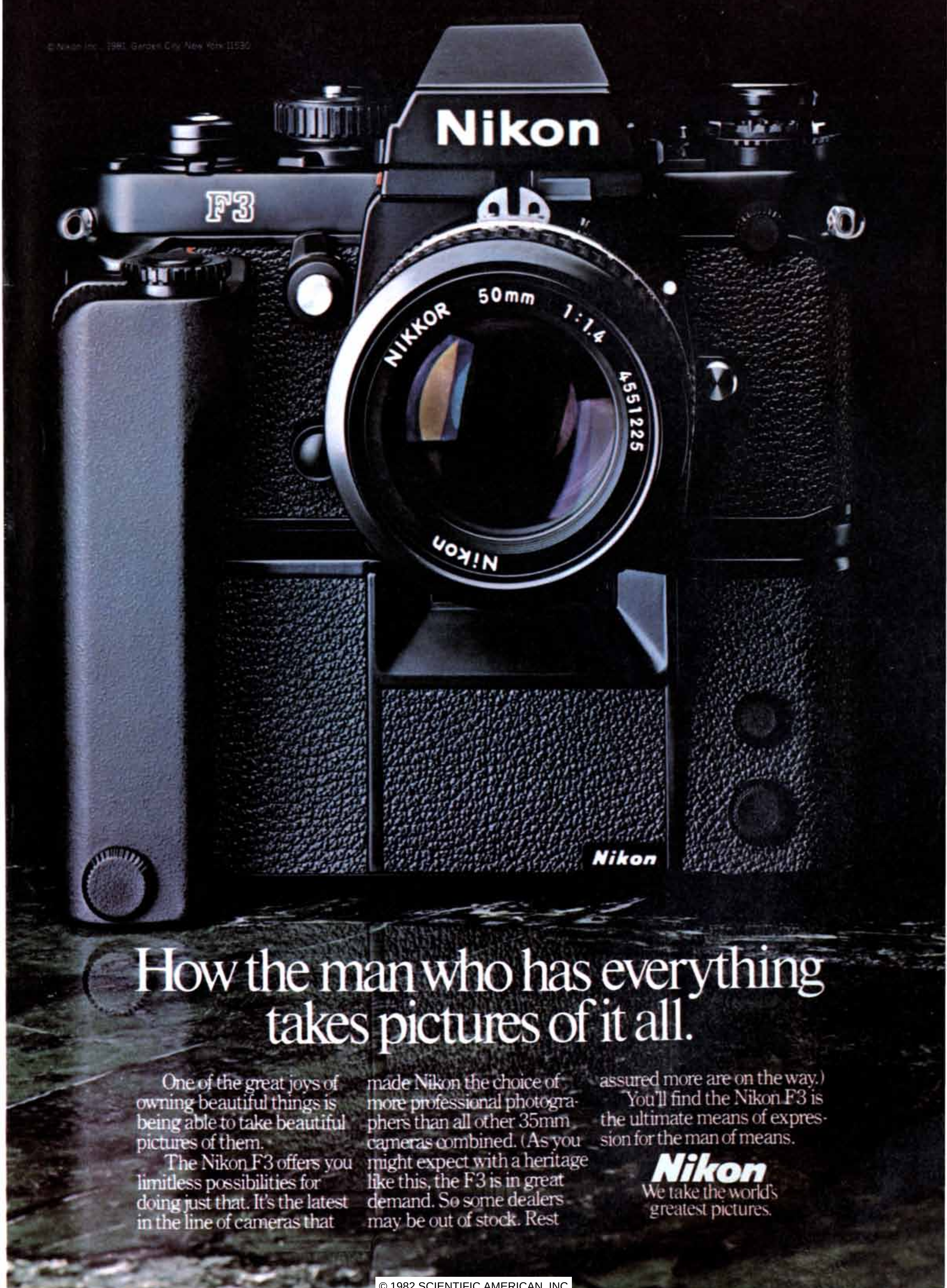
To be sure, the hypothetical mechanism of a nozzle is in conflict with the observations made of several radio sources. For one thing, it seems likely that the VLBI jets are produced in regions no more than a few light-years in size. Yet if a gas cloud were dense enough to confine the pressure of a jet in such a small place, the cloud would be a spectacular source of X rays. Furthermore, the alignment of the VLBI jet and the large-scale jets in the extended radio sources suggests that the collimating mechanism must be a good gyroscope: it must have a long "memory" for a fixed direction in space. It is not clear that a gas cloud can be a good gyroscope.

These difficulties are alleviated if the collimation is assumed to happen in the region surrounding a black hole whose mass may lie anywhere between a million solar masses and several billion. The presence of such a black hole at the center of radio galaxies is supported by several independent items of circumstantial evidence. First, some of the most active radio sources show day-to-day changes in their optical flux. This suggests that the flux comes from a region no larger than the distance light can travel in a day. That distance is only 10 times greater than the radius of a black hole whose mass is a billion solar masses. At X-ray energies great changes in flux in less than an hour have been reported. Second, a group of optical astronomers led by Wallace L. W. Sargent and Peter J. Young at Cal Tech have investigated the light from the center of the galaxies M87 and NGC 6251. They have argued that the distribution and the spectrum of the light indicate the presence of a compact mass exceeding a few billion solar masses. Third, a black hole represents the deepest possible gravitational "well." Therefore a black hole is best able to effect a high-efficiency conversion of the energy of infalling gas into radiant or kinetic energy.

If the black hole is spinning, it will distort the space around it by an effect that can be predicted by means of the general theory of relativity. The distortion will cause the gas surrounding the hole to precess around the spin axis of the hole. The gas at different distances will precess at different rates. As a result two things will happen. First the viscous drag between neighboring rings of gas




**DIFFERENT ORIENTATIONS of cosmic jets may give them different appearances because of relativistic aberration. Here each radio source consists of two jets emerging in opposite directions from the center of a galaxy. The center radiates equally in all directions, but the gas in the jets moves fast enough for the jets to radiate preferentially in the direction of the motion. If the jets are parallel to the line of sight (*a*), a distant observer perceives a core of radio emission surrounded by a radio halo. If the jets are almost parallel to the line of sight (*b*), the observer detects only the core and the nearer jet; the far jet is invisible because little of its radiation is directed toward the observer. If the jets are perpendicular to the line of sight (*c*), the observer detects only the core because aberration makes both jets invisible.**


Nikon
F3
NIKKOR 50mm 1:1.4 4551225
Nikon
Nikon
How the man who has everything takes pictures of it all.
One of the great joys of owning beautiful things is being able to take beautiful pictures of them.
The Nikon F3 offers you limitless possibilities for doing just that. It's the latest in the line of cameras that made Nikon the choice of more professional photographers than all other 35mm cameras combined. (As you might expect with a heritage like this, the F3 is in great demand. So some dealers may be out of stock. Rest assured more are on the way.)
You'll find the Nikon F3 is the ultimate means of expression for the man of means.
Nikon
We take the world's greatest pictures.

Many happy returns.
IMPORTED
Canadian Club
Blended Canadian Whisky
C.C.
FEDERAL
TAX
REFUND
United States Treasury
PAY
TO THE
Canadian Club
"The Best In The House".
6 Years Old. Imported in Bottle from Canada by Hiram Walker Importers Inc., Detroit, Mich. 86.8 Proof. Blended Canadian Whisky. ©1982

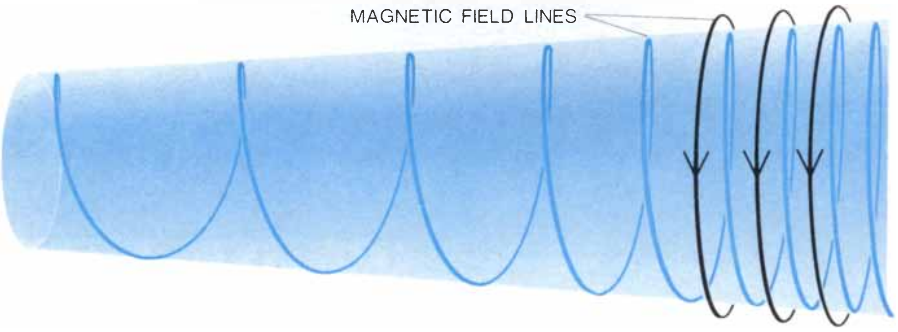
will make the gas settle into a disk in the equatorial plane of the hole. Then the same viscous drag that caused the disk to form will make matter drift slowly inward through the disk. The drift will feed jets with mass and will also power the jets by the release of gravitational energy. Presumably the jets will be launched in the directions parallel to the black hole's spin axis. Since the spinning black hole has a great deal of angular momentum, these directions are fairly stable.

The structure of the accretion disk will be governed by the rate at which gas is supplied by viscous drag. If the rate of supply is large, the pressure of the radiation produced in the innermost part of the disk will puff up the innermost part. As a result a thick torus, or donut-shaped volume, of gas will surround the black hole. The surface of the torus near the spin axis of the hole will define two narrow funnels in each of which a jet can be collimated. Each jet would be driven outward through its funnel largely by the pressure of the radiation field inside the funnel and perhaps in part by mechanical energy brought to the surface of the funnel by convection in the gas inside it. One difficulty persists. A disk supported by radiation pressure implies an optical and ultraviolet luminosity far greater than is generally observed at the center of the galaxies associated with double radio sources. There is nonetheless reason to think that in SS433, the jet-emitting object within our galaxy, the gas supply is very large. A torus supported by radiation remains a viable explanation for the production of jets by that source.

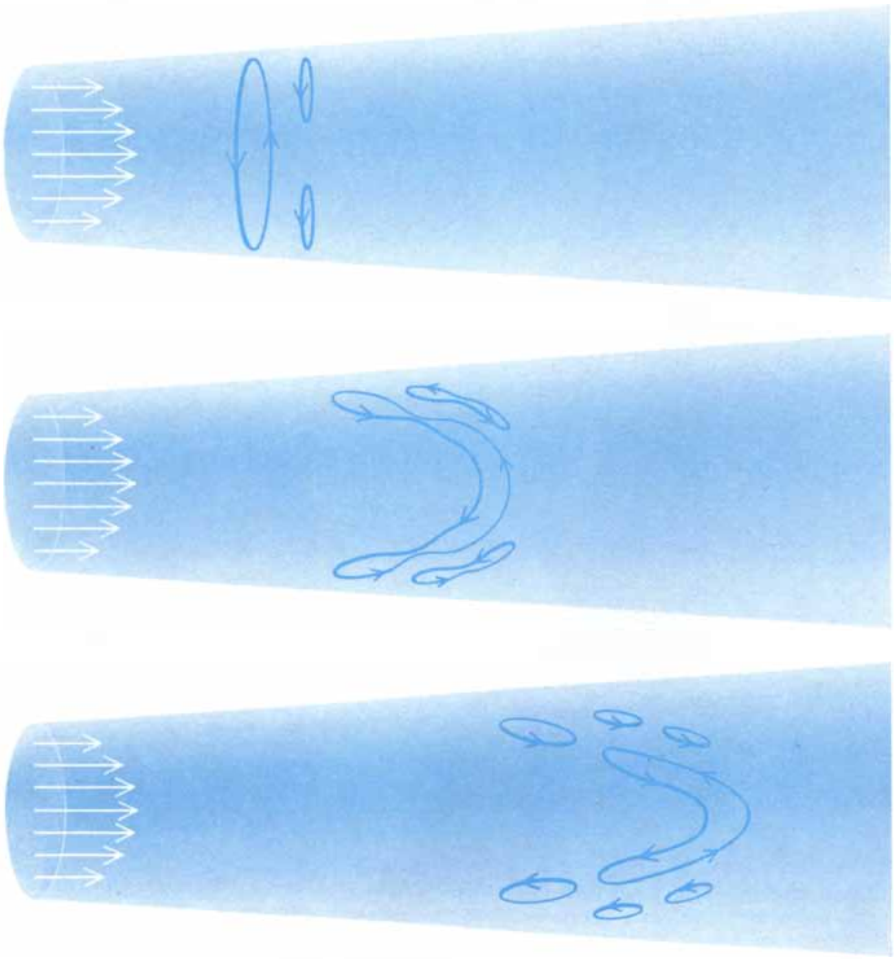
A variant on the radiation-supported torus is the gas-supported torus, which can be produced at a low rate of gas supply. Here the radiation generated in the innermost part of the disk escapes freely and cannot be responsible for inflating it. Nevertheless, the viscous friction in the disk is great enough for the gas to liberate gravitational energy by sinking toward the black hole faster than it can radiate energy. This means the gas gets hotter as it falls. The heating causes it to fill a toruslike volume.

The power for the jets need not be derived from the infalling gas but may be drawn from the black hole itself. In particular a significant quantity of mechanical energy can be extracted from a spinning black hole. Suppose the gas flowing toward the hole has a magnetic field. Eventually some of the field lines penetrate the surface of the hole. The lines become loosely attached to the hole, and as the hole spins it drags the field lines along with it. These lines can therefore do work on the matter near the black hole. Indeed, the field lines attached to the hole are rather like ropes around a windlass. As for the black hole itself, it acts like a giant flywheel that is gradually being spun down.

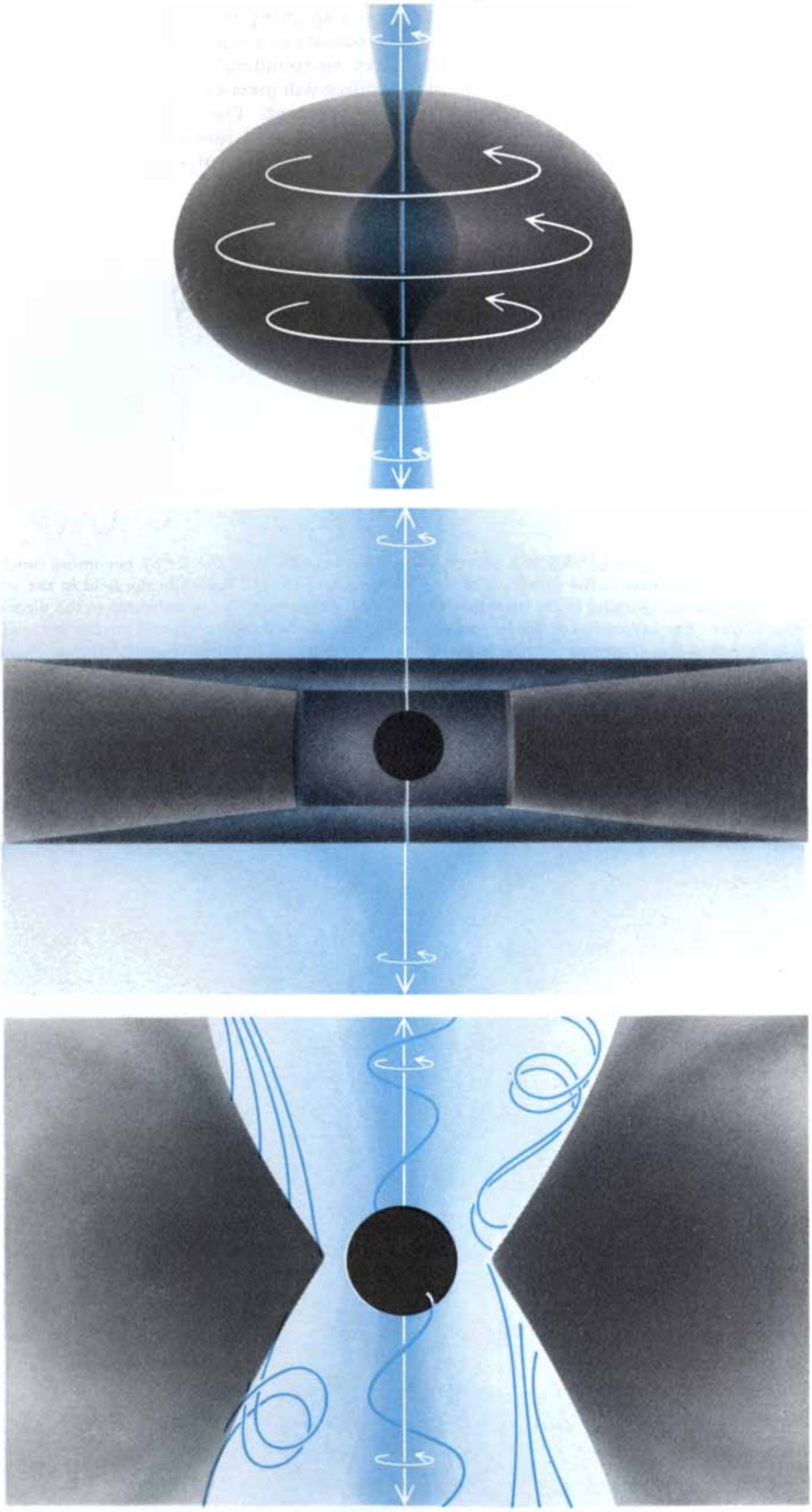
We have not yet touched on how the cosmic jets are fueled. Fundamentally the most powerful radio sources require at least a few stellar masses of gas each year unless the source is powered in part by a spinning black hole. The normal processes of stellar evolution in the galaxy surrounding the core of the radio source will meet a certain amount of this requirement. The best available estimates suggest, however, that they fall short by a factor of at least 10. Besides,




**MAGNETIC FIELD LINES in a jet respond to the expansion of the jet by becoming more nearly perpendicular to the direction of flow of the gas in the jet. Basically the field in the jet has a component parallel to the direction of flow and a component perpendicular to the direction of flow. The expansion of the jet spreads the former more than the latter. Thus the loops of a helical field line (representing the superposition of a parallel field and a perpendicular field) can incline toward the perpendicular with distance from the core of the radio source. Perpendicular loops of the field can exert a tension on the jet; they may thus arrest the jet's expansion.**



**SHEAR IN THE JET has a different effect on the jet's magnetic field: it makes the parallel field dominant. The shear, caused by lack of uniformity in the velocity of the gas in the jet (*white arrows*), elongates the loops in the perpendicular field. The elongated regions may ultimately break away from their parent loops and close up to become small isolated loops.**



**JETS CAN FORM** by any one of several hypothetical mechanisms. The nozzle mechanism (*top*) assumes that a rotating cloud of gas surrounds a buoyant cloud of hot gas. The hot gas escapes along the cloud's axis of rotation. The accretion-disk mechanism (*middle*) assumes that the gravitational field of a supermassive black hole collects a gaseous disk. The jets, powered by the pressure resulting from the gas's electromagnetic radiation, are launched along the spin axis of the hole. The accretion-torus mechanism (*bottom*) assumes that radiation pressure (or the pressure of the hot gas) puffs up the inner part of the accretion disk around a black hole. The jets are powered by this pressure, by convective energy in the puffed-up gas or by the spinning black hole, which gives energy to the jets by dragging magnetic field lines through them.

the violent activity in the core will probably blow gas away. A different suggestion is that the strong gravitational fields at the center of a galaxy will create a dense cluster of stars. Indeed, the stars there may be packed together some 10 million times more densely than the stars near the sun. Such clustering means that stars are much likelier to collide. In addition stars can be torn apart by tidal forces if they come too close to a central supermassive black hole. Still, detailed calculations have shown that it is hard for the required cluster of stars to build up and persist.

A third possibility has observational support in the form of galaxies with double or even multiple centers. It is a process that has been termed cannibalism by Jeremiah Ostriker of Princeton University. Here the center of a "cannibal" galaxy consumes a smaller companion galaxy: a "missionary" galaxy. The outer parts of the missionary are quickly stripped away by tidal forces. The inner parts prove more difficult to digest. Thus the inner parts sink toward the cannibal's center. The idea that cannibalism fuels jets is attractive because radio galaxies are commonest at great distances. They were commoner, therefore, in the past, and specifically at times when galaxies were closer together and likelier to interact. A galaxy was 1,000 times more likely to be a radio source when the universe was a fourth its present age. Moreover, it is known that among nearby galaxies the ones that happen to be in places where the density of galaxies is high are far more likely to be radio sources than the ones that are isolated.

If radio sources indeed are fueled by cannibalism, an explanation of the inversion-symmetrical jets emerges. Suppose both the missionary galaxy and the cannibal galaxy have a central supermassive black hole; then the missionary's black hole will eventually settle into a binary orbit around the cannibal's. Suppose too the missionary's black hole maintains jets that are collimated along the hole's spin axis. Normally the hole would point its spin axis along a fixed direction in space. Now, however, the gravitational field of the companion black hole will distort the space in its neighborhood. As a result a distant observer would find that the spin axis precesses. If the spin axis precesses, so do the jets.

Thanks mainly to the construction of the large linked interferometers and the simultaneous development of very-long-baseline interferometry, radio astronomers have made great strides toward elucidating the structure of the double radio sources, including their cosmic jets. On the interpretive side, however, we remain embarrassingly ignorant. We lack a clear idea of the com-

# WHERE DO BLACK DOCTORS COME FROM?


**Nearly half of all black doctors practicing in the U.S. today are graduates of Meharry.**

**Sure, there are black** medical students at Harvard and Stanford and Johns Hopkins. But no single private medical school in the U.S. produces as many black physicians as Meharry Medical College.

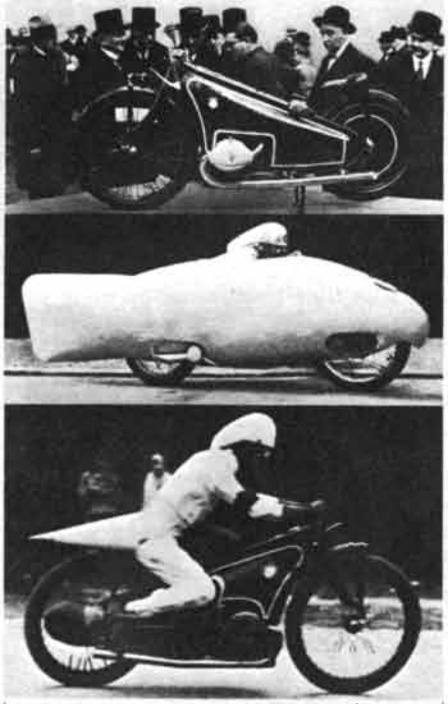
Ninety percent of our students require financial aid to some degree. To make more funds available for scholarships and student loans, and to maintain the quality of our training, our goal is to raise $27 million.

**Meharry must depend** largely on private and corporate contributions. We are eligible for only limited funding under Federal and State programs—and, as a professional school, receive no resources from the United Negro College Fund.

**Won't you help?** Because without Meharry, a lot of qualified youngsters couldn't even consider a medical career.

**OPEN WIDE FOR MEHARRY**
MEDICAL COLLEGE

Send your tax-deductible contribution, in any amount, to **Meharry Medical College, Second Century Fund, Nashville, TN 37208.**
The need is urgent, and your support is appreciated.



TOP: BMW is the rage of the 1930 Paris Motor Show.
CENTER: In 1937 BMW again builds the fastest motorcycle in the world, clocking 279.5 kph.
BOTTOM: BMW's set 9 world records in 1932.

resorting to needless complexity."

Indeed, the design of the BMW twin has grown not increasingly complex with the years but ingeniously simple—the hallmark of a brilliantly engineered machine.

It is the simple genius of the opposed twin which makes a BMW significantly less expensive to maintain and repair than most motorcycles. And for that matter, far less likely to need repair in the first place.

With the cost of a motorcycle mechanic now hovering at about $24 an hour; with the price of a mere tune-up of certain multi-cylinder motors upward of $100, the BMW twin is becoming more appreciated by the day.

Its rewards, however, are not purely financial.

THE JOYS OF BEING ONE OF THE FEW MOVING PARTS ON A BMW.

Cycle Guide flatly states that "no engine configuration known to man is better at aiding the low 'cg' required in a flick-left-flick-right bike than the opposed twin."

Small wonder then that BMW's are renowned as extraordinary handling machines.

For, in contrast to motorcycles encumbered by the top-heaviness of their own engines, the BMW's cornering prowess is actually enhanced by its motor design.

Providing its rider with a control so exhilarating as to inspire the motorcycle editor of AutoWeek to write of one BMW:

"Long after a clumsy Japanese hyperbike would have wrestled you into exhaustion, the R 100 RS urges you to uncharted apexes; the only limit to your enthusiasm is the amount of gas in the bike's 5.8-gallon fuel tank."

1982 MOTORCYCLES THAT WON'T BE DATED BY 1983.

There is an inescapable irony in how quickly so-called "motorcycles of the future" become things of the past.

While BMW's, machines whose basic engine design has remained unchanged for over a half-century, continue to endure.

It is this timelessness of design, of course, that marks the difference between a motorcycle that is perceived as a classic. And one which simply is not.

A perception which, on any given day, can be measured with the dollars-and-cents yardstick of the NADA's Used-Motorcycle Handbook. Where, over the past 3 years, BMW's have retained the astonishing average of 77.7% of their original purchase price.

Which provides additional proof that the best way to determine a motorcycle's future is to simply look at its past.

**THE LEGENDARY MOTORCYCLES OF GERMANY.**

BMW

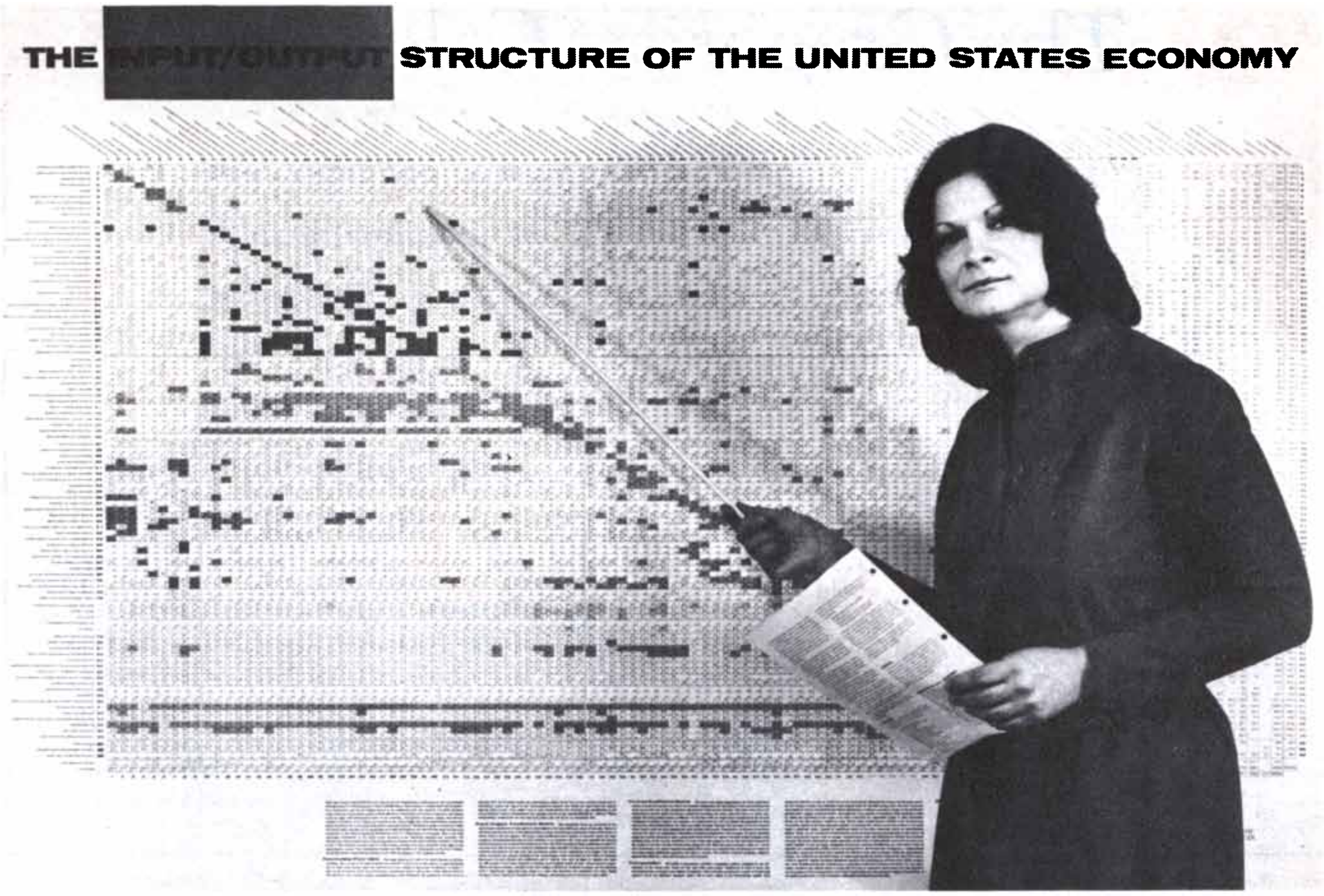
position of the jets, of how fast they are moving, of how they are confined and of why they are so stable. Most important of all, we do not really understand how they are made. We have only a wealth of hypotheses. Some are ingenious, some are improbable but many seem quite reasonable.

The pervasiveness of cosmic jets suggests that the raw materials of a deep gravitational well, a supply of gas and some spin are all that are needed for a galactic nucleus to expel its exhaust in opposite directions. Jets are nonetheless heterogeneous. Some appear to be fast, others slow. Some appear to be powerful, others feeble. Some are long, others short. Perhaps theorists have been trying too hard to find one mechanism that accounts for everything. A diverse set of processes might find counterparts in nature and always lead to cosmic jets.

What are the prospects for progress in the study of cosmic jets? On the observational side the prospects are bright. The comparison of extragalactic jets with the jets in our galaxy should yield insight into which aspects of the production and propagation of jets are independent of scale. In addition to SS433 the X-ray source Scorpius X-1 (only 600 light-years away) shows clear evidence of double structure. Moreover, a number of clouds of interstellar gas and dust have offered tantalizing hints that jets may arise in conjunction with the formation of stars.

So far the Very Large Array has mapped only a small fraction of the radio sources accessible to it, and only a very few such sources have been mapped at the highest available sensitivity and resolution. Moreover, the technique of very-long-baseline interferometry has now come of age. The next stage is to construct large arrays of radio telescopes that are dedicated to VLBI. At optical wavelengths one can look forward to the launching of the Space Telescope in 1985. This instrument, which will have more than 20 times the resolving power of comparable telescopes on the ground, will reveal much about quasars and may discover several more jets at optical wavelengths.

On the theoretical side much work remains. Detailed calculations of the gas flow around massive black holes can be done with computers. Such studies are handicapped, however, by our ignorance of the microscopic properties of plasma under cosmic conditions. The experimental approach may prove very fruitful. In particular, aerodynamicists routinely experiment with jets of gas. Typically the jets they make have a velocity no more than a few times the speed of sound in them. Greater speeds may be needed. It is an exciting prospect that the strange and beautiful shapes in the radio sky might be reproduced in a wind tunnel.




# WHAT MAKES THE U.S. ECONOMY TICK?

The editors of SCIENTIFIC AMERICAN have prepared a wall chart displaying for the 1980's the Input/Output Structure of the U.S. Economy based on the latest interindustry study from the U.S. Department of Commerce.

The SCIENTIFIC AMERICAN Input/Output wall chart does for economics what the table of elements does for chemistry. It answers at a glance questions about the linkage between the microeconomics of the firm and the macroeconomics of the system; about the web of technological interdependencies that tie industry to industry; about the industry-by-industry direct and indirect consequences of swings in public and private spending; about the impact of change in technology, and about any other topic you can think of. You are rewarded by surprise as well as by confirmation of your hunches. For teaching and practical and theoretical studies, here is a powerful, graphic tool.

In the familiar format of the SCIENTIFIC AMERICAN Input/Output wall charts for the 1960's and 1970's, the wall chart for the 1980's measures 65" × 52" and is printed in eight colors. Each of the nearly 10,000 cells in the 97-sector interindustry matrix shows (1) the interindustry commodity flow, (2) the direct input/output coefficient and (3) the "inverse" coefficient. Where the direct input/output coefficient exceeds .01, the cell is tinted in the color code of the industrial bloc from which the input comes. This device, combined with triangulation of the matrix, brings the structure of interindustry transactions into graphic visibility.

A supplementary table displays, industry by industry, the capital stock employed; the employment of managerial, technical-professional, white-collar and blue-collar personnel; the energy consumption by major categories of fuel, and environmental stress measured by tons of pollutants.

The editors of SCIENTIFIC AMERICAN are happy to acknowledge the collaboration, in the preparation of this wall chart, of Wassily Leontief, originator of input/output analysis—for which contribution to the intellectual apparatus of economics he received the 1973 Nobel prize—and director of the Institute for Economic Analysis at New York University.

Packaged with the chart is an index showing the BEA and SIC code industries aggregated in each of the 97 sectors.

**SCIENTIFIC AMERICAN**

415 Madison Avenue, Dept. I/O, New York, N.Y. 10017

YES, send me the new SCIENTIFIC AMERICAN 1980's INPUT/OUTPUT wall chart—printed in eight colors on heavy paper measuring 65" × 52".

I understand my chart(s) will be shipped prepaid, uncreased*, flat-rolled and in a sturdy mailing tube. (Allow 4-6 weeks for delivery)

I enclose a check for $____ for ____ chart(s) at $35 each, saving me $5 each in handling charges.

Bill me for $____ for ____ chart(s) at $40 each.

Name ________________

Company ________________

Street Address ________________

City, State, Zip Code ________________

*Charts shipped international require one fold. 5/82

Make check payable to SCIENTIFIC AMERICAN. Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York.

# The Gregorian Calendar

*Pope Gregory XIII began the modern calendar 400 years ago in order to correct the accumulating drift in the Julian calendar and keep Easter in the spring*

by Gordon Moyer

In 1582 Pope Gregory XIII introduced the calendar named in his honor; it is the system of reckoning days currently in civil use throughout the world. Before the inception of the Gregorian calendar Western civilization depended on a system instituted by Julius Caesar, called the Julian calendar. For more than 16 centuries this system remained in service in spite of an accumulating discrepancy between the mean length of its year, 365.25 days, and the tropical year, the average interval between successive transits of the apparent sun through the vernal equinox. By 1582 the error in the Julian system had grown to approximately 11 days. This defect was of principal concern to the pope; if the Julian calendar had continued in service, Easter would eventually have been celebrated in the summer.

Pope Gregory assembled an eminent body of astronomers, mathematicians and clergymen for the reform of the Julian calendar. The commission faced a fundamental problem inherent in all civil calendars: for obvious reasons a calendar designed for ordinary service must have a whole number of days; it cannot simply leave a fraction of a day dangling at the end of the year. One of the leading members of the commission, the Jesuit astronomer Christoph Clavius, put it succinctly: "Annum civilem necessario constare ex diebus integris" ("Civil years necessarily consist of integral days"). This basic requirement is the root of all the difficulties in the construction of an accurate calendar. It is also primarily why the creation of a perfect calendar, one that would never err by even a single day, is impossible.

In principle the Gregorian calendar is merely a slightly altered version of the Julian calendar. The commission decreed that 10 days would be eliminated from the year 1582 in order to restore the date of the vernal equinox to March 21; over the many centuries the date had regressed to March 11. To curb the drift of the equinoctial date a plan was adopted whereby in three out of every four centurial years—for example 1700, 1800and 1900—the leap day that would have been added in the Julian system would henceforth be omitted. These revisions, promulgated in the papal bull of February 24, 1582, set off a wide barrage of polemics, incited furious debate among scholars and prompted the man in the street to question whether birds would now know when to fly south for the winter. One merely has to leaf through the great *Bibliographie Générale de l'Astronomie* published in 1887 by Jean-Charles Houzeau and Albert-Benoit Lancaster to gain an idea of the enormous number of treatises that were written for and against the reform.

The controversy was as much religious as it was academic. This was the age of the Reformation; Protestant countries rejected the new calendar, denouncing it as a papal scheme to bring their rebellious fold back under the jurisdiction of Rome. The accusation was not entirely unfounded. Gregory XIII was a vigorous—in fact, a ruthless—promoter of the Counter Reformation. He assisted Philip II of Spain in his scourging of the Protestant Dutch and gleefully accepted the head of the leader of the French Huguenots after the infamous St. Bartholomew's Day Massacre, a blood bath that the pope celebrated as a Catholic victory, ordering a medal struck in its commemoration. Gregory probably saw it as an opportune time to impose a calendar reform on the Christian world, which he did by threatening to excommunicate anyone who refused to accept it.

Opposition to the calendar was not entirely a consequence of the religious strife of the day. Many learned men acknowledged the need for calendar reform (the accumulating error of the Julian calendar had been observed for centuries), but they were not persuaded that the Gregorian system was a significant advance over the "Old Style." Indeed, the preeminent mathematician François Viète, often called the father of modern algebra, condemned the reform as a corruption of the Julian calendar.

Leading scientists of the 16th century, Viète among them, argued that the Gregorian calendar was astronomically unsound. This opinion was shared by two of the bitterest critics of the reform, Michael Maestlin and Joseph Justus Scaliger. Maestlin, an astronomer, was one of the first to openly espouse the Copernican theory and is famous for having been Johannes Kepler's professor at Tübingen. Scaliger, also a renowned academic, was an extraordinary scholar; fluent in a dozen languages, he was a classicist, historian, philologist and chronologer. Colleagues called him a "sea of sciences," a "bottomless pit of erudition." He became an archenemy of Clavius, the principal defender of the Gregorian reform.

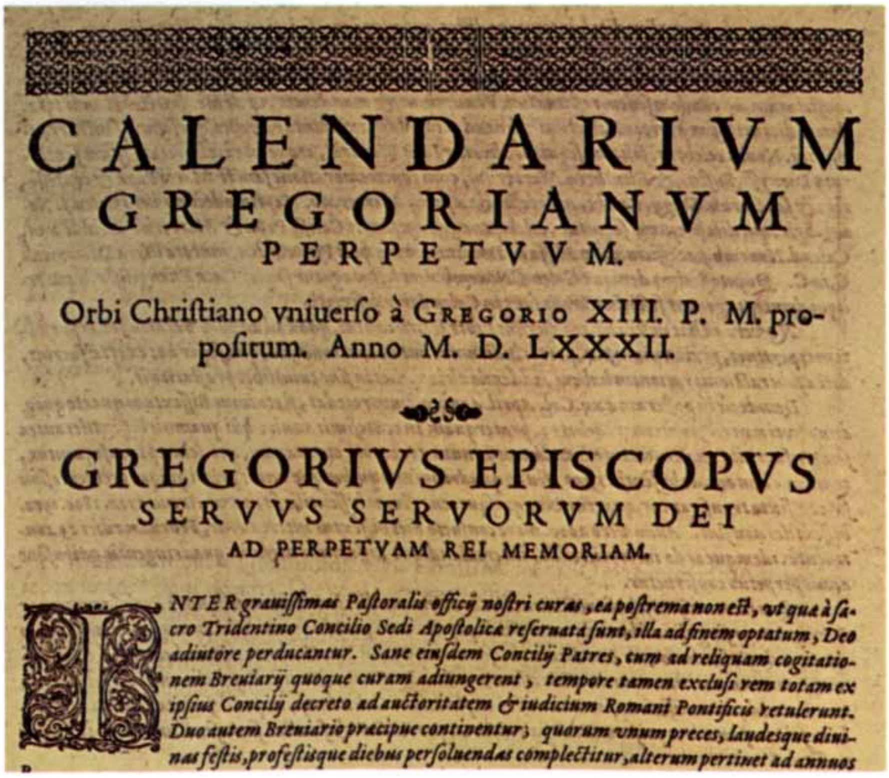
In 1595 Scaliger published a commentary on the *Canon Paschalis* of Hippolytus, a fourth-century work on the reckoning of the date of Easter. To it he appended a pungent criticism of the Gregorian calendar that is more than twice the length of the main work. Scaliger argued for his own plan of reform, which in fact would have produced a slightly more accurate calendar but a somewhat more complex one.

Clavius was quick to respond to Scaliger's pronouncements. In the same year he published a stinging rebuttal, *Iosephi Scaligeri elenchus, et castigatio calendarii Gregoriani.* Scaliger's criticisms from his work on the *Canon Paschalis* and Clavius' responses are presented in alternating paragraphs. Clavius' remarks are frequently amusing in their vehemence, but they are seldom more outrageous than Scaliger's invectives. He lashed out at the Bavarian Clavius, calling him a "German fat-belly" and a "beast."

Scaliger raised objections to both the civil and the ecclesiastical parts of the calendar. He maintained that the tables utilized in computing the date of Easter, called epact tables, were erroneous, and he protested that the reformed rules for observing leap years were not sufficient

**DETAIL OF A PAINTING** emphasizes the reason for the reform of the Julian calendar. A member of the commission appointed by Pope Gregory XIII in 1576 to study the need for reform is shown pointing to a chart indicating the 10-day error astronomers determined had accumulated in the Julian calendar since the time of the First Christian Council of Nicaea in A.D. 325. The upper arc in the diagram represents a segment of the calendar year divided into days, the lower arc a corresponding section of the tropical year (the year measured by successive passages of the sun through the vernal equinox). The zodiacal signs of Libra and Scorpio indicate the month of October, from which the 10 days in excess of the mean sun were to be removed. In the full painting the pope is shown sitting on a throne to the left. The painting is by an unknown artist whose work was commissioned in 1582, the year of the reform, by Scipio Turamini, a member of the municipal council of Siena. The photograph was made in the Archives of the City of Siena, which still has the full painting.

**POPE GREGORY XIII** is commemorated by a statue atop the entrance to the Palazzo Púbblico in Bologna. The portal was designed by the 16th-century architect Galeazzo Alessi to commemorate the pope's election because Gregory was born in Bologna. He became pope on May 13, 1572, and 10 years later instituted the reform of the 1,600-year-old Julian calendar.



CALENDARIVM
GREGORIANVM
PERPETVVM.

Orbi Chriſtiano vniuerſo à GREGORIO XIII. P. M. propoſitum. Anno M. D. LXXXII.

GREGORIVS EPISCOPVS
SERVVS SERVORVM DEI
AD PERPETVAM REI MEMORIAM.

*INTER grauiſſimas Paſtoralis officij noſtri curas, ea poſtrema non eſt, vt quæ à ſacro Tridentino Concilio Sedi Apoſtolicæ reſeruata ſunt, illa ad finem optatum, Deo adiutore perducantur. Sane eiuſdem Concilij Patres, cum ad reliquam cogitationem Breuiarij quoque curam adiungerent, tempore tamen excluſi rem totam ex ipſius Concilij decreto ad auctoritatem & iudicium Romani Pontificis retulerunt. Duo autem Breuiario præcipue continentur; quorum vnum preces, laudesque diuinas feſtis, profeſtiſque diebus perſoluendas complectitur, alterum pertinet ad annuos*

**PAPAL BULL** of February 24, 1582, ordered Christians throughout Europe to adopt the Gregorian calendar on October 15, 1582 (or in 1583). The bull was reprinted in Tome V of the *Opera mathematica* of Christoph Clavius, a Jesuit astronomer who was a leading member of the papal commission studying the calendar reform and later the principal defender of the reform. This photograph shows the first page of the bull as it appears in the Clavius volume.

to keep the date of the vernal equinox consistently on March 21. Indeed, the calendar does embody a small fluctuation in the equinoctial date, amounting to a few days. Even so, the calendar succeeds in keeping the date on or near March 21 every year and will continue to do so for many centuries.

Maestlin, like Scaliger, did not approve of the Gregorian system for determining Easter. The Council of Nicaea in A.D. 325 had decreed that Easter should be celebrated on the same date by all Christians. It was decided by the church that the date should be the first Sunday after the 14th day of the moon (approximately the full moon) that falls on or next after the date of the vernal equinox, which at the time of the council was assumed to be fixed at March 21. The cumbersome rule is still applied today in establishing the date of Easter and subsequently all other movable feasts. Computing the Easter date is unquestionably the most intricate operation in the calendar. Even Carl Friedrich Gauss, who as a young man worked on the problem, did not succeed in developing a complete algorithm for deriving the date.

Easter can fall on any Sunday from March 22 to April 25 in any year. Incredibly, it takes 5,750,000 years for all the dates of Easter to repeat in the same order. The table of epacts is employed to determine the age of the moon on January 1 for every year in a cycle of 7,000 years. Since the average interval between one new moon and the next, called a lunation, is a little more than 29.53 days, the maximum number of epacts is 30 (29.53 rounded to the nearest whole number).

Knowing the moon's age on January 1, it is a simple matter to ascertain the dates of all the new and full moons throughout the year. It is then quite easy to find the 14th day of the moon occurring on or after March 21. The Sunday after the 14th day is Easter. The cycle of epacts, however, is only an approximation to the true or astronomical moons; an epact can differ by as many as three days from the actual lunar phase, although its usual deviation is only a day or two. The prime advantage of an artificial lunar cycle is its simplicity. The true motion of the moon is much too complex to reduce to a simple and convenient table, one that could be readily understood by clerics trained in the computus, the science of reckoning Easter practiced throughout the Middle Ages and the Renaissance.

Clavius was quite satisfied with close approximations. He repeatedly urged the need for simplicity in the calendar, noting that it was absurd to think everyone could be trained in astronomy. Nevertheless, Maestlin advocated the use of

ROMANI
CALENDARII
A GREGORIO XIII.
P. M. RESTITVTI
Explicatio
S.D.N. CLEMENTIS VIII.
P.M. IVSSV EDITA.
Auctore
CHRISTOPHORO CLAVIO
BAMBERGENSI SOCIETATIS IESV.
Accessit confutatio eorum, qui Calendarium aliter instaurandum esse contenderunt.

**CHRISTOPH CLAVIUS was portrayed in 1606, when he was 69, in an engraving (*left*) by Franciscus Villamoena. Clavius was a distinguished mathematician who staunchly defended the Gregorian reform. The title page of Clavius' definitive work on the calendar is shown at the right. The work was commissioned by Clement VIII, who became pope in 1592; he was the fifth pontiff following Gregory.**

exact astronomical calculations in deriving the date of the paschal (Easter) full moon. Kepler, unlike his former professor, supported Clavius' reasonably accurate method of following the moon, declaring that "Easter is a feast, not a planet."

Largely through his tireless defense of the calendar Clavius earned for himself an honored place in the history of science, so much so that the largest crater on the moon is named for him. In all he wrote five defenses of the calendar. The first, published in 1588, was an answer to criticisms made by Maestlin. Clavius' definitive calendrical work, however, is an 800-page tome, *Romani calendarii a Gregorio XIII P. M. restituti explicatio,* first published in 1603. Much of his defense against Maestlin is repeated, often verbatim, in this later treatise.

Clement VIII, who became pope in 1592, had enjoined Clavius to explain and justify the new calendar to the Christian world. The *Explicatio* is the copious result. It is an imposing technical work. The Edinburgh scholar Alexander Philip, the author of one of the few substantial texts on the calendar in English, wrote that Clavius, "like a cuttle-fish," actually clouded his subject in "the ocean of ink with which he surrounded it." There is some merit to the criticism; Clavius is at times long-winded and repetitious, but he is seldom inaccessible, and other scholars have remarked on his readability. He was, in fact, one of the best-selling scientific authors of the late Renaissance.

All Clavius' treatises on the calendar were reprinted and collected in Tome V of his *Opera mathematica,* which was published in 1612, the year he died. Tome V also includes a reprint of the *Kalendarium Gregorianum perpetuum,* the official publication on the reform issued by the church in 1582 and written by the calendar commission under Pope Gregory. Also reprinted is the papal bull, "*Inter gravissimas . . .*" ("In the gravest concern . . .") of February 24, 1582, announcing to all the Christian princes that the new system would officially begin on October 15 of that year. It prescribed the removal of 10 days from the calendar: "In order therefore to restore the vernal equinox to its former place, which the Fathers of the Nicene Council put at XII Kalend. Aprilis [March 21], we prescribe and command as concerning the month of October in the year 1582 that 10 days inclusive from the 3rd Nones [October 5] to the day before the Ides [October 14] be taken away." Clavius and the other members of the commission avoided a disruption of the days of the week in this expulsion of days; Thursday, October 4, 1582, the last date in the Julian calendar, was immediately followed by Friday, October 15.

In his defense against Maestlin, Clavius explained that there was nothing mysterious about the choice of month; October simply included the fewest religious feasts and therefore an omission in that month was the least disruptive for the church. He also described it as the month that presented the fewest problems for business.

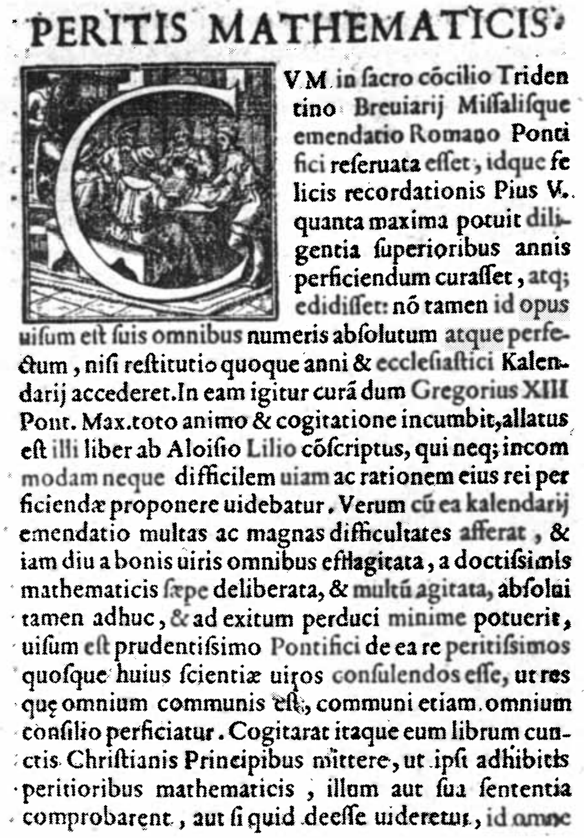
The elimination of 10 days was not an astronomical necessity. The date of the vernal equinox could just as well have remained at March 11. The real problem was in preventing whatever date was chosen from continuing to drift. How, then, did the plan enacted by Pope Gregory succeed in keeping the date on or about March 21? That is the central technical question to be asked in regard to the reform. Although the major concern of Gregory's commission had been the restitution of the date of Easter, that objective could not be achieved until the date of the equinox had been stabilized.

It is conceivable that no reform would have taken place in 1582 if the paschal celebration had not been at issue. The error of approximately 11 days that had built up in the Julian calendar between the time of the Council of Nicaea and 1582 was still a minor discrepancy that had not yet revealed itself as an appreciable difference between the days of the calendar and the seasons. In fact, if the Julian calendar had continued in use to the present, its error would be only about two weeks, still not large enough to be perceived in the Northern Hemisphere as a shift in the calendar dates of spring toward summer. Yet the error was significant in regard to Easter, since its observance depends on a fixed date of the vernal equinox.

The steady regression of the equinoctial date was almost entirely a consequence of the initial discrepancy between the length of the Julian year and the length of the tropical year. On the advice of the Greco-Egyptian scholar Sosigenes, Julius Caesar in instituting the Julian calendar in 45 B.C. established a plan whereby every common year consisted of 365 days and every year perfectly divisible by four consisted of 366. The extra day was originally added just before February 25, which was known as *ante diem sexto Kalendas Martius,* the sixth day before the calends of March. The extra day was therefore called *bissexto* [second six] *Kalendas Martius,* from which the term bissextile derives, a word still used to refer to the leap year and the leap day. In the Gregorian reform the leap day was shifted to the last day of February.

At the time of the inception of the Julian calendar the length of the tropical year was not a well-established quantity, at least not among Western astronomers. Sosigenes' plan for introducing an extra day every four years provided a mean calendar year of 365.25 days. In 45 B.C., however, the length of the tropical year was approximately 365.24232 days, or nearly 11 minutes four seconds shorter than the Julian year. This initially small discrepancy accumulated until the difference was no longer a matter of minutes but of days. It was then that the error began to reveal itself as a gradual regression of the dates of the equinoxes and solstices.

The accumulation of error on the Julian calendar was accelerated by the gradual decrease in the length of the tropical year, a phenomenon consistently ignored by scholars—even astronomers—treating the technical aspects of the calendar. The decrease follows a geometric progression that must be taken into account in computing the accuracy of a solar calendar. If the initial deviation between the Julian calendar and the mean sun had remained at 11 minutes four seconds (if the calendar had invariably gained that amount of time on the sun every year), the error would have amounted to one mean solar day in a little more than 130 years. The drift would have shown up as a shift of the equinoctial date back one calendar day from its position 130 years earlier. Actually, however, the regression of the dates of the equinoxes and solstices had been faster. By the time of the Gregorian correction the calendar was erring



PERITIS MATHEMATICIS.

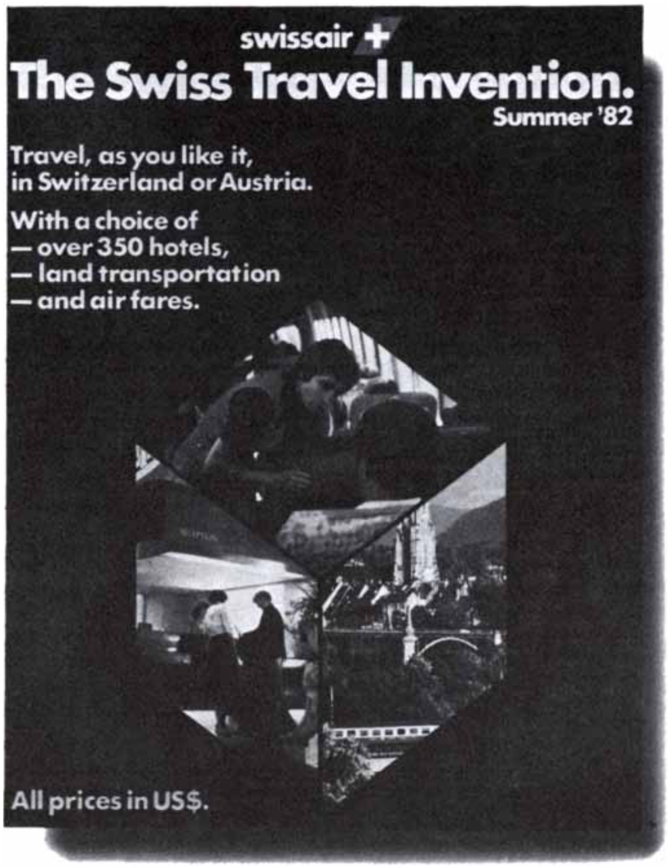
CVM in ſacro cōcilio Tridentino Breuiarij Miſſaliſque emendatio Romano Pontifici reſeruata eſſet, idque felicis recordationis Pius V. quanta maxima potuit diligentia ſuperioribus annis perficiendum curaſſet, atq; edidiſſet: nō tamen id opus uiſum eſt ſuis omnibus numeris abſolutum atque perfectum, niſi reſtitutio quoque anni & eccleſiaſtici Kalendarij accederet. In eam igitur curā dum Gregorius XIII Pont. Max. toto animo & cogitatione incumbit, allatus eſt illi liber ab Aloiſio Lilio cōſcriptus, qui neq; incommodam neque difficilem uiam ac rationem eius rei perficiendæ proponere uidebatur. Verum cū ea kalendarij emendatio multas ac magnas difficultates afferat, & iam diu a bonis uiris omnibus efflagitata, a doctiſsimis mathematicis ſæpe deliberata, & multū agitata, abſolui tamen adhuc, & ad exitum perduci minime potuerit, uiſum eſt prudentiſsimo Pontifici de ea re peritiſsimos quoſque huius ſcientiæ uiros conſulendos eſſe, ut res quę omnium communis eſt, communi etiam omnium conſilio perficiatur. Cogitarat itaque eum librum cunctis Chriſtianis Principibus mittere, ut ipſi adhibitis peritioribus mathematicis, illum aut ſua ſententia comprobarent, aut ſi quid deeſſe uideretur, id omne

**ORIGINATOR of the Gregorian calendar was Aloisius Lilius (the Latinized name of Luigi Giglio), who is portrayed at the left. On his death in 1576 his plan for the reform of the calendar was rescued from oblivion by his brother Antonio and presented to Pope Gregory XIII for review. The plan was adopted by the pope and his advisory commission with only a slight modification by Christoph Clavius. At the right is the first page of Lilius' compendium setting forth his plan for the calendar. The work was found by the author in the National Central Library in Florence, catalogued as being by an anonymous author, although the page reads, "This book written by Aloisio Lilio." The illuminated *C* depicts the pope presiding at a meeting of some of the members of his calendar-reform commission.**

**You don't have to settle for a vacation package. Create your own with The Swiss Travel Invention.** It gives you the best of both worlds. Pre-paid travel so you know what your costs will be, yet freedom to roam as your whims dictate.

You travel independently through Switzerland or Austria with pre-paid accommodation vouchers at guaranteed rates. They're welcome at over 350 hotels with as little notice as a call ahead the night before arrival.

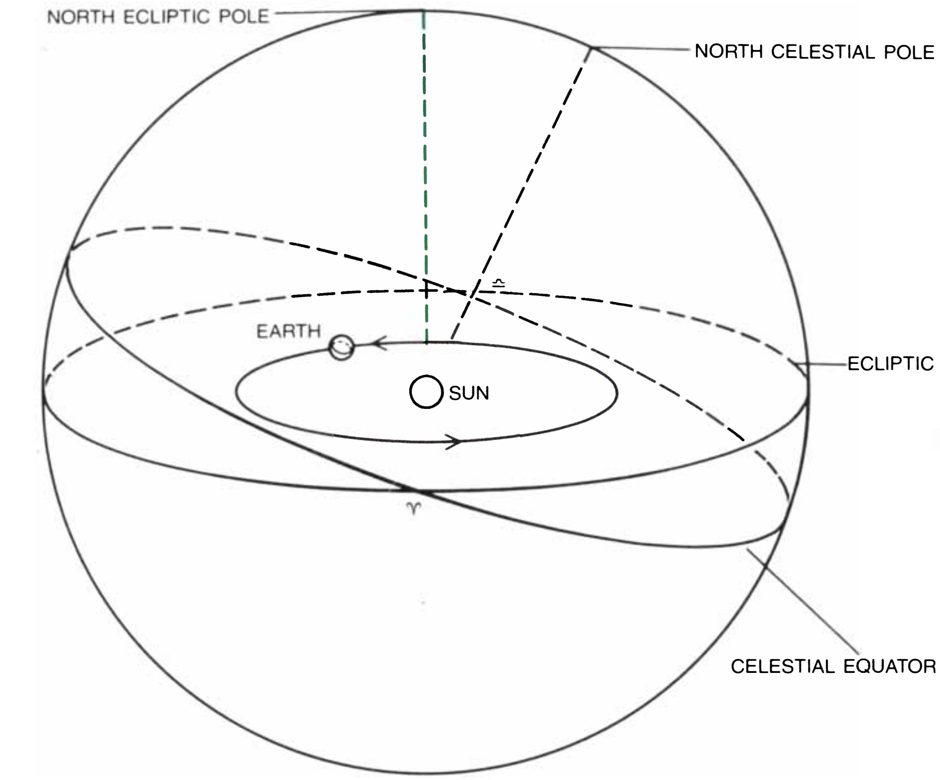
So you can make your plans as you go, according to your budget.

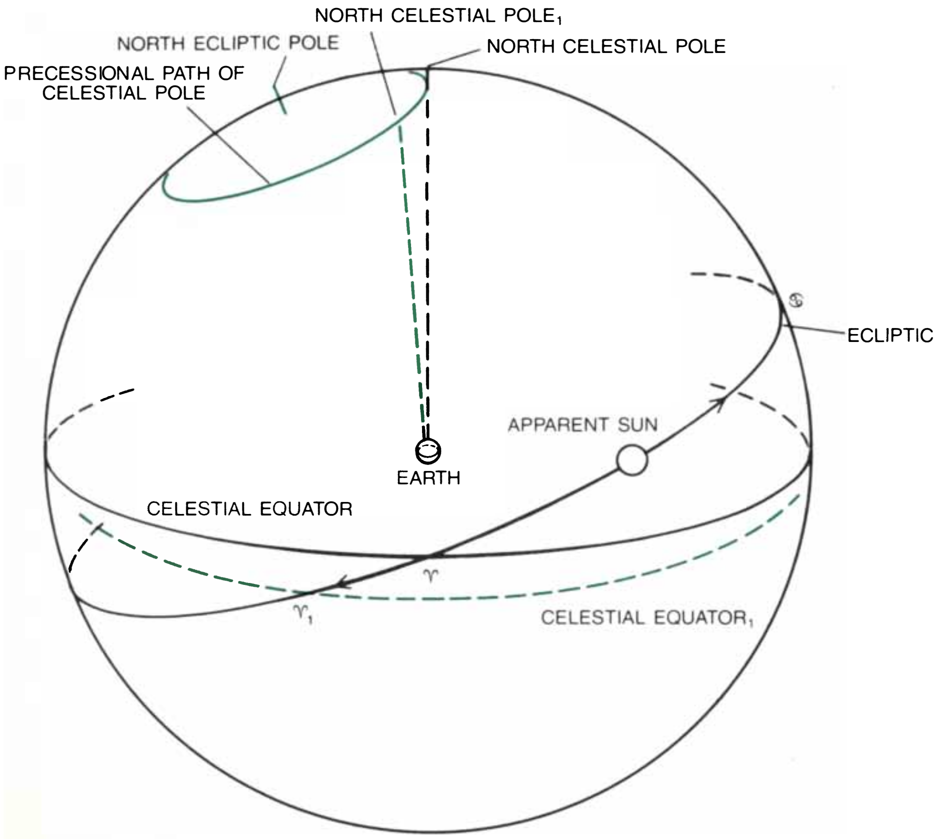



Hotels range from a comfortable $14 per night per person to a luxurious $53.

Airfares range from Apex to First Class, and land travel is by car or rail. Again, your choice. Your pace.

Call your travel agent or **swissair**




**TROPICAL YEAR is the average interval between two consecutive transits of the apparent sun through the vernal equinox (*top*). The vernal equinox (♈) is the intersection of the ecliptic, the plane of the earth's orbit and the celestial equator (the projection of the terrestrial equator onto the celestial sphere). To an observer on the earth the sun seems to revolve about the earth (*bottom*). The sun at the vernal equinox moves along the ecliptic toward the summer solstice (♋) while the vernal equinox precesses in the opposite direction. After one year the sun has not made a complete circuit before it is again back at the vernal equinox in its shifted position (♈$_1$). Because of tidal friction the earth is gradually slowing in its rotation, resulting in a corresponding increase in the precession of the equinoxes. The consequence is that the length of the tropical year is diminishing gradually. The mean year of the calendar is 365.2425 days.**

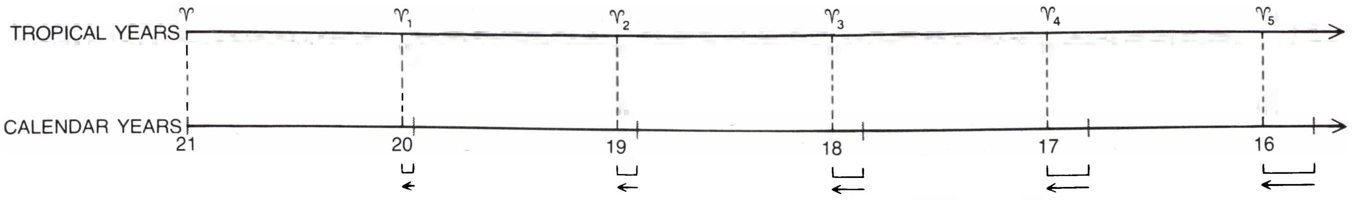
with respect to the mean sun at a rate of about one day every 128.5 years.

The decrease in the length of the tropical year becomes progressively more significant as a function of time. It is one of the circumstances precluding the construction of a perfect calendar. The main reason, though, as Clavius pointed out, is that a calendar designed for ordinary use must provide an integral number of days, whereas the tropical year has a complex fractional component. The matter boils down to trivial arithmetic: a fractional value can never be expressed as an integer. A solar calendar can only approximate the length of the tropical year; the fractional part is accounted for by intercalation, that is, the periodic insertion of intercalary, or leap, days in the calendar. The closer an intercalation approximates the fractional part of the tropical year, the more successful the calendar will be at retaining the vernal equinox at the same date over an extended period.

The plan of reform adopted by Pope Gregory meshes quite neatly with the Julian system and curbs the drift of the date of the equinox so effectively that the calendar will not lose a single day on the sun in well over 2,000 years. The plan that forms the basis of the Gregorian civil calendar was devised not by Clavius or any other member of the commission but by a lecturer in medical science at the University of Perugia, who unfortunately never lived to see his plan become law. His name was Luigi Lilio (often written Giglio), Latinized as Aloisius Lilius. Clavius called him the *primus auctor* of the new calendar. Before the calendar came to be generally known as the Gregorian it was often referred to as the Lilian calendar. Lilius not only developed the new intercalation for the commission but also constructed the table of epacts, which Clavius later modified to better conform to the new leap-year rules.

It was Lilius who recommended the removal of 10 days from the calendar, either all at once or over a period of 40 years beginning in 1584, by dispensing with the leap days that would normally be added every fourth year in that interval. Lilius left it to the commission to choose between these alternatives. Apparently it was Clavius who decided that the days should be extracted all at once from October.

Lilius labored on the details of his reform for some 10 years. The development of a convenient and fairly accurate method of reckoning Easter consumed most of his efforts. Far less of a problem was the construction of a system of intercalation that would better approximate the length of the tropical year. On his death in 1576 his *Compendium novae rationis restituendi kalendarium* (*Compendium of the New Plan for the Restitu-*




**REGRESSION of the date of the vernal equinox results when the mean length of the calendar year is longer than the tropical year. Several vernal equinoxes are depicted. The first one falls on March 21, but over a period of time the discrepancy between the calendar year and the tropical year results in the regression shown by the brackets. For simplicity the gradual shortening of the tropical year is ignored.**

*tion of the Calendar*), historically the most important document on the subject of the calendar, was presented in manuscript by his brother Antonio Lilio to Pope Gregory. It was then circulated for review among an international group of distinguished clerics and scholars. Lilius' work was praised for its accuracy and simplicity, and among the scores of proposals that had been offered it was clearly the one most favored for adoption.

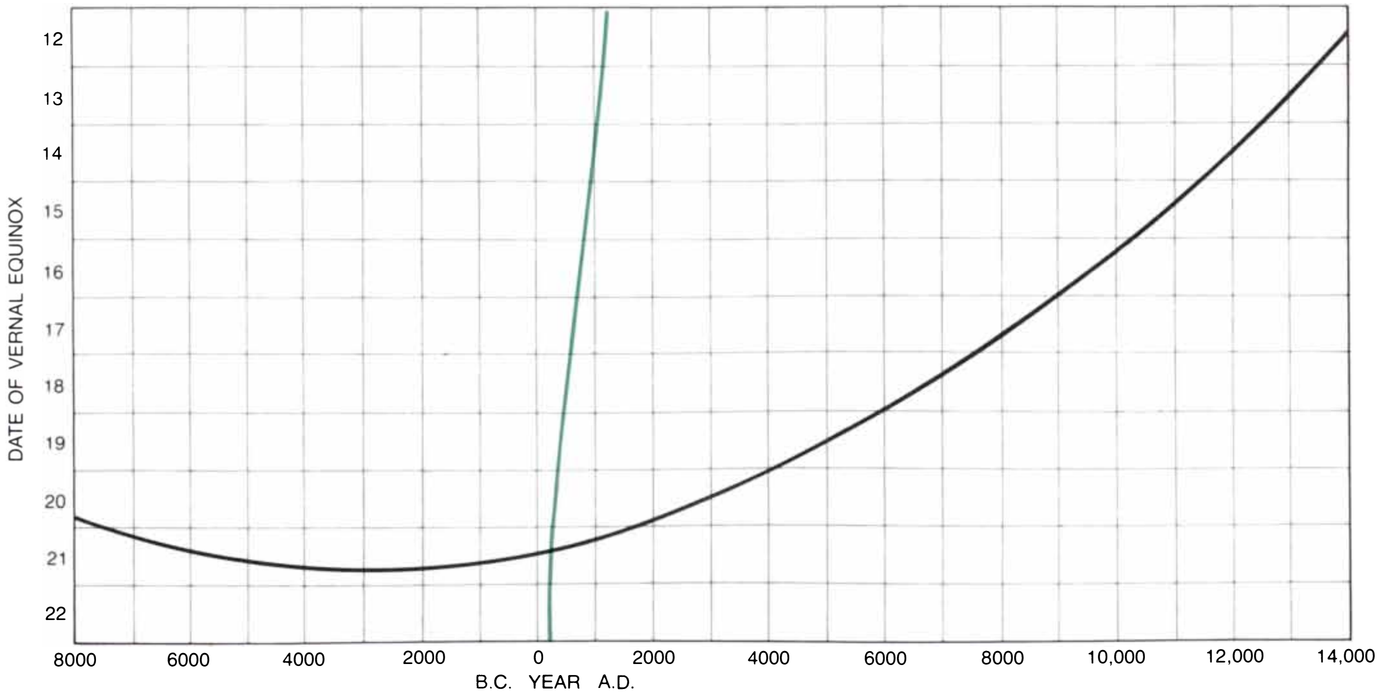
(In a note published in 1974 in *Journal for the History of Astronomy* Noel M. Swerdlow of the University of Chicago said it appeared that Lilius' treatise was lost. He added, "It is, however, not impossible that the *Compendium* survives still undiscovered in manuscript." Swerdlow recently wrote to me to say that Thomas Settle of the Polytechnic Institute of New York had learned that a printed version might be in the archives of the National Central Library in Florence. Indeed, a copy does survive in that library, where it is catalogued as being by an anonymous author. Yet on the first page of that slender volume Lilius explicitly states he is the author. Having confirmed the existence of the *Compendium* in Florence, I searched for it in a number of other Italian cities. The rare treatise, whose whereabouts is curiously never cited by scholars, is also in the archives of the Biblioteca Comunale degli Intronati in Siena and in the Vatican Library. Each library's copy is bound in a varying collection of separately published calendrical works by contemporaries of Lilius, the most notable among them being Alessandro Piccolomini.)

The intercalation Lilius proposed in the *Compendium* was simple: the suppression of three leap days in centurial years not perfectly divisible by 400. The Gregorian intercalation therefore follows the Julian system with the single exception that centurial years such as 1700, 1800 and 1900, which would have been leap years in the Julian calendar, became common years (with the leap day dropped). The Gregorian calendar reduces the number of intercalary days to 97 in 400 years, as opposed to 100 intercalary days in 400 Julian years.

The change is small but profound. It brings the mean length of the calendar year into much closer agreement with the tropical year, providing a mean calendar year of 365.2425 days. In 1582 the length of the tropical year was nearly 365.24222 days, which differed from the Gregorian year by only a little more than 24 seconds. If the length of the year itself were not gradually diminishing, the calendar under the plan devised by Lilius would retain the date of the vernal equinox at or about March 21 for more than 3,550 years.

Clavius himself estimated that the calendar would deviate by one day in the year 5084. Considering the decline in the length of the year, which was quite unknown to 16th-century astronomers, the calendar will remain accurate to within one mean solar day for a much shorter time, some 2,417 years, or until about the year 4317. (Extrapolations are based on the determination of the tropical year at epoch 1900 by the American astronomer Simon Newcomb; his figure was 365.24219879 mean solar days.)

How Lilius arrived at a value of 365.2425 days remains a mystery. Discouragingly the *Compendium* sheds no new light on the question. The most reli-




**JULIAN AND GREGORIAN CALENDARS are compared with respect to the drift of the date of the vernal equinox. In the Julian calendar the drift (*color*) was quite rapid. In the Gregorian calendar it is slower (*black*) because the calendar more closely approximates the tropical year. The tropical year equaled the Gregorian year of 365.2425 days in about 3000 B.C. Since then the drift has accelerated.**

able values of the tropical year known to Lilius were given in the *Alphonsine Tables* of 1252, Copernicus' *De revolutionibus orbium coelestium* of 1543 and the *Prutenic Tables* of 1551. Each source gives a value of about 365 days five hours 49 minutes 16 seconds; they differ from one another by less than a second, and yet all are approximately four seconds longer than the Gregorian year. None will produce an intercalation of 97 days in 400 years. Swerdlow recognized, however, that all three quantities expressed as sexagesimal fractions—the form in which they would be taken from a table of the mean motion of the sun—are identical with the second fractional place. The form of expressing the sexagesimal fraction is 365; 14, 33 (the second fractional place being 33). Converting the sexagesimal fraction into a common one gives $365^{97}/_{400}$, or 365.2425 days. It is as good a theory as any for the origin of the length of the Gregorian year.

Lilius was not the first to propose an intercalation of 97 days in 400 years. The same plan was put forward in 1560 by Petrus Pitatus, an astronomer of Verona. It is not known whether Lilius borrowed from Pitatus, but the intercalation is simple enough to have been derived independently.

Lilius' system confines the date of the vernal equinox within rather narrow limits. The date can be March 21, March 20 and even March 19, although it has not fallen on March 19 since about the turn of the century. Notwithstanding what is said in many textbooks on astronomy, the vernal equinox as a result of the Gregorian rules for leap year falls more often on March 20 than it does on March 21.

A consequence of the Gregorian intercalation is that all the dates in the calendar repeat in a cycle of precisely 146,097 days, which equals 400 Gregorian years. In other words, every date in 1583 will be repeated in 1983; the same is true for 1584 and 1984 and so on in multiples of 400 years. The calendar completes its first grand cycle on October 15 of this year.

Lilius and Clavius succeeded where others had failed. The Gregorian calendar affords a highly satisfactory compromise between essential accuracy and much-desired simplicity. For more than 800 years attempts to improve the Julian calendar were made by such able men as Roger Bacon, Nicolas of Cusa, Regiomontanus, Johannes Schöner and Paul of Middelburg. Each scholar had recorded the growing disparity between the calendar and the sun, but for one reason or another—including political strife, governmental indifference, untimely death—nothing was done about it until in 1572 a former professor of law from Bologna named Ugo Buoncompagni became Pope Gregory XIII.

# The Regeneration of Potato Plants from Leaf-Cell Protoplasts

*A new approach to the cloning, or asexual propagation, of plants starts with living cells that have been stripped of their outer wall. The method yields variants that promise future crop improvements*

by James F. Shepard

In plants as in animals variation from one individual to another is generally brought about through the shuffling of genes in sexual reproduction. The adaptive value of such variation is suggested by the elaborate reproductive organs of the flowering plants and by the equally elaborate strategies that have evolved to facilitate cross-pollination. Even so, not all flowering plants are totally dependent on sexual methods of reproduction. Some species have evolved an additional asexual reproductive capability, embodied in tissues that are anatomically distinct from the flower parts. The asexual, or vegetative, reproductive organ of such a plant can be an underground stem, a modified root or even a leaf that has the ability to develop into a complete plant. Whatever the origin of the vegetative tissue, asexual reproduction tends to preserve the phenotype, or physical characteristics, of the parent in its offspring.

Individual organisms that arise asexually from the somatic, or body, cells of the parent rather than from the specialized sexual cells are called clones, and the propagation of species by such methods is called cloning. I shall describe here a new experimental approach to cloning; perhaps surprisingly, it yields potentially useful forms of phenotypic variation in the regenerated plants. The technique, called protoplast cloning, has been developed primarily for potato plants, but in principle it could be applied to a wide range of crops. Further study is needed to determine what role, if any, protoplast cloning may play in future crop-improvement programs. Because the topic of cloning has drawn considerable attention lately, I shall first explain the meaning of the term in this context.

A clone is often presumed to be a carbon-copy replica of the parent: "a chip off the old block." At first such a conclusion seems reasonable. In nature clones derived from the same parent are in many cases very similar to one another, and so they might be expected to have functionally equivalent genomes. This commonly held idea, however, is a misconception.

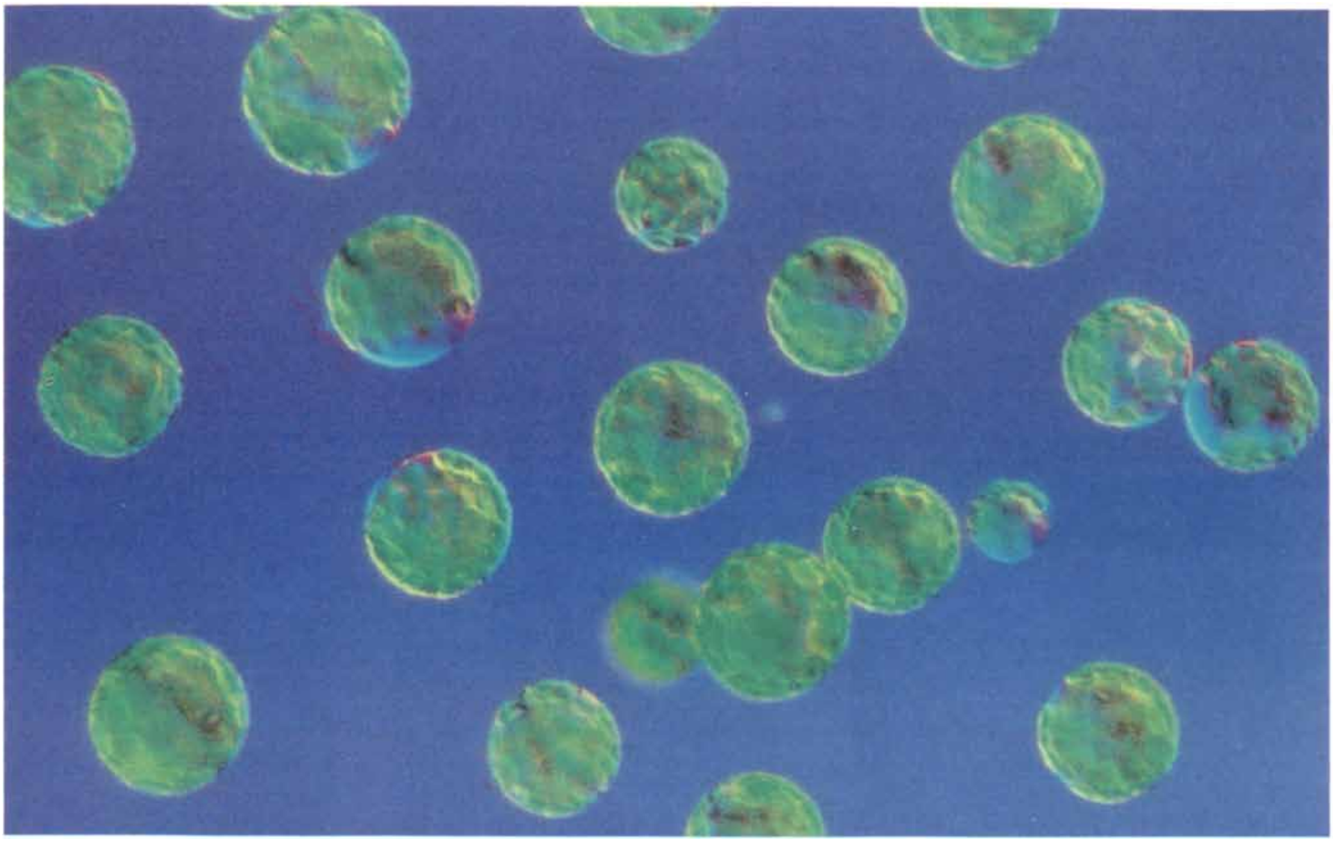
The term clone is derived from the Greek κλών, meaning a slip, or twig, suitable for plant propagation. In 1903 Herbert J. Webber of the U.S. Department of Agriculture proposed that the transliterated word "clon" be adopted to designate those plants that are propagated asexually and (in his words) "are simply parts of the same individual." Soon afterward "clon" was altered in spelling to clone, and since then usage of the term has expanded to the point where it is now applied to all asexually reproduced forms of life. Indeed, the word has been applied even to the reproduction of DNA, the genetic material; thus one now speaks somewhat loosely of the cloning of genes in bacteria.

Today most scientific reference works continue to define a clone in the more restricted sense as an individual organism derived asexually from a single cell through mitosis, the form of cell division in which the daughter cells retain the same number of chromosomes as the parent cell has. (Mitosis is the usual form of cell division among somatic cells; it is distinguished from meiosis, or sexual-cell division, in which each daughter cell receives half as many chromosomes as the parent cell.) This definition is not meant to imply either phenotypic or genetic homogeneity in the resulting population; in other words, clones from a single parent need not be identical in appearance or have exactly the same genetic composition. Indeed, in some species a clone can differ noticeably from its parent; it follows that the genome of a clone, like that of an organism reproduced sexually, must be equipped in some way to generate variation. This second source of variation has significantly extended the range of options available to the plant breeder in his quest for improved plant varieties.

Several crop species are propagated asexually to preserve essential varietal characteristics, either because the plants are sexually infertile or because their genomes are too complex for their phenotype to remain uniform in sexually reproduced progeny. In these species clones that differ in some obvious way from the parent sometimes appear. Such divergent individuals are called somatic variants, bud sports or simply sports; they result from permanent genetic changes in specialized meristem cells: the rapidly dividing cells at the tip of a growing stem, branch or root that generate all or part of the new plant. Many important varieties of clonally propagated crop plants have arisen from such vegetative mutations. Familiar examples are the pink grapefruit, the navel orange, the nectarine and several varieties of potato. In other plants, for instance the sweet potato, sports appear with a frequency that can be as high as 2 percent; as a result the maintenance of varietal purity through conventional cloning is a continuing problem.

A number of genetic mechanisms can lead to the appearance of a plant sport. Changes in the number of chromosomes per cell nucleus, "point" mutations (which alter only a single gene on one of the chromosomes) and modifications of extranuclear genes (that is, genes in cellular organelles such as chloroplasts or mitochondria) have all been implicated. Most somatic variants have not yet been characterized genetically, and it is probable that still more mechanisms will be recognized as potential contributors to the phenomenon.

An impediment to a better understanding of somatic variation is the fact that many species of crop plants have polyploid genomes: each cell nucleus contains more than two sets of chromosomes. In an asexually propagated species polyploidy does not preclude a high frequency of somatic change, but it does make the genetic analysis of any alter-



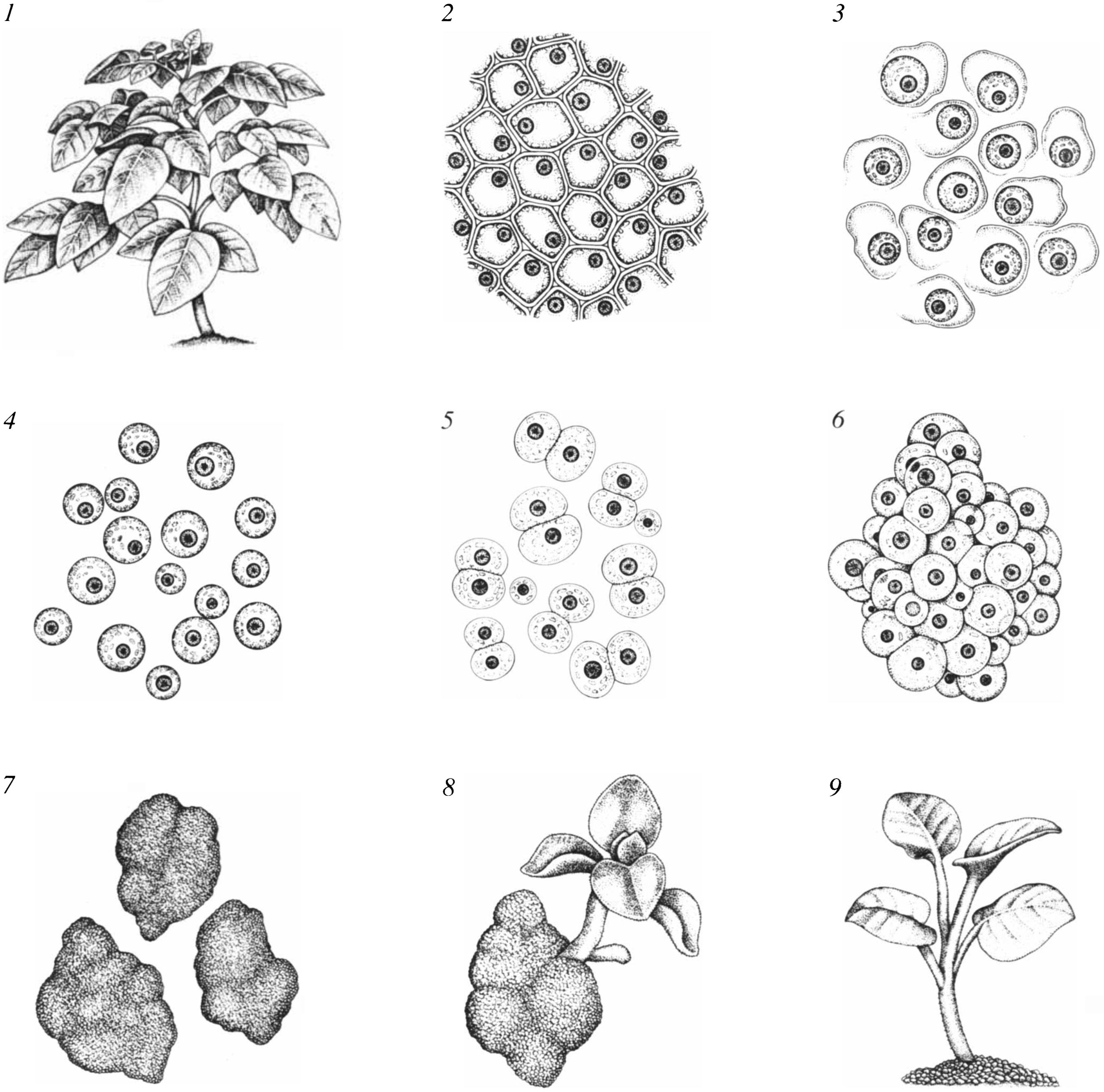
**PROTOPLASTS** prepared from the leaf cells of a potato plant contract into spheres when they are separated from their supporting wall. The protoplasts thrive in a culture medium that supplies all the nutrients they need to synthesize new cell walls, grow and divide. Each protoplast contains a nucleus and a number of chloroplasts (cellular organelles that make chlorophyll and hence appear as green spots).

**FIELD OF PROTOPLAST-DERIVED CLONES** (called protoclones) of Russet Burbank potato plants was photographed at the Potato Research Farm near Grand Forks, N.D. Each row consists of plants propagated asexually by planting the tubers of a single plant regenerated from an isolated protoplast. Variations are evident both in the height of the plants and in the number of flowers per plant.

ation very difficult, particularly when the genome is complex and when many genes control the expression of the modified trait.

In recent years plant cloning has been refined to the point where a single cell removed from the body of a plant can be cultured and then induced to regenerate a complete individual. This developmental potential of single cells, a property known to cell biologists as totipotency, was first suggested by the results of experiments with cultured carrot cells done by Frederick C. Steward of Cornell University about 25 years ago. By about 1965 totipotency was firmly established for comparable tissue cultures of tobacco and other plant species; thereafter it was demonstrated for somatic cells and sexual cells isolated directly from numerous plants.

The final conceptual advance in this area was reported in 1971 by Itaru Takebe and his colleagues in Japan. They employed a combination of two enzymes, pectinase and cellulase, to dissociate tobacco leaves into living but wall-less plant cells called protoplasts. The isolated protoplasts were cultured in a medium that promoted growth and cell division. In the last step of their procedure the masses of cultured cells, called calluses, were induced to regen-




**CLONING PROCEDURE** employed by the author and his colleagues at Kansas State University to regenerate a complete potato plant from a leaf-cell protoplast is illustrated in this sequence of drawings. Small terminal leaves are first removed from a young potato plant (*1*). The leaves are placed in a solution containing a combination of enzymes capable of dissolving the cell wall (*2*). Another substance in the solution causes the protoplasts to withdraw from the cell wall and to become spherical, thereby protecting the living protoplasm during the disintegration of the wall (*3*). The isolated protoplasts are next transferred to a culture medium (*4*), where they grow, synthesize new cell walls and begin to divide (*5*). After about two weeks of culture each protoplast has given rise to a clump of undifferentiated cells, called a microcallus (*6*). The microcalluses are transferred to a second culture medium, where they develop into full-size calluses (*7*). At this stage the cells of the callus begin to differentiate, forming a primordial shoot (*8*). The shoot develops into a small plant with roots in a third culture medium and is then planted in soil (*9*). Micrographs of several key steps appear on the next two pages.

erate small shoots, which eventually grew into entire plants.

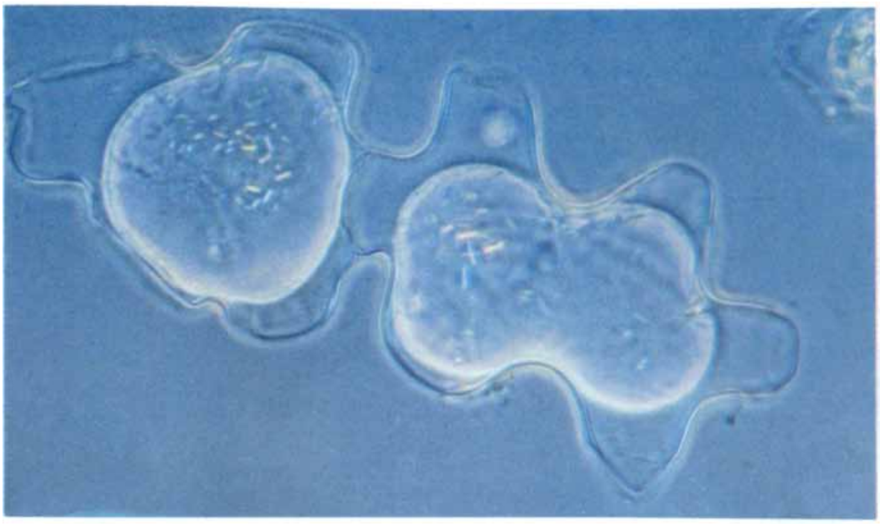
A key finding in these initial tobacco-protoplast experiments was that more than 90 percent of the protoplast-derived clones, or protoclones, were remarkably similar to the parent both in appearance and in chromosome complement. This similarity had generally not been observed in earlier experiments, in which tobacco plants had been regenerated from the protoplasts of callus cells maintained in long-term tissue cultures. In the earlier work a high proportion of the clones had been malformed, dwarfed and marked by severe chromosomal abnormalities. Thus leaf-cell protoplasts of tobacco appeared to offer a source of genetically stable cells with which one could do various experiments without having to contend with a background of preexisting or spontaneous genetic changes.

This advance suggested that if suitable procedures could be devised for the regeneration of protoplast-derived plants from other species, notably the food crops, their protoclones should also be genetically uniform (that is, unless the protoplasts were intentionally subjected to mutagenic treatment or genetic manipulation before they were cultured). Follow-up experiments confirmed that clones from leaf-cell protoplasts of a few other plant species were also fairly uniform in appearance and did not express a significant frequency of somatic change.

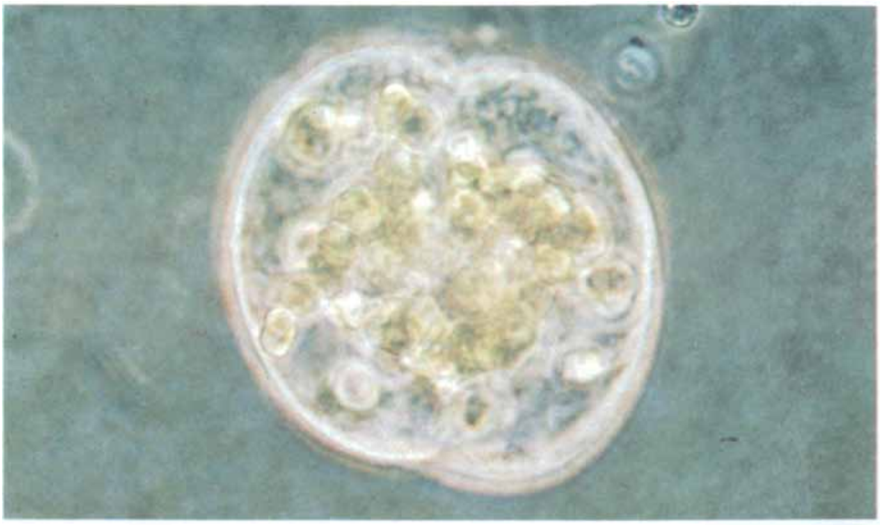
In 1973 my colleagues and I began working with leaf-cell protoplasts of tobacco. We soon found phenotypic differences between some members of the protoplast-derived clonal populations. In the absence of any mutagenic treatment variegated color patterns appeared in the leaves of roughly one out of every 250 regenerated plants. In subsequent sexual cross-breeding experiments the altered characteristics were usually passed to the offspring only through the maternal cells of the plant. Such a pattern of inheritance suggests that the mutations were in some genetic element outside the cell nucleus.

These findings implied that the tobacco protoclones with a normal chromosome number were not all identical. Accordingly we were encouraged to believe that if similar techniques could be applied to a crop species that naturally undergoes frequent somatic mutation, potentially useful forms of genetic change might ensue. To test this hypothesis we chose a common commercial variety of potato named the Russet Burbank.

The origin of the Russet Burbank potato dates back to 1875, when the botanist Luther Burbank selected a seedling from the progeny of a single potato fruit, or berry. The seedling was asexually propagated through the planting of



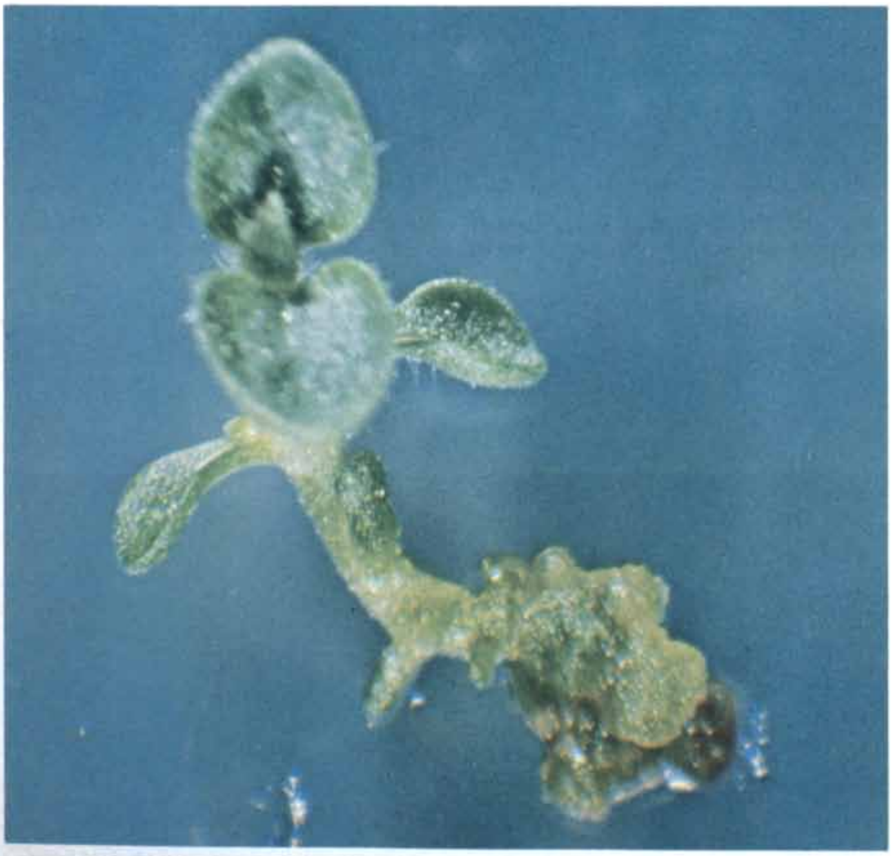
**PROTOPLASTS WITHDRAW** from cell walls in this micrograph of a pair of leaf cells. The protective contraction of the protoplasm is triggered by an osmoticum: a substance that induces the cell to lose water through osmosis. The cells are shown enlarged about 1,000 diameters.



**GROWING PROTOPLAST** has synthesized new cell walls and is in the process of dividing for the first time in this phase-contrast micrograph, which was made after six days of culture.

**MICROCALLUS** of undifferentiated cells grown from a single leaf-cell protoplast of potato was stained with a fluorescent dye called acridine orange and illuminated with ultraviolet radiation to make this micrograph. The nucleus of each cell appears as a bright yellow spot.

**ARRAY OF CALLUSES** derived from leaf-cell protoplasts of potato are seen growing in a Petri dish about four weeks after they were transferred to the second, shoot-inducing culture medium. The calluses usually begin to turn green at this stage of the regeneration process.



**PRIMORDIAL SHOOT** emerges from the mass of undifferentiated callus cells after six weeks of incubation in the shoot-inducing medium. The magnification is about 20 diameters.

tubers (the familiar edible portion of the potato plant, which is not a root but a modified underground stem); repeated cloning of this plant type eventually gave rise to the variety designated Burbank. Soon after the turn of the century a sport was selected from the Burbank line that produced tubers with a russet, or reddish brown, skin. The variant was given the name Russet Burbank; today it is the most widely grown variety of potato in the U.S., accounting for almost 40 percent of the total production.

Burbank's achievement in providing the basis for a new potato cultivar from just a few seeds of one plant can be better appreciated if one considers that in the past 50 years more than 20 million potato seedlings have been evaluated by plant breeders in the U.S., but none has led to a commercial variety as successful as the Russet Burbank sport. It seems reasonable to predict, therefore, that given such a sound genetic foundation on which to build, definite opportunities exist for developing improved versions of the cultivar through asexual means. The mere selection of naturally occurring sports is neither an efficient approach nor a realistic one, however, because the frequency of somatic change for individual traits is low, and few if any spontaneous improvements in disease resistance, for example, have been recognized.

As an alternative we set out to measure the frequency of somatic variation in clonal populations derived from single protoplasts; the task would then become one of determining whether any advantage could be gained over the method of selecting sports at the plant level. To test this approach we first had to develop techniques for isolating the protoplasts of potato cells and for regenerating complete plants from them. In 1977 we published a preliminary description of a successful procedure, and since then we have refined the techniques and extended them to other potato cultivars.

The regeneration of a potato plant from a leaf-cell protoplast begins with the enzymatic digestion of the leaf tissues to release a large number of individual protoplasts. Up to this point the procedure resembles the methods already employed for isolating leaf-cell protoplasts from other plants. From here on, however, a series of specific cultural steps is required to initiate and sustain the division of the potato protoplasts and to direct the development of the calluses that arise from the proliferation of each protoplast's progeny.

When the protoplast-derived calluses are subjected to the proper sequence of cultural conditions, they exhibit a compact growth habit and become intensely green in color. The cell masses, which at this stage are typically between three and five millimeters across, are then

transferred to another culture medium designed to stimulate the development of shoots. Between 10 and 40 percent of the calluses form primordial shoots while growing on this medium. The calluses with shoots are transferred to a third culture medium, where the shoots develop into complete plants with stems, leaves and roots.

Unlike the tobacco plants cloned from protoplasts in the earlier studies, most of the potato plants regenerated from protoplasts of the Russet Burbank cultivar are not identical with the parent or with one another. There are many instances of variation among the protoplast-derived potato clones; to simplify the discussion, however, one can regard the variant protoclones as falling into either one of two general classes. In one class are the wild aberrants: those protoclones that exhibit gross structural anomalies and are generally lacking in vigor. Close examination of the cell nuclei of such protoclones suggests that in many of them the chromosome number has been altered. Such oddities appear in all populations of Russet Burbank protoclones; they have not been analyzed in depth, however, and they will not be considered further in this discussion.

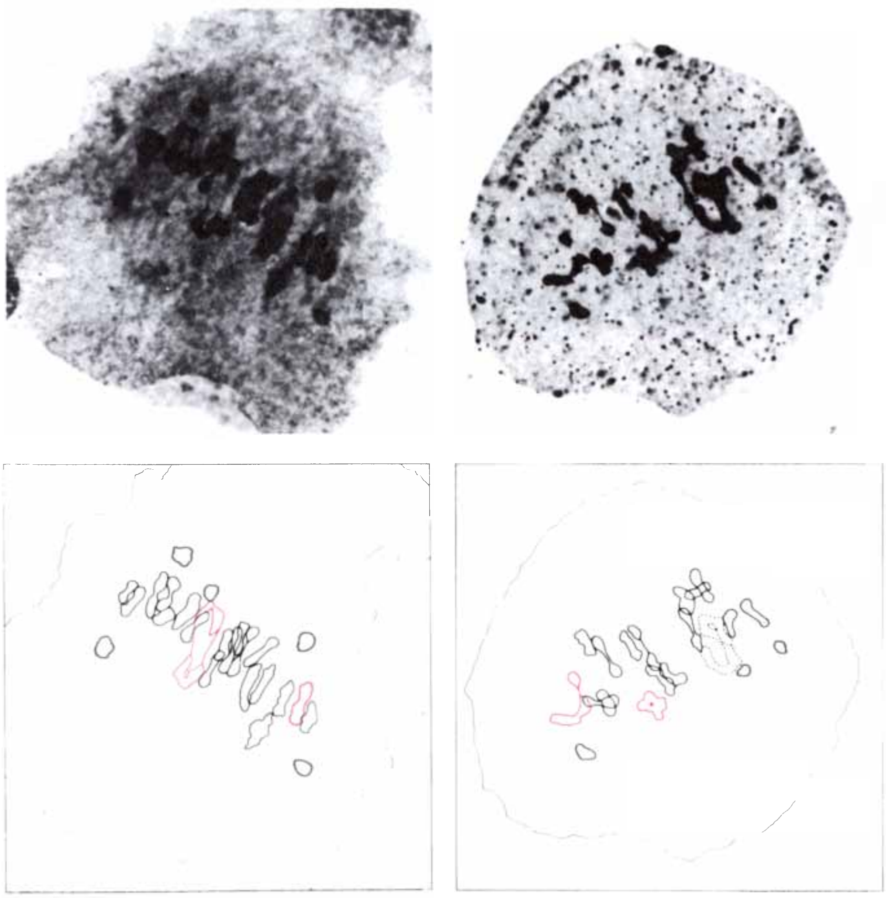
The variant Russet Burbank protoclones in the other class are called phenotypic variants, meaning that the differences between individuals are subtle and that most of the definitive characteristics of the Russet Burbank cultivar have been retained. Studies of selected phenotypic variants suggest that most of them have retained the original chromosome number (48), although it is still possible some of them may differ slightly in either chromosome number or composition, or in both.

One kind of change that has appeared in the phenotypic-variant class of protoclones is an altered susceptibility to disease-causing organisms. In 1978 Ulrich Matern and Gary A. Strobel of Montana State University and I found that a few protoclones of Russet Burbank were less susceptible than the parent variety to damage by the fungus *Alternaria solani,* the pathogen responsible for the disease of potato plants known as early blight. Although most of the remaining protoclones were similar to the parent in their reaction to the fungus, still others were more severely affected. The differences between protoclones were in susceptibility to the pathogen and were detectable only after inoculation with the fungus. Otherwise the protoclones were fairly uniform in appearance.

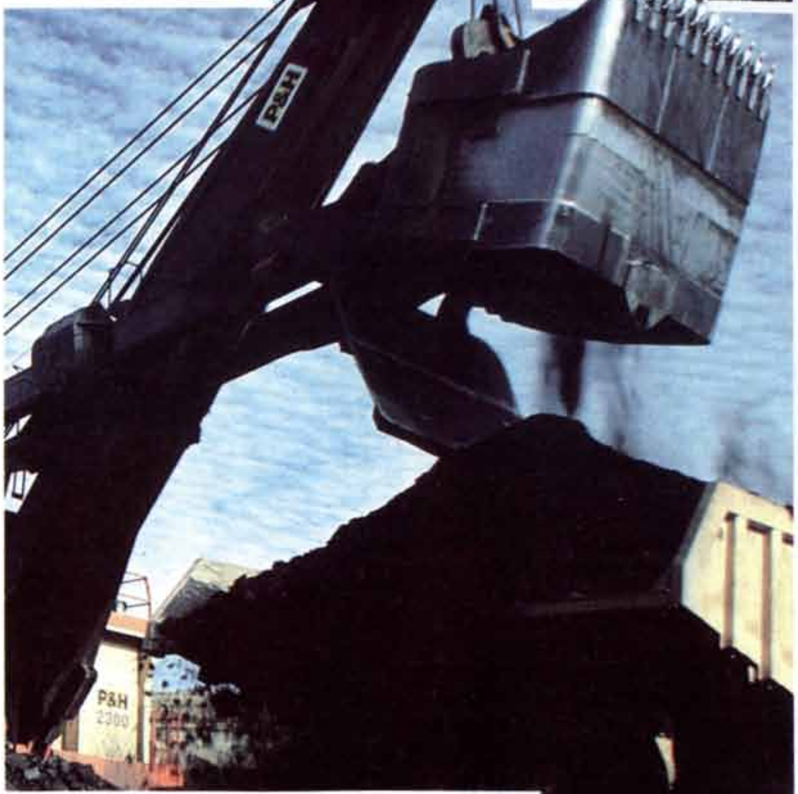
Potato plants are also susceptible to late blight, a disease caused by the fungus *Phytophthora infestans.* This disease, which caused the Irish "potato famine" of the 1840's, is still a major threat to commercial potato farming. Genes for resistance to the fungus have been introduced into some cultivars from wild relatives of the potato through conventional breeding techniques, but the fungus itself is highly variable, and strains emerge that can quickly overcome a simply inherited gene for resistance to the disease. Longer-lasting and broader-spectrum protection has sometimes been achieved by amalgamating several genes into what is termed a multigenic form of resistance.

In 1980 Elias Shahin, Dennis Bidney and I showed that protoclones of Russet Burbank vary in their resistance to the late-blight fungus. Small stem cuttings from each protoclone were inoculated with the pathogen, as were comparable cuttings from the parent. Roughly 2 percent of the protoclones were found to be more resistant to the tested strain of *P. infestans* than the Russet Burbank plants were at the same stage of development. Moreover, when the plants were propagated through their tubers, the progeny of the resistant protoclones retained their original level of resistance to the fungus.

Individual protoclones with enhanced resistance to late blight differed in the degree of resistance to the pathogen. None of them, however, expressed the extreme form of resistance, which is characterized by a hypersensitive local reaction to the inoculation; it is this form of resistance that is conveyed by the major dominant resistance genes in some of the wild relatives of the potato. Hence it is highly unlikely that point mutations of dominant genes are responsible for the resistance observed in the protoclones, unless the genes involved are of a different type from those currently recognized in the potato. Mu-



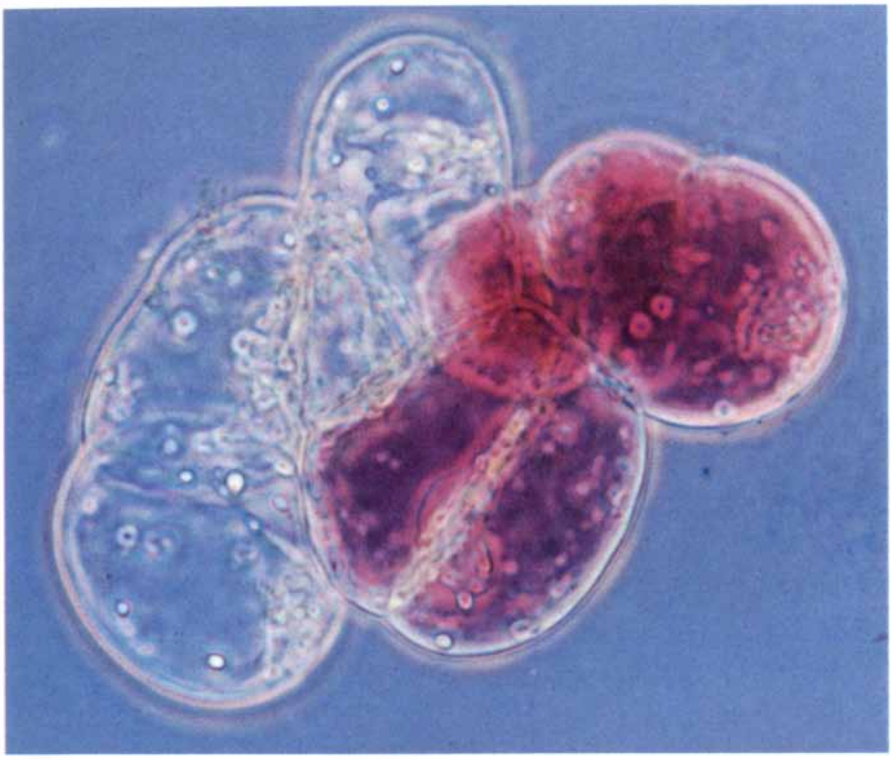
**EVIDENCE OF GENETIC CHANGE can be detected in the chromosomes of two pollen cells, one cell from a Russet Burbank potato plant (*left*) and the other from a protoclone regenerated from a leaf cell of the same plant (*right*). The micrographs show the pairing of chromosomes that takes place during an early phase of meiosis (sexual-cell division). The various dark shapes result from the association of homologous chromosomes (those carrying genes that serve the same function) or homologous segments of chromosomes. Differences in the pairing patterns of the two cells can be attributed to structural modifications or rearrangements in some of the chromosomes of the protoclonal line. In the cell of the parental Russet Burbank plant there are five unpaired chromosomes (called univalents), 18 associations of two chromosomes each (bivalents), one association of three chromosomes (a trivalent) and one association of four chromosomes (a quadrivalent). In the accompanying map the univalents are shown in gray, the bivalents in black, the trivalent in light color and the quadrivalent in dark color. In the protoclonal cell there are two univalents, 15 bivalents, no trivalents, two quadrivalents and one association of eight chromosomes. The eight-chromosome chain, represented by the broken black outline, is suggestive of a translocation of genetic information between nonhomologous chromosomes. Events of this kind are not observed in the parental variety. The photographs and maps were supplied by Bikrum S. Gill and Lauren Kam of Kansas State University.**



P&H







**RED PIGMENT APPEARS** in cells that have been cultured from protoplasts isolated from the leaves of a potato variety named Bison, which produces tubers with a characteristic red skin. Only some of the cells in this micrograph, made after 14 days of culture, have begun to accumulate the pigment. The expression of the gene (or genes) controlling the synthesis of the red pigment can serve as a useful genetic marker in certain kinds of laboratory test.

**TUBER DEVELOPS** at the apex of a primordial shoot growing from a protoplast-derived callus of Russet Burbank cells. Special nutritional and environmental conditions were required to induce the tip of the shoot to swell and form the tuber. The technique may lead to a new way of studying tuber formation under controlled conditions of temperature and period of daylight.

tations of recessive genes for disease resistance would probably not be expressed. The reason is that potato plants are tetraploid: each somatic cell has four sets of chromosomes. For a recessive gene to be expressed it would have to be present in all four sets, a coincidence that is extremely unlikely. It is important to point out, however, that the evaluation of late-blight resistance in these protoclones has been done only in the laboratory and not under field conditions. It is therefore uncertain whether the resistant reactions will be useful in commercial potato production.

Screening protoclones for differential reactions to some pathogens, as described above, can be a fairly rapid technique for testing whether stable phenotypic differences exist within protoclonal populations. A second group of traits requires quantification over the course of several years before data can be considered significant. Protoclones in this category differ from the parent in complex or quantitative horticultural traits, which are defined by the joint action of many genes.

Such traits are exquisitely sensitive to environmental influences and must therefore be measured repeatedly under an adequate range of conditions. For example, the number of leaves on each plant and their total surface area are considerably reduced in most plants when there is not enough moisture for long periods. A number of genes control this response, presumably in an overall adaptation that protects the plant against excessive water loss. This is an extreme example but not an isolated one; a plant responds to environmental stimuli throughout its lifetime by altering a wide range of physiological processes to accommodate fluctuations not only in moisture but also in light intensity, hours of daylight, ambient temperature, atmospheric humidity and available nutrients.

In collaboration with Gary A. Secor of North Dakota State University my colleagues and I have evaluated a population of 65 selected Russet Burbank protoclones grown under field conditions for several years. Quantification of numerous horticultural traits has firmly established that stable phenotypic variation exists in this population for almost all the complex traits measured. Moreover, when all the traits were considered, no protoclone of the 65 tested was exactly like another, and none was an exact copy of the parent plant. It is significant that of all the horticultural traits measured in this experiment even the genetically most complex ones, such as tuber yield, were found to be subject to variation. This observation suggests that simple mechanisms of genetic change, such as point mutations, cannot be the sole cause—or even the primary cause—of the variation.

There is further evidence to support

# hexa·photo·cybernetic
## The Possibilities are Endless.

Shutter-Priority 1000 5.6

Six-mode exposure control. System versatility. Newer electronics for wider applications.



The Canon A-1 is one of the world's most advanced automatic SLR cameras. Combining the finest in optical and mechanical engineering with the most sophisticated electronics, it's technology applied to give you the ultimate in creative control. At the touch of a button.

Depending on your subject, you can choose from six independent exposure modes to achieve the results you want:

1 Shutter-Priority: You select the shutter speed, to freeze the action and prevent camera shake or create an intentional blur. The A-1 automatically selects the appropriate lens opening.

2 Aperture-Priority: Control the area in focus by selecting the lens opening for the effect you want. The A-1 matches with the right speed.

3 Programmed: When you need to shoot fast, just focus. The A-1 will select *both speed and aperture* for great results.

4 Stopped-Down: For extreme close-up or specialized photography, a bellows, a microscope or almost anything can be attached to the A-1. It's still automatic.

5 Flash: Totally automatic flash photography, of course, with a wide variety of Canon Speedlites to choose from.

6 Manual: Yes. For those times when you absolutely want to do it all yourself. To experiment. To explore the possibilities.

There are over forty fine Canon lenses ranging from Fish Eye to Super Telephoto, plus accessories to meet every need. If you can't photograph your subject with a Canon A-1, it probably can't be photographed.

From the sophistication of its LED viewfinder display, to a ruggedness that allows up to five-frame-per-second motor drive, the Canon A-1 represents an incredible technology. At a price that makes owning one a definite possibility.

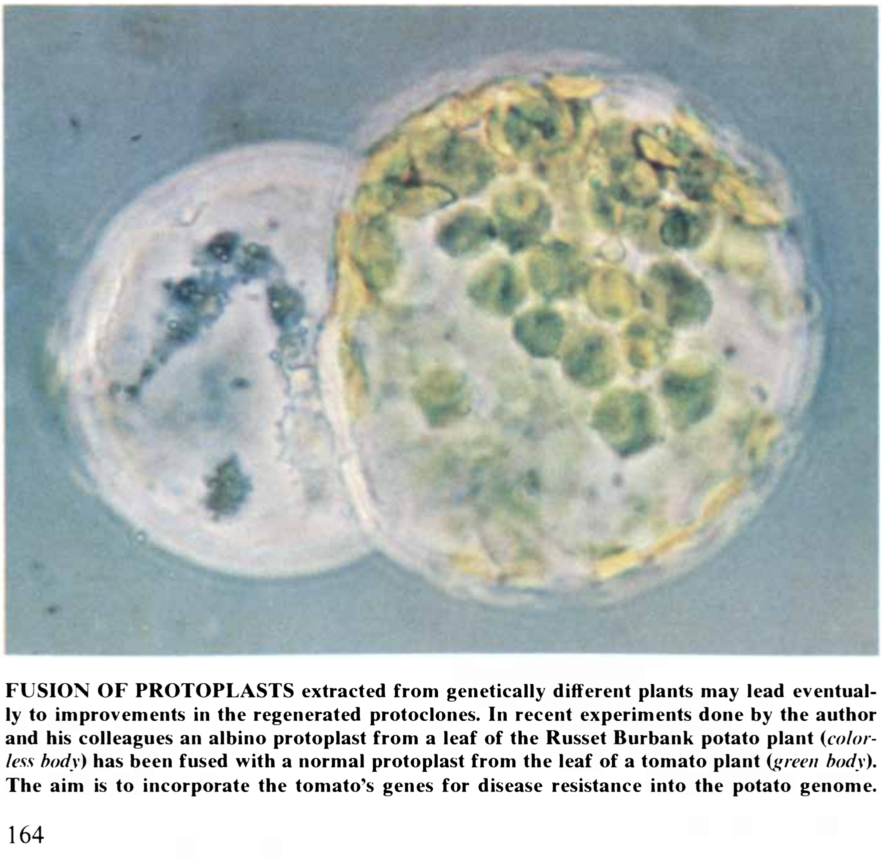
**Canon® A-1**

Canon USA, Inc., One Canon Plaza, Lake Success, New York 11042 • 140 Industrial Drive, Elmhurst, Illinois 60126 • 6380 Peachtree Industrial Blvd., Norcross, Georgia 30071 • 123 Paularino Avenue East, Costa Mesa, California 92626 • Bldg. B-2, 1050 Ala Moana Blvd., Honolulu, Hawaii 96814 • Canon Canada, Inc., Ontario this conclusion in the observed stability of flower color. Russet Burbank has matching recessive genes for white flowers. A dominant point mutation for, say, pink flowers could be indicative of a single-gene mutation. To date flower color has been evaluated for several thousand protoclones, but in no case has there been a departure from the normal white flowers. It follows that if identifiable point mutations for simply inherited traits do take place, they must appear at a substantially lower frequency than other types of change do. This conclusion is of prime conceptual importance because complex genetic traits require the expression of many genes; if point mutations were the only mechanism for the genetic modification of protoclones, there would be little opportunity for significantly improving traits such as field performance and the size and shape of tubers. Through mechanisms of genetic modification other than point mutation numerous plant traits become subject to amelioration through protoplast culture and plant regeneration.

Of all the traits measured in protoclonal populations so far, tuber yield is of particular interest. It is almost always desirable to increase the yield of a crop if commensurate increases are not also necessary in the energy, labor and other inputs of production. In potato farming yield means the weight of tubers produced per hectare of land, with the stipulation that the tubers be of a quality suitable for their intended use. Tuber quality in turn is defined by factors such as uniformity of shape and size, susceptibility to disease, storage characteristics and suitability for processing. It is essential that there be no compromise in such qualitative traits when increases in tuber weight are sought. In other circumstances the elimination of a single flaw in tuber quality can be as significant economically as increasing the weight of the tubers.

Statistical analysis of data from a 1979 field plot in North Dakota revealed that none of the 65 protoclones evaluated was superior to the parental clone in total tuber weight. The tuber yield of one protoclone, designated No. 307, did increase by 25 percent, but inconsistencies between different plantings rendered the increase insignificant. In 1980 follow-up field plots at the same location showed that once again protoclone No. 307 exceeded the yield of the parental Russet Burbank clone, but again the differences were not statistically significant. More years of experience are needed to determine whether No. 307 or any other protoclone offers a genuine yield advantage over the parent.

What is clear is that superior protoclonal performance at one geographic site is no guarantee of a similar advantage at another site. For example, in 1980 the same protoclones we tested in North Dakota were planted at a field plot in Colorado by Richard Zink of Colorado State University. He found that protoclone No. 307 did less well there than the parent, whereas other clones that did poorly in North Dakota equaled or exceeded the yield of the parental Russet Burbank clone. Site specificity has been a common phenomenon for the 65 protoclones over the past five years, and therefore it appears to be predictable. If it is, the Russet Burbank protoclones would resemble those varieties of potato in which the yield advantage among competitive lines is commonly realized only in certain geographic or environmental regions.



**FUSION OF PROTOPLASTS extracted from genetically different plants may lead eventually to improvements in the regenerated protoclones. In recent experiments done by the author and his colleagues an albino protoplast from a leaf of the Russet Burbank potato plant (*colorless body*) has been fused with a normal protoplast from the leaf of a tomato plant (*green body*). The aim is to incorporate the tomato's genes for disease resistance into the potato genome.**

Given the experimental data obtained so far for protoclones of the Russet Burbank potato, one is left with several questions. The first is: Will a similar range of phenotypic variation be observed in potatoes other than the Russet Burbank? Our group has observed similar variability in protoclones of three additional potato lines. On the other hand, Gerhard Wenzel and his colleagues at the Max Planck Institute for Plant Genetics in Cologne have reported that for the diploid potato lines they studied, phenotypic variability was not a regular feature of the protoclonal populations. Since they did not make detailed measurements of horticultural or disease-resistance traits, however, the apparent discrepancy between their results and ours has yet to be resolved. It is possible that the specific genotypes studied or differences in cultural methodology could influence the frequencies and types of variation.

A second question has to do with the origin of the variability: Do the genetic changes that appear in the protoclones result from genetic differences in leaf cells that were already present before the cells were isolated as protoplasts, or does the process of protoplast culture induce genetic change, or both? The question is a difficult one. It has been known for some time that gross genetic changes, such as alterations in chromosome number, can appear in the leaf cells of some plants. If plants are regenerated from cells with many such alterations, they usually fall into the wild-aberrant class. Point mutations and somatic recombination (the rearrangement of genes in a somatic cell) have also been described in leaf cells. Hence both gross and subtle forms of genetic change are known to occur in the leaf cells of some plant species; if protoplasts of the altered cells are cultured, the genetic changes can be individually recovered by plant regeneration.

Recently David Ingram and Richard Brettel of the University of Cambridge suggested that the observed variability of potato protoclones is evidence of considerable genetic variation in the leaf-cell populations of all vegetatively propagated plants. If this hypothesis is correct, they speculated, the leaves represent a vast repository of essentially untapped genetic variation. The idea is attractive, but other possibilities too could explain the observed variability. For example, the process of protoplast

isolation and the initial cultural conditions certainly must be traumatic to a cell that has evolved under totally different conditions. Such a shock could foster errors in the first round (or rounds) of DNA replication, which could later be expressed in the whole plant as some form of phenotypic variation.

Moreover, protoplasts are not cultured singly; at the outset there are typically 10,000 or more cells per milliliter. Some of the cells die and liberate their contents into the culture medium. Plant cells synthesize substances that are known to be mutagens, such as alkaloids and flavonoids; if these substances became sufficiently concentrated in the culture medium, they could contribute to genetic alteration.

One way to investigate this possibility is to compare plants regenerated from protoplasts with those regenerated from whole-cell calluses grown from the same potato leaves. To date our results suggest that plant populations regenerated from leaf calluses have approximately the same percentage of wild aberrants as protoclonal populations do; phenotypic variants, however, have not been detected under greenhouse conditions. Final evaluation of these clones will be done under field conditions to determine whether any variants are present in the population. In any case the range and frequency of phenotypic variation seem substantially greater in plants grown from protoplasts than in those grown from leaf-cell calluses.

The most fundamental question is: What genetic mechanisms in somatic potato cells are responsible for the observed phenotypic variation in the protoclones? For now one can only speculate. As I mentioned above, potato cultivars are genetically very complex, and most of the variable traits studied so far are multigenic in nature. The traditional method of identifying the source of a genetic change is through analysis of sexual progeny for inheritance of the variable trait. In potato plants the approach is largely ineffective because as much variability or more for the given trait is present in the progeny of a self-pollinating line as is present in the protoclonal populations. A genetic study by Bikrum S. Gill of Kansas State University, however, has revealed that although each of the protoclones studied to date has 48 chromosomes, structural changes are apparent in the chromosomes of some of the cells [*see illustration on page 160*]. It is premature to conclude that alterations in chromosome structure or other such effects are primary contributors to the observed variability, but they do deserve serious consideration.

Will the regeneration of plants from protoplasts find a prominent place in crop-improvement schemes? It seems that definite opportunities exist, but much further study is needed. A few hundred protoclones cannot by themselves serve as a basis for evaluating the worth of a method. Moreover, it is the possible usefulness of the protoclonal method itself, and not whether one protoclone or another can solve a particular agricultural problem, that is the central issue. When information becomes available for protoclones of other crop plants, we shall be in a better position to evaluate the overall applicability of protoplast culture to crop improvement.

Nevertheless, the fact that important quantitative plant traits are subject to modification is a critical aspect of the technique's potential, and the prospects for success will brighten as it becomes possible to select cells and colonies in the laboratory more effectively. This would enable the experimenter to screen millions of genotypes within the confines of a single Petri dish and to recover the rare individual plant with a beneficial alteration. Some workers have already accomplished this task in tobacco, the petunia and a few other plants by screening protoplast populations for alterations in simply inherited traits such as resistance to herbicides. Complex traits are far more difficult to identify in this way, however, and selective systems must rely on some physiological or genetic peculiarity of the species.

As an example of a potentially useful trait of this kind, consider the potato cultivar named Bison, which produces tubers with a red skin. When leaf-cell protoplasts of this variety are cultured under certain conditions, they lose their green color and accumulate the red pigments known as anthocyanins. The expression of the gene (or genes) for anthocyanin production could serve as a genetic marker in screening experiments where the identification of the Bison genome would be valuable.

It is also possible to induce small protoplast-derived potato calluses with primordial shoots to form tubers. By further developing and standardizing this phenomenon it should be possible to devise selective conditions for protoclones whose ability to form tubers is less sensitive to the effects of temperature or period of daylight.

Finally, there is the wide range of possibilities for creating new hybrid lines by fusing the protoplasts of genetically different plants prior to regeneration. Preliminary experiments along these lines are being done by our group and others. For example, we have recently succeeded in fusing a leaf-cell protoplast from a potato plant with one from a tomato plant. The goal of this particular effort is to introduce specific genes (for instance, those for disease resistance) from the tomato to the potato. Such an achievement would make possible the shuffling of genes between plant species that are now sexually incompatible, thereby increasing the size of the germ-plasm pool available to the plant breeder.

# THE JAPANESE BEETLE.

Thirty-two years ago a homely, funny shaped car found a home in America.

A little car, that promised maximum efficiency, dependability and economy, at a truly economical price.

But over the years even Volkswagens change. To the point that the German Beetle has gone the way of the American Passenger Pigeon. Leaving American passengers to pay a lot more for a Volkswagen than they used to.

At Isuzu, we've been working for over 65 years with a single philosophy in mind. To build the toughest, most economical and dependable car any carmaker can build.

A practical car. That can deliver over 40* estimated MPG, over 50 MPG on the highway. Yet go for under $6,800 in the showroom.**

Of course, driving a practical car like an Isuzu Diesel requires giving certain things up.

Like the necessity of paying for expensive tune-ups. Or the luxury of paying for what some car companies regard as options. (We think there's nothing optional about having front disc brakes or a diesel fuel/water separator.)

A car that's quick enough to be one of the fastest in its class, yet durable enough to last for years.

The Japanese Gas and Diesel Beetle from Isuzu.

Some people may consider its exterior a trifle plain. But then, an Isuzu has always had thick skin.

ISUZU

*Isuzu Diesel Coupe [41] EST MPG, 51 MPG highway. Use estimated MPG figures for comparison. Your own mileage may vary due to speed, weather, trip length and road conditions. California Gas MPG will be less. **Priced from $6,799. Manufacturer's suggested retail price P.O.E. excluding tax, license, transportation and California extended emissions warranty.

# The $149^{95} personal computer.


## Introducing the Sinclair ZX81

If you're ever going to buy a personal computer, now is the time to do it.

The new Sinclair ZX81 is the most powerful, yet easy-to-use computer ever offered for anywhere near the price: only $149.95* completely assembled.

Don't let the price fool you. The ZX81 has just about everything you could ask for in a personal computer.

**A breakthrough in personal computers**

The ZX81 is a major advance over the original Sinclair ZX80—the world's largest selling personal computer and the first for under $200.

In fact, the ZX81's new 8K Extended BASIC offers features found only on computers costing two or three times as much.

Just look at what you get:

- Continuous display, including moving graphics
- Multi-dimensional string and numerical arrays
- Mathematical and scientific functions accurate to 8 decimal places
- Unique one-touch entry of key words like PRINT, RUN and LIST
- Automatic syntax error detection and easy editing
- Randomize function useful for both games and serious applications
- Built-in interface for ZX Printer
- 1K of memory expandable to 16K

The ZX81 is also very convenient to use. It hooks up to any television set to produce a clear 32-column by 24-line display. And you can use a regular cassette recorder to store and recall programs by name.

*Plus shipping and handling. Price includes connectors for TV and cassette, AC adaptor, and FREE manual.

**If you already own a ZX80**

The 8K Extended BASIC chip used in the ZX81 is available as a plug-in replacement for your ZX80 for only $39.95, plus shipping and handling—complete with new keyboard overlay and the ZX81 manual.

So in just a few minutes, with no special skills or tools required, you can upgrade your ZX80 to have all the powerful features of the ZX81. (You'll have everything except continuous display, but you can still use the PAUSE and SCROLL commands to get moving graphics.)

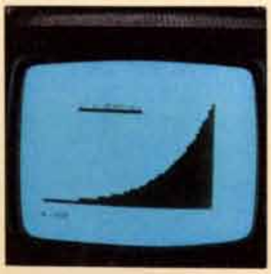
With the 8K BASIC chip, your ZX80 will also be equipped to use the ZX Printer and Sinclair software.

**Order at no risk****

We'll give you 10 days to try out the ZX81. If you're not completely satisfied, just return it to Sinclair Research and we'll give you a full refund.

And if you have a problem with your ZX81, send it to Sinclair Research within 90 days and we'll repair or replace it at no charge.

**Does not apply to ZX81 kits.



**NEW SOFTWARE:** Sinclair has published pre-recorded programs on cassettes for your ZX81, or ZX80 with 8K BASIC. We're constantly coming out with new programs, so we'll send you our latest software catalog with your computer.

**ZX PRINTER:** The Sinclair ZX Printer will work with your ZX81, or ZX80 with 8K BASIC. It will be available in the near future and will cost less than $100.

**16K MEMORY MODULE:** Like any powerful, full fledged computer, the ZX81 is expandable. Sinclair's 16K memory module plugs right onto the back of your ZX81 (or ZX80, with or without 8K BASIC). Cost is $99.95, plus shipping and handling.

**ZX81 MANUAL:** The ZX81 comes with a comprehensive 164-page programming guide and operating manual designed for both beginners and experienced computer users. A $10.95 value, it's yours free with the ZX81.


# The $99^{95}$ personal computer.

## Introducing the ZX81 kit

If you really want to save money, and you enjoy building electronic kits, you can order the ZX81 in kit form for the incredible price of just $99.95.* It's the same, full-featured computer, only you put it together yourself. We'll send complete, easy-to-follow instructions on how you can assemble your ZX81 in just a few hours. All you have to supply is the soldering iron.

**How to order**

Sinclair Research is the world's largest manufacturer of personal computers.

The ZX81 represents the latest technology in microelectronics, and it picks up right where the ZX80 left off. Thousands are selling every week.

We urge you to place your order for the new ZX81 today. The sooner you order, the sooner you can start enjoying your own computer.

To order, simply call our toll free number, and use your MasterCard or VISA.

To order by mail, please use the coupon. And send your check or money order. We regret that we cannot accept purchase orders or C.O.D.'s.

**CALL 800-543-3000.** Ask for operator #509. In Ohio call 800-582-1364. In Canada call 513-729-4300. Ask for operator #509. Phones open 24 hours a day, 7 days a week. Have your MasterCard or VISA ready.

These numbers are for orders only. For information, you must write to Sinclair Research Ltd., 2 Sinclair Plaza, Nashua, NH 03061.

**sinclair**

| AD CODE **05SA** | PRICE† | QTY. | AMOUNT |
|---|---|---|---|
| ZX81 | $149.95 | | |
| ZX81 Kit | 99.95 | | |
| 8K BASIC chip (for ZX80) | 39.95 | | |
| 16K Memory Module (for ZX81 or ZX80) | 99.95 | | |
| Shipping and Handling | 4.95 | | $4.95 |
| | | TOTAL | |

**MAIL TO:** Sinclair Research Ltd., One Sinclair Plaza, Nashua, NH 03061.

NAME ______________________

ADDRESS ______________________

CITY/STATE/ZIP ______________________

† U.S. Dollars

# THE AMATEUR SCIENTIST

*What makes you sound so good when you sing in the shower?*

by Jearl Walker

Even a bad singer sounds good in the shower, at least to himself. I certainly qualify as unskilled in singing, but in the shower my voice is a thing of beauty, making it easy for me to overrate my ability. The shower stall is so kind to musical sounds that a friend of mine practiced her violin lessons in one.

The secret of the shower stall as an aid to musical performance lies in its acoustic resonance, which greatly increases the strength of the sound. Like any other sound, a sound generated by the vocal apparatus travels through the air as a wave of variations in pressure. If the voice could be held steady at one frequency (or pitch), the wave could be represented by a sinusoidal curve. The curve depicts the regular variation in the pressure of the air along the path of the sound wave from relatively high (above the ambient pressure) to relatively low.

The variation results from slight oscillations of the air molecules parallel to the path of the sound wave. The oscillations generate a fairly high density of air (a high pressure) at some places and a fairly low density (a low pressure) at others. The entire pattern of high and low pressure travels away from the mouth at the speed of sound, which at room temperature is about 340 meters per second.

An associated factor is the frequency of the sound, which is measured by the number of times per second a high-pressure section of the wave passes an imaginary point along the path. If the number is 500, the frequency is 500 hertz (cycles per second). The range of hearing for a young human adult is roughly from 20 hertz to 20,000.

A wave moving from one place to another is called a traveling wave. Another type of wave, the standing wave, is important to singing and acoustic resonance. Consider a cylindrical pipe open at both ends. In the simplest approximation its diameter makes no difference with respect to resonance, but its length is important. A traveling sound wave of a single frequency is fed continuously into the pipe. Imagine following one high-pressure segment of the wave along the pipe. At the far end of the pipe part of the sound is reflected backward, even though the pipe is open. The rest of the sound continues on outward. The reflected part now overlaps the waves still moving toward the far end of the pipe and interferes with them. At a given instant a particular section of the air has a tendency to rise in pressure because of the wave going one way and to fall in pressure because of the wave going the other way. The result is a constantly shifting, complex pattern of pressure variations. The sound coming out of the tube is no louder than the sound that went into it.

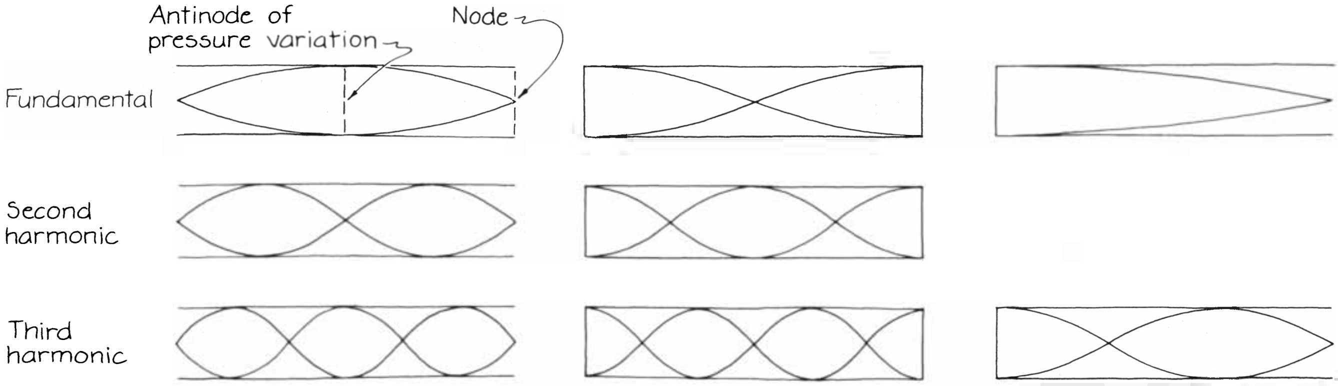
At certain frequencies a much simpler pattern of pressure variations develops. The pattern is the harmonic frequencies for the pipe. These frequencies are integral multiples of the lowest frequency: the fundamental. The pattern is a standing wave. When such a wave is created, the sound coming out from the pipe is much louder than the sound that went into it. This condition is called acoustic resonance.

Suppose the fundamental frequency is directed into the pipe. The interference of the waves traveling in opposite directions ensures that at the ends of the pipe (where the waves reflect) the pressure never varies. Such a place is called a node. The pressure changes most in the middle of the pipe, varying smoothly from high to low (high when the oppositely traveling waves interact to increase the pressure, low when they interact to decrease it). Such a place is an antinode. The pattern of pressure variations along the pipe is called a standing wave because the nodes and antinodes stand still.

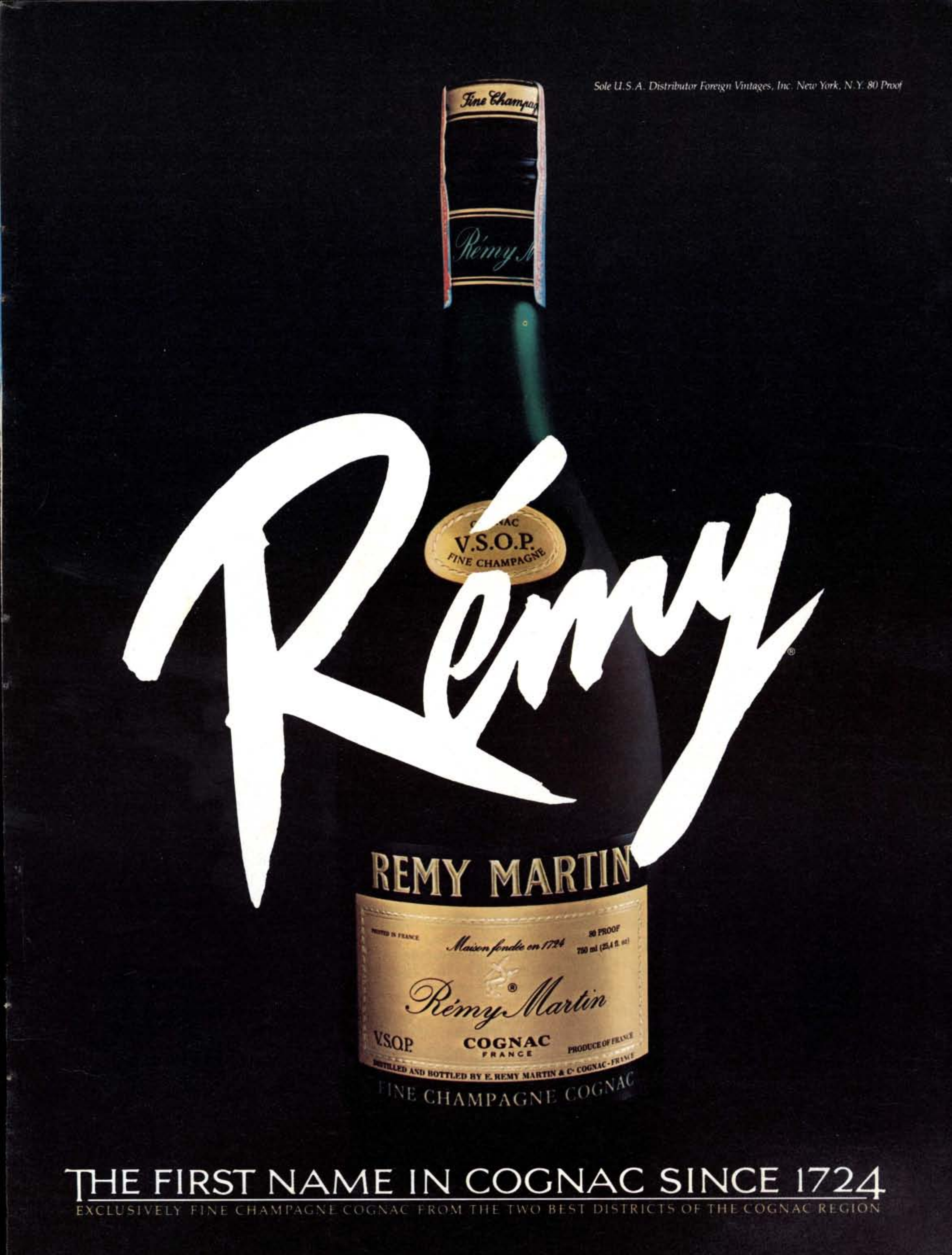
A standard graphical representation of standing waves in a pipe appears in the illustration below. The points where the curves cross are nodes and the points where the curves have the greatest separation are antinodes. (One can be misled by the standard representation into thinking that the molecules oscillate up and down in the pipe. Actually the direction of their oscillation is along the length of the pipe.)

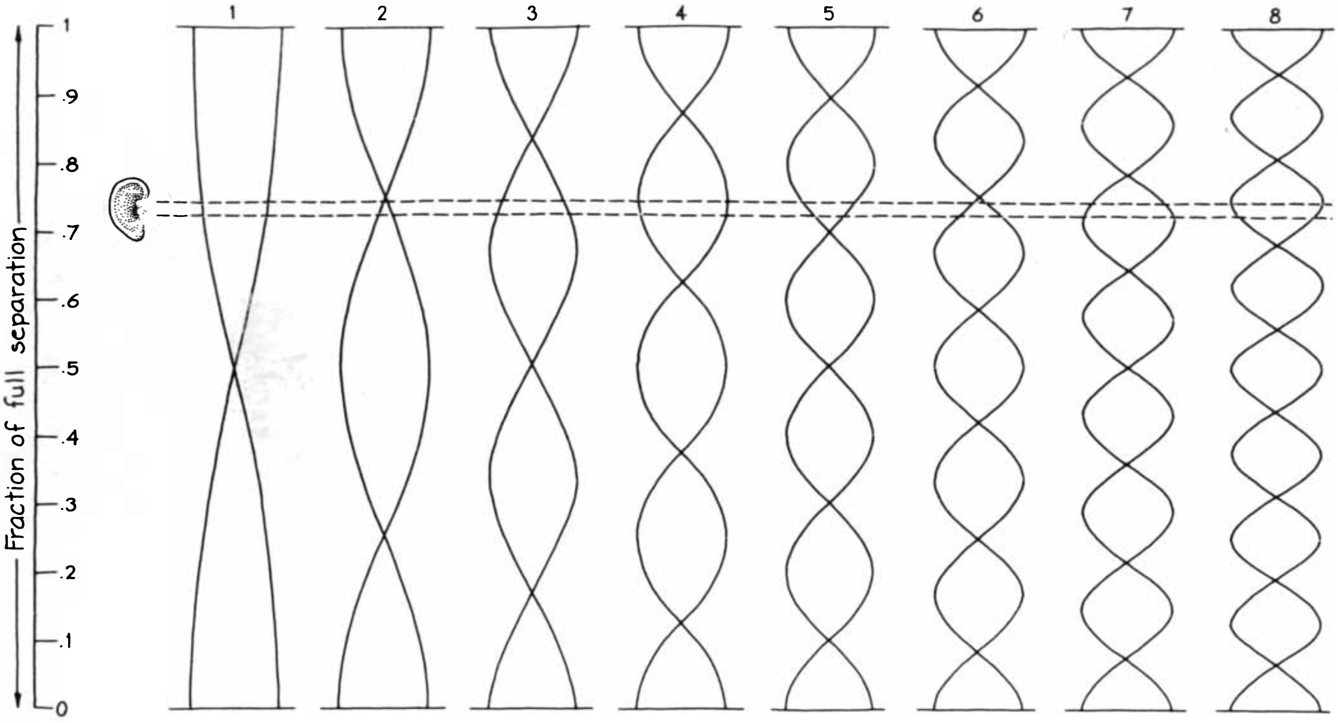
Things get more complicated as the higher frequencies of a harmonic series are directed into the pipe. With the second harmonic (twice the frequency of the fundamental) the standing wave shows three nodes, one node at each end and one in the middle. Antinodes lie between the nodes. The third harmonic (three times the frequency of the fundamental) has four nodes and three antinodes.

Acoustic resonance is significant because it creates a large standing wave of




*Acoustic resonances in three types of pipe*


Sole U.S.A. Distributor Foreign Vintages, Inc. New York, N.Y. 80 Proof
Fine Champagne
Rémy Martin
V.S.O.P.
FINE CHAMPAGNE
Rémy®
REMY MARTIN
Maison fondée en 1724
80 PROOF
Rémy Martin
V.S.O.P.
COGNAC
FRANCE
PRODUCE OF FRANCE
DISTILLED AND BOTTLED BY E. REMY MARTIN & C° COGNAC - FRANCE
FINE CHAMPAGNE COGNAC
THE FIRST NAME IN COGNAC SINCE 1724
EXCLUSIVELY FINE CHAMPAGNE COGNAC FROM THE TWO BEST DISTRICTS OF THE COGNAC REGION




*The first eight harmonics between the floor and the ceiling of a shower stall*

sound, which is louder than the sound directed into the pipe. The pipe is therefore an amplifier. In my simple model the values of the harmonic series in a pipe depend only on the speed of sound and the length of the pipe. Heating the air, which makes the sound travel faster, shifts the harmonic frequencies upward. Shortening the pipe has the same effect. That is why the fundamental is higher in a short organ pipe than it is in a long one.

So far I have discussed only a pipe with both ends open, but for singing in the shower two other configurations are important. One of them is a pipe with both ends closed; its harmonic series is the same as the series in the open pipe if the pipe is the same length. The standing waves are different but their frequencies are the same. Now an antinode appears at each end of the pipe because the pressure variation is highest at a wall. A picture of the first harmonic (the fundamental) shows an antinode at each end and a node in the middle.

When the pipe is closed at only one end, the harmonic series is subtly modified. A drawing of the first harmonic still follows the rules I have outlined, showing an antinode at the closed end and a node at the open end. The same is true for the next harmonic, but the result is that its frequency is three times the frequency of the first harmonic. Sometimes this standing wave is called the second harmonic because it is the second-lowest frequency possible for resonance in such a pipe. I find this designation confusing. Since the harmonic's frequency is three times that of the fundamental, I shall continue to call it the third harmonic. Only the odd harmonics (first, third, fifth and so on) are possible for resonance in a pipe with one open end. Standing waves for the even harmonics are impossible because of the rules governing pressure variations at the ends of a pipe.

Singing entails creating a resonance in what is essentially a pipe with one open end. The pipe is not the straight, cylindrical one I have been describing, but the model still serves well. The sound originates with the passage of air pushed out of the lungs through the vocal cords (which are not cords but thin membranes). The air emerges from the vocal cords in a series of pulses. The frequency of the pulses depends mainly on the tension of the cords; a higher tension results in more frequent bursts of air and so in higher frequency.

The vocal cords vibrate in a harmonic series of frequencies, yielding a harmonic series of sound waves encompassing the fundamental and all the higher harmonics. The most intense wave is the fundamental.

These waves pass through the vocal tract, consisting of the larynx, the pharynx and the mouth. The vocal tract can be modeled as a pipe with one closed end (at the vocal cords) and one open end (at the mouth). It therefore has a harmonic series of frequencies at which standing waves can be excited. Among the frequencies entering the vocal tract from the vocal cords are some that fall on or near these resonant frequencies. Those frequencies are relatively loud.

A singer generates the loudest sound when the fundamentals of the vocal cords and those of the vocal tract match. A good singer can achieve the match, probably with little conscious effort, in several ways. The tension of the vocal cords can be adjusted somewhat to regulate the fundamental at that point. Once that fundamental is established, however, further matching must be done by varying the shape of the vocal tract.

Here the physics of the acoustic resonance is not as simple as it is in the pipe models, where I ignored the effect of diameter. By appropriate reshaping of the vocal tract a singing adult male can shift the fundamental of the tract to any frequency between 250 and 700 hertz. The third harmonic (bear in mind that only the odd harmonics develop in this kind of pipe) varies between 700 and 2,500 hertz.

I am unskilled in matching the fundamental of my vocal tract to the one being generated in my vocal cords. If I sought to project to an audience a note of, say, 500 hertz, I would have to scream to make it sufficiently audible. A well-trained singer would make the match and so would benefit from the amplification arising from the resonance of the vocal tract. An advantage of singing in the shower is that the unskilled vocalist is aided by the resonances generated between the surfaces of the shower stall.

The stall is in effect a pipe with both ends closed. It differs from the other pipes I have been describing in having three pairs of ends: (1) the floor and the

ceiling, (2) one set of parallel walls and (3) the other set of parallel walls. (I treat the shower door or curtain as a solid wall.) A standing wave of sound can be excited between any pair of walls. A depiction of the fundamental for, say, the floor-ceiling pair shows antinodes at the floor and the ceiling and a node halfway between the two. The second harmonic (a pipe with both ends closed can have even harmonics) has three antinodes (floor, ceiling and halfway between) and two nodes (between the antinodes). Similar patterns arise for the other two pairs of walls.

Although the association of singing in the shower and acoustic resonance has long been known, the only study I have found of it is an abstract of a talk given by Daniel W. Haines of the University of South Carolina at the Second Conference on the Teaching of Acoustics and the Physics of Sound and Music in April, 1976. His work revealed several interesting features of the phenomenon. For example, suppose the aim is to generate resonance between the floor and the ceiling of the shower stall. If the source of the sound is halfway between the two, there cannot be any resonance because that is the location of a node in the associated standing wave. A node, which is by definition the absence of pressure variations, cannot arise where the sound source itself is creating variations in pressure. To be closer to an antinode of a standing wave one would do better to sing near the floor or the ceiling (although it might be hazardous and might expose one to the ridicule of an onlooker, however friendly).

The second harmonic and other even harmonics are possible if the source of sound is halfway between the floor and the ceiling of the shower stall. Any other height for the source might give rise to many harmonics, some more intense than others, depending on how close the source is to an antinode. The only impossible harmonics are those where the source and a node coincide. (My explanation here may be faulty. If the source is quite small and radiates sound throughout the stall, its location is probably unimportant. I think, however, that the human head is large enough for its location to matter.)

Haines also noted that the singer can sound good or bad to himself according to the position of his ears. If they are at a node, the vocalist cannot hear the standing wave (since hearing is actuated by pressure waves on the eardrums). If they are at an antinode, the impact on the eardrums is at a maximum.

The first eight harmonics between the floor and the ceiling are shown in the illustration on page 172. The mouth and ears of the singer are approximately a fourth of the distance from the ceiling to the floor. In theory he could both excite and hear the fundamental because his mouth and ears are not at a node of the associated standing wave. Since they are also not at an antinode, he cannot excite or hear a strong fundamental. The second harmonic cannot be excited or heard because his head is at one of its nodes. Among the standing waves depicted in the illustration the third, fourth, seventh and eighth harmonics would come out best.

The singer is somewhat more mobile when it comes to the standing waves between a pair of walls. By moving toward or away from a wall one can modify the spectrum of harmonics that are excited and heard. At the center of the stall the fundamental from each pair of walls would be absent, since both fundamentals have nodes there. The second harmonics have antinodes at that position and would come into play. By moving closer to a wall the singer would eliminate the second harmonic but would excite the third.

The harmonics of the walls are complicated by the horizontal separation between the mouth and the ears. For some harmonics the mouth might be in the right place to excite a standing wave while the ears were at a node. (To study this problem adequately one must also consider the distortion of the standing wave by the head.)

A further complication arises because the harmonic frequencies are not precisely as they are shown in the illustration above. The numbers there depend on the speed of sound, which may change as the air in the stall gets warmer from the shower. Even without any change in the speed of sound the frequencies are not sharp but extend over a range of about 10 hertz. A standing wave is excited most efficiently when the singer's voice is pitched at a frequency listed in the table; the wave is weaker if he is slightly off the theoretical value.

Another complication results from the fact that the singer's body occupies space in the stall, reflecting some of the sound. It is best to assume that this complication is secondary to the principal mechanism of creating standing waves. Modeling the resonances as if the stall is a pipe with closed ends at least makes possible an approximation of the resonant frequencies.

The frequencies tabulated in the illustration above are for my shower stall, which has flat tiled walls, floor and ceiling. One wall is a glass door. I calculated these harmonic frequencies (1,000 hertz or less) assuming the speed of sound to be 346 meters per second in the stall. The frequency of the fundamental is then the speed divided by twice the distance between a pair of reflecting surfaces.

Since the floor and the ceiling are farther apart than the walls, the floor-ceiling fundamental is lower in frequency than the fundamentals for the walls. The higher harmonics are calculated by multiplying the fundamental frequency by an integer. The second harmonic is twice the fundamental, the third three times the fundamental and so on.

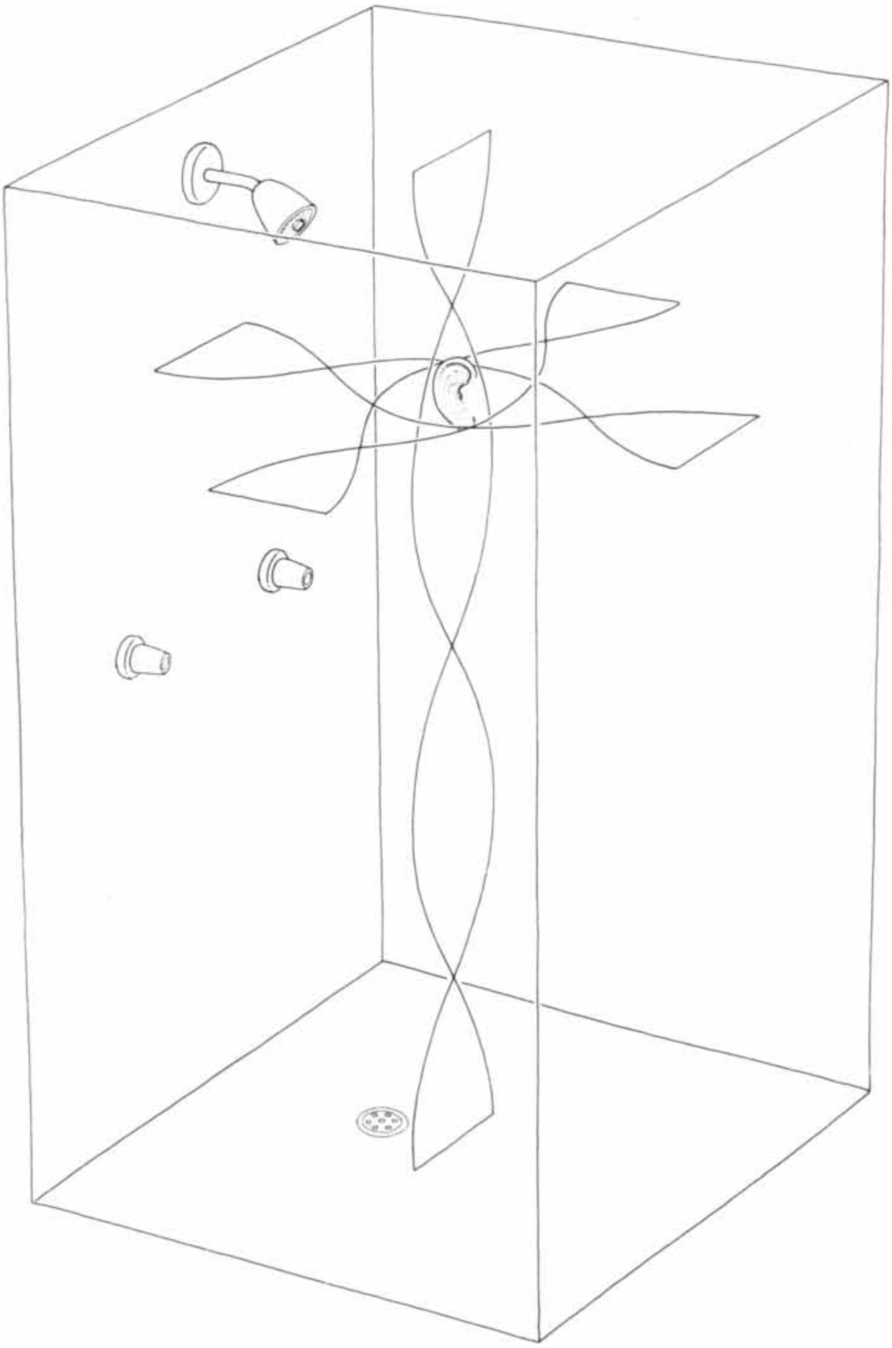
Now, suppose I begin to sing in the shower, somehow adjusting my vocal cords so that their fundamental is 330 hertz but failing to match the fundamental of my vocal tract. The vocal cords deliver frequencies at integral multiples of the fundamental, namely 330, 660, 990 hertz and so on. Suppose the fundamental of my vocal tract is set at 450 hertz. Its next resonance is at 1,350 hertz (the third harmonic). Outside the shower stall my singing would sound weak, since none of the frequencies from the vocal cords resonate in the vocal tract. Inside the stall, however, the frequencies at 330 and 660 hertz create

| | Floor to ceiling, 2.095 meters | Wall to wall, 1.375 meters | Wall to wall, 1.010 meters |
|---|---|---|---|
| Harmonic | Frequency (hertz) | Frequency (hertz) | Frequency (hertz) |
| 1 | 83 | 126 | 171 |
| 2 | 165 | 252 | 343 |
| 3 | 248 | 377 | 514 |
| 4 | 330 | 503 | 685 |
| 5 | 413 | 629 | 856 |
| 6 | 495 | 755 | 1,028 |
| 7 | 578 | 881 | |
| 8 | 661 | 1,007 | |
| 9 | 743 | | |
| 10 | 826 | | |
| 11 | 908 | | |
| 12 | 991 | | |
| 13 | 1,074 | | |

*The harmonic frequencies for the stall* standing waves between the floor and the ceiling, the lower frequency exciting the fourth harmonic and the higher one the eighth. (The sound at 990 hertz should excite the 12th harmonic, but my mouth is at the node for that resonance.) I therefore sound great in spite of the absence of resonance from my vocal tract.

Wanting to experiment with the harmonics in my shower stall but not being able to hold a note for very long and not being skilled in determining the frequency of the note, I substituted for my voice an audio oscillator and a loudspeaker. The oscillator emitted a sinusoidal wave at whatever frequency I set on the dial. The signal, strengthened by an amplifier, was directed to a speaker cone I could put anywhere in the stall. The cone had an aperture of 19 by 12 centimeters; a smaller cone would not emit the low frequencies.

If you decide to experiment in this way, bear in mind that around water an electrical appliance is dangerous. I turned off the water at the valve supplying the shower. Another danger is that the sound of a standing wave excited in the stall can be loud enough to hurt your ears, particularly if the amplifier is set at high volume. I wore a set of headphones of the type that keeps out noise.



*Patterns for a few harmonics in a shower stall*

By tuning the oscillator through the range from about 200 to 1,000 hertz I could hear the resonances come and go. When the oscillator was out of resonance, the sound in the stall was relatively weak. At the resonances it increased, sometimes dramatically.

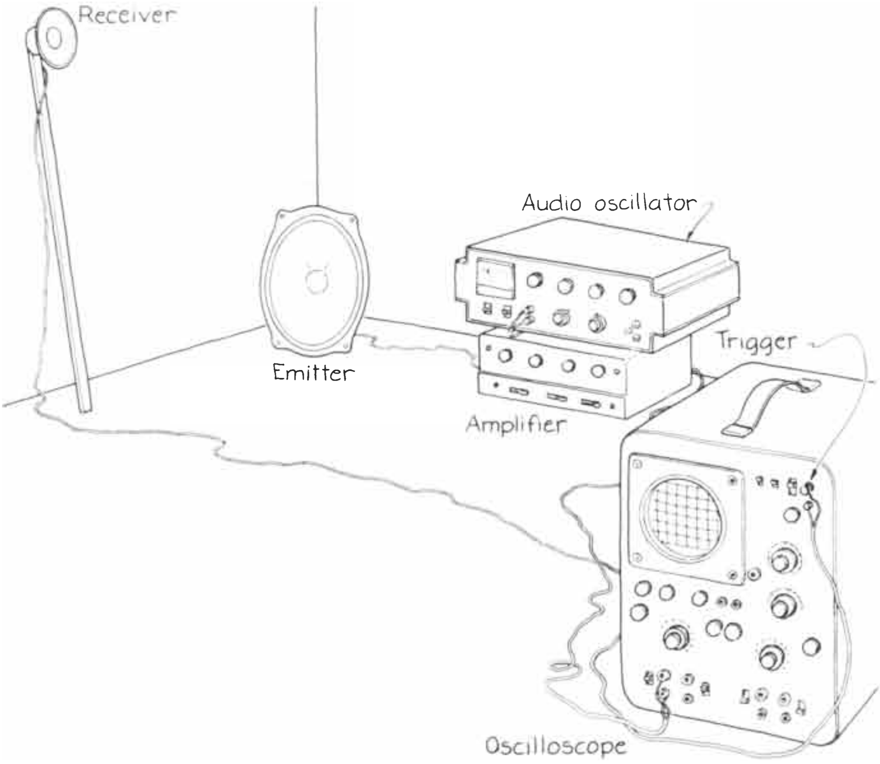
Keeping the oscillator at some resonant frequency, I could alter my perception of the sound by moving my head around in the stall. As my ears moved from an antinode to a node the level of sound dropped sharply. With a few experimental moves I could usually determine whether the standing wave was between a pair of walls (and, if it was, which pair) or between the floor and the ceiling.

The loudspeaker's efficiency at exciting a standing wave depended partly on how I pointed it. Pointed upward it could only weakly excite a standing wave between the walls even though it was set at the right frequency. I found that when I put the speaker at an angle in a corner on the floor, the sound that spread from the cone was directed more or less correctly for any possible standing wave.

To assess the resonance frequencies more objectively I installed an oscilloscope and a second speaker, which was slightly smaller than the emitting speaker. (A much smaller speaker would fail to pick up the lower frequencies.) This speaker intercepted some of the sound waves in the stall and generated a wave pattern on the screen of the oscilloscope. The vertical component of the pattern reflected the amplitude of the sound wave; the horizontal axis was the oscilloscope's time sweep.

I connected the external trigger ports of the oscilloscope to the audio oscillator so that the sweep across the screen was synchronized with the output from the oscillator. When I set the frequency from the oscillator at a harmonic for the shower stall, the pattern on the oscilloscope increased in amplitude. I measured the range of frequencies exciting a resonance by finding the two frequencies (one below the optimum frequency and one above it) at which the pattern was reduced to half the maximum amplitude.

To map out the standing waves in the stall I moved the receiving speaker either horizontally or vertically between the walls. Looking for a harmonic between the floor and the ceiling, I pointed the speaker upward and moved it vertically. For a standing wave between two walls I pointed the speaker at one wall and moved it back and forth horizontally. In this way I could ascertain whether a resonance was a standing wave between the walls or a standing wave between the floor and the ceiling. To minimize distortion caused by the movement




*A setup for measuring harmonics electronically*

of my body I taped the receiving speaker to a meter stick that I maneuvered in the stall. Even so, my body and the equipment distorted the standing waves enough for me to notice it.

I found and identified most of the harmonics in my table of values except for some of the lower frequencies at which the speakers were inefficient. The frequency at which I got the best pattern on the oscilloscope was not always identical with the theoretical value, but the range around that frequency usually fell across the values in the table. I should have liked to see how the harmonic frequencies shifted as I warmed the air in the stall by turning on the hot water, but for such an experiment the danger of electrocution is too great.

One's perception of the harmonics in a shower stall differs from my more objective assessment of them with electronic apparatus. The ear perceives a pure frequency as being made up of several frequencies because the perceptual apparatus of hearing does not respond to sound waves precisely. The system is said to be nonlinear, which means that in receiving and processing a signal it distorts it.

Because of the nonlinearity of the hearing process some people can hear frequencies that are not there. Suppose two frequencies reach the ear; the listener not only hears those two frequencies but also may perceive the difference between them. For example, if the two signals are at 500 and 650 hertz, a signal of 150 hertz may also be perceived. Now let $f1$ and $f2$ stand for the frequencies actually sounded; the hearer may also perceive tones at values of $2f1 - f2$, $3f1 - 2f2$ and so on. In other words, signals of low frequency may be perceived when only higher frequencies are actually reaching the ear.

These difference tones might enrich one's singing in the shower. Suppose I sing with the fundamental of my vocal cords at 289 hertz. The integral multiples of the fundamental are 578, 867 and so on. As usual, however, I fail to achieve a match between the fundamental of my vocal tract and the fundamental of the vocal cords. Instead the tract has a fundamental of 330 hertz and thus a third harmonic of 990.

The tract amplifies the sound at 289 hertz badly and at 578 worse, making the singing weak. The sound at 578 hertz, however, excites the seventh harmonic between the floor and the ceiling. A difference tone may then enhance the sound at 289 hertz. The difference between the 289-hertz signal (weakly amplified by my vocal tract) and the 578-hertz signal (amplified by the stall) is 289 hertz. Provided I can perceive a difference tone, the signal at 289 hertz sounds louder than it actually is.

I shall leave further investigation of difference tones to you. If you find interesting effects, I should like to hear from you. I should also like to be told about any flaws in my basic argument about why someone singing in the shower often sounds much better than he does when he is singing somewhere else.

# BIBLIOGRAPHY

*Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.*

## METAMAGICAL THEMAS


COSMIC VIEW: THE UNIVERSE IN FORTY JUMPS. Kees Boeke. John Day Company, 1957.

THE LORE OF LARGE NUMBERS. Philip J. Davis. Random House, Inc., 1961.


## LIFE EXPECTANCY AND POPULATION GROWTH IN THE THIRD WORLD


CAN HEALTH AND NUTRITION INTERVENTIONS MAKE A DIFFERENCE? Davidson R. Gwatkin, Janet R. Wilcox and Joe D. Wray. Overseas Development Council Monograph No. 13, February, 1980.

THE POLICY IMPLICATIONS OF FIELD EXPERIMENTS IN PRIMARY HEALTH AND NUTRITION CARE. Davidson R. Gwatkin, Janet R. Wilcox and Joe D. Wray in *Social Science & Medicine,* Vol. 14C, No. 2, pages 121–128; June, 1980.

HOW MANY DIE? A SET OF DEMOGRAPHIC ESTIMATES OF THE ANNUAL NUMBER OF INFANT AND CHILD DEATHS IN THE WORLD. Davidson R. Gwatkin in *American Journal of Public Health,* Vol. 70, No. 12, pages 1286–1289; December, 1980.

INDICATIONS OF CHANGE IN DEVELOPING COUNTRY MORTALITY TRENDS: THE END OF AN ERA? Davidson R. Gwatkin in *Population and Development Review,* Vol. 6, No. 4, pages 615–644; December, 1980.


## QUARKONIUM


GAUGE THEORIES OF THE FORCES BETWEEN ELEMENTARY PARTICLES. Gerard 't Hooft in *Scientific American,* Vol. 242, No. 6, pages 104–138; June, 1980.

OBSERVATION OF THE RADIATIVE TRANSITION $\psi \rightarrow \gamma$ E(1420). D. L. Scharre et al. in *Physics Letters,* Vol. 97B, No. 2, pages 329–332; December, 1980.

RECENT EXPERIMENTAL ADVANCES IN POSITRONIUM RESEARCH. Arthur Rich in *Reviews of Modern Physics,* Vol. 53, No. 1, pages 127–165; January, 1981.

OBSERVATION OF AN $\eta'_c$ CANDIDATE STATE WITH MASS $3592 \pm 5$ MEV. C. Edwards et al. in *Physical Review Letters,* Vol. 48, No. 2, pages 70–73; January 11, 1982.


## NMR IMAGING IN MEDICINE


IMAGE FORMATION BY INDUCED LOCAL INTERACTIONS: EXAMPLES EMPLOYING NUCLEAR MAGNETIC RESONANCE. P. C. Lauterbur in *Nature,* Vol. 242, No. 5394, pages 190–191; March 16, 1973.

NUCLEAR MAGNETIC RESONANCE IMAGING IN MEDICINE. Edited by L. Kaufman, L. E. Crooks and A. R. Margulis. Igaku-Shoin, Ltd., Tokyo, 1981.

PROCEEDINGS OF AN INTERNATIONAL SYMPOSIUM ON NMR IMAGING, BOWMAN GRAY SCHOOL OF MEDICINE, WINSTON-SALEM, N.C., OCTOBER 1–3, 1981. Edited by N. Karstaedt, R. L. Witcofski and C. L. Partain. Bowman Gray School of Medicine Press, 1982.

PRINCIPLES OF NMR IMAGING. I. L. Pykett, J. H. Newhouse, F. S. Buonanno, T. J. Brady, M. R. Goldman, J. P. Kistler and G. M. Pohost in *Radiology,* Vol. 143, pages 157–168; April, 1982.


## THE GENETICS OF ANTIBODY DIVERSITY


THE MOLECULAR BASIS OF ANTIBODY FORMATION: A PARADOX. W. J. Dreyer and J. Claude Bennett in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 54, No. 3, pages 864–868; September, 1965.

EVIDENCE FOR SOMATIC REARRANGEMENT OF IMMUNOGLOBULIN GENES CODING FOR VARIABLE AND CONSTANT REGIONS. Nobumichi Hozumi and Susumu Tonegawa in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 73, No. 10, pages 3628–3632; October, 1976.

SEQUENCES OF FIVE POTENTIAL RECOMBINATION SITES ENCODED CLOSE TO AN IMMUNOGLOBULIN $\kappa$ CONSTANT REGION GENE. Edward E. Max, J. G. Seidman and Philip Leder in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 76, No. 7, pages 3450–3454; July, 1979.

THE ORGANIZATION OF IMMUNOGLOBULIN GENES AND THE ORIGIN OF THEIR DIVERSITY. P. Leder, E. E. Max and J. G. Seidman in *Immunology 80,* edited by M. Fougereau and J. Dausset. Academic Press, Inc., 1980.


## THE ASYMMETRY OF FLOUNDERS


THE AMBIDEXTROUS UNIVERSE. Martin Gardner. Basic Books, 1964.

ANIMAL ASYMMETRY. A. C. Neville. Institute of Biology's Studies in Biology No. 67, Edward Arnold, 1976.

ON THE BIOLOGICAL BASIS OF HUMAN LATERALITY, I: EVIDENCE FOR A MATURATIONAL LEFT-RIGHT GRADIENT. II: THE MECHANISMS OF INHERITANCE. M. C. Corballis and M. J. Morgan in *The Behavioral and Brain Sciences,* Vol. 2, pages 261–336; 1978.

FLATFISHES AND THE INHERITANCE OF ASYMMETRIES. David Policansky in *The Behavioral and Brain Sciences,* in press.


## COSMIC JETS


THE STRUCTURE OF EXTENDED EXTRAGALACTIC RADIO SOURCES. George Miley in *Annual Reviews of Astronomy and Astrophysics,* Vol. 18, pages 165–218; 1980.

THE BIZARRE SPECTRUM OF SS 433. Bruce Margon in *Scientific American,* Vol. 243, No. 4, pages 54–65; October, 1980.

RELATIVISTIC JET PRODUCTION AND PROPAGATION IN ACTIVE GALAXIES. M. J. Rees, M. C. Begelman and R. D. Blandford in *Annals of the New York Academy of Sciences,* in press.


## THE GREGORIAN CALENDAR


THE CALENDAR: ITS HISTORY, STRUCTURE AND IMPROVEMENT. Alexander Philip. Cambridge University Press, 1921.

PAPI ASTRONOMI EPATTE: LUIGI GIGLIO, ASTRONOMO "PRIMES AUCTOR" DELLA RIFORMA GREGORIANA DEL CALENDRIO. Giulio Aromolo. Istituto della Stampa, Naples, 1963.

FURTHER ADJUSTMENT OF THE GREGORIAN CALENDAR YEAR. E. R. Hope in *The Journal of the Royal Astronomical Society of Canada,* Part I, Vol. 58, No. 1, pages 3–9, February, 1964; Part II, Vol. 58, No. 2, pages 79–87, April, 1964.

DIE GREGORIANISCHE KALENDERREFORM IM URTEIL ZEITGENÖSSISCHER ASTRONOMEN. Peter Aufgebauer in *Die Sterne,* Vol. 45, Nos. 4/5, pages 118–121; 1969.


## THE REGENERATION OF POTATO PLANTS FROM LEAF-CELL PROTOPLASTS

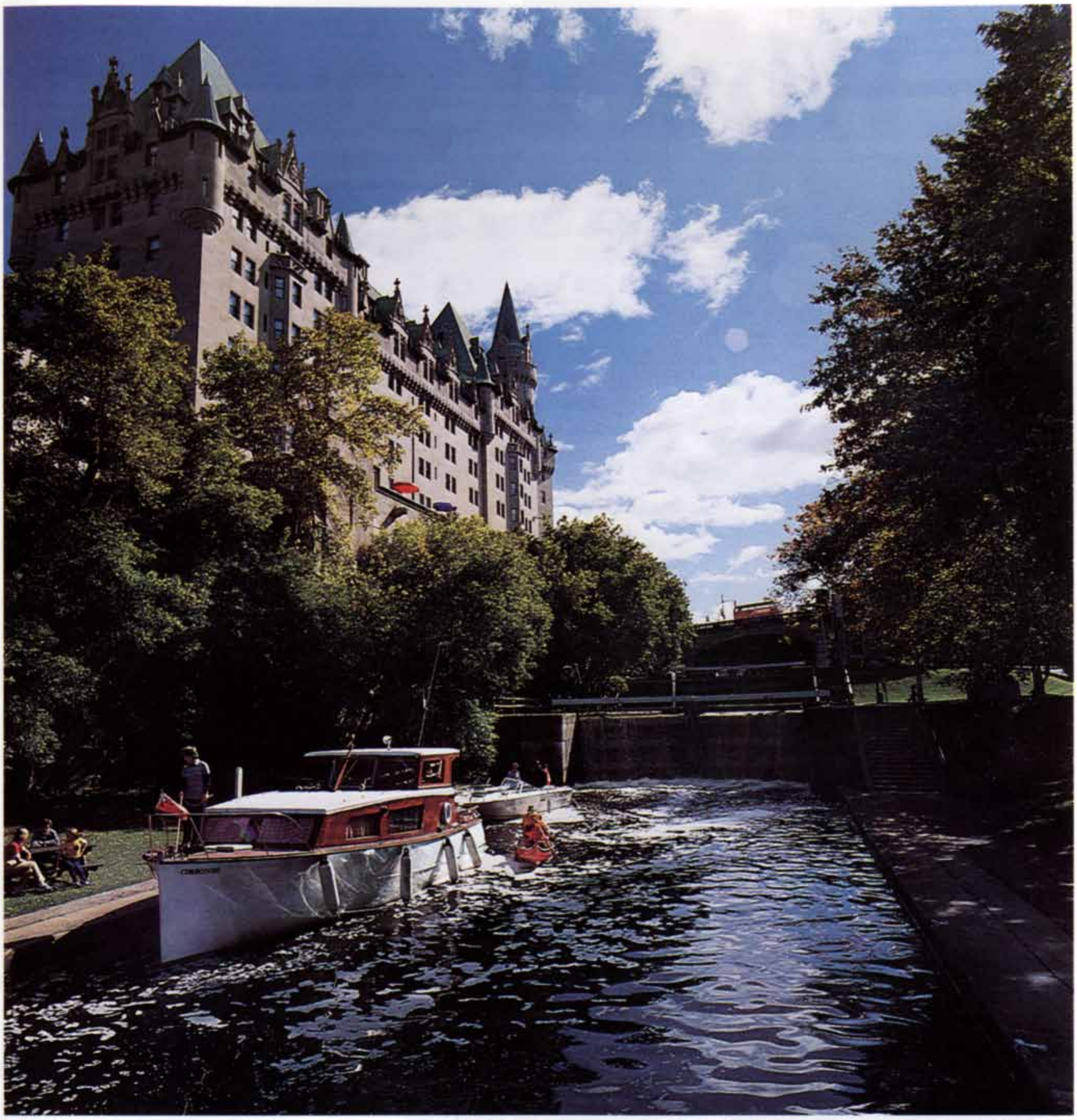

POTATO PROTOPLASTS IN CROP IMPROVEMENT. James F. Shepard, Dennis Bidney and Elias Shahin in *Science,* Vol. 208, No. 4439, pages 17–24; April 4, 1980.

PROTOPLASTS AS SOURCES OF DISEASE RESISTANCE IN PLANTS. James F. Shepard in *Annual Review of Phytopathology,* Vol. 19, pages 145–166; 1981.


## THE AMATEUR SCIENTIST


TONE QUALITY. John Backus in *The Acoustical Foundations of Music.* W. W. Norton & Company, Inc., 1969.

THE PERCEPTION OF TWO PURE TONES. John S. Rigden in *Physics and the Sound of Music.* John Wiley & Sons, Inc., 1977.



Ottawa, Ontario/Canada

# Room with a view

It's called a château and looks like a castle, but it's actually a hotel and please be our guest. Canada's capital abounds with far-away feelings from double-decker bus tours and sedate canal cruises to Changing the Guard on Parliament Hill. The splendours of our heritage are yours to discover in Ottawa, Ontario/Canada. Treat yourself royally! Call TOLL FREE 1-800-828-8585 or from New York State 1-800-462-8404 or write: Ontario Travel, Queen's Park, Toronto  M7A 2E5, Ontario/Canada.

ONTARIO
yours to discover!

# What if you chose
# as a technical

## "At AMP, Inc., an HP 1000 computer cuts product test costs by 96 percent."

AMP's Communication Products Division, in Winston-Salem, N.C. manufactures electrical and electronic connectors. Last year, thousands of products passed through an extensive battery of environmental tests, all controlled and processed by an HP 1000 computer system.

Group Systems Manager Anthony L. Chitea says, "With the HP system, we traded manual probing and a costly time-share program for automatic test, real-time data, and a 30-fold increase in productivity. Testing procedures which previously took three hours to complete are performed on the HP 1000 in less than 15 minutes. Based on this savings alone, our system easily paid for itself in less than a year.

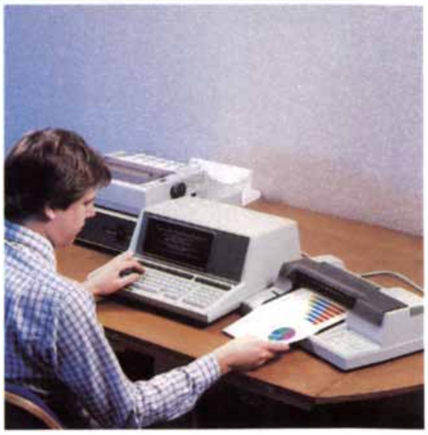
"By enhancing productivity, we're able to increase testing volume 60 times over, while reducing costs per test by 96 percent."

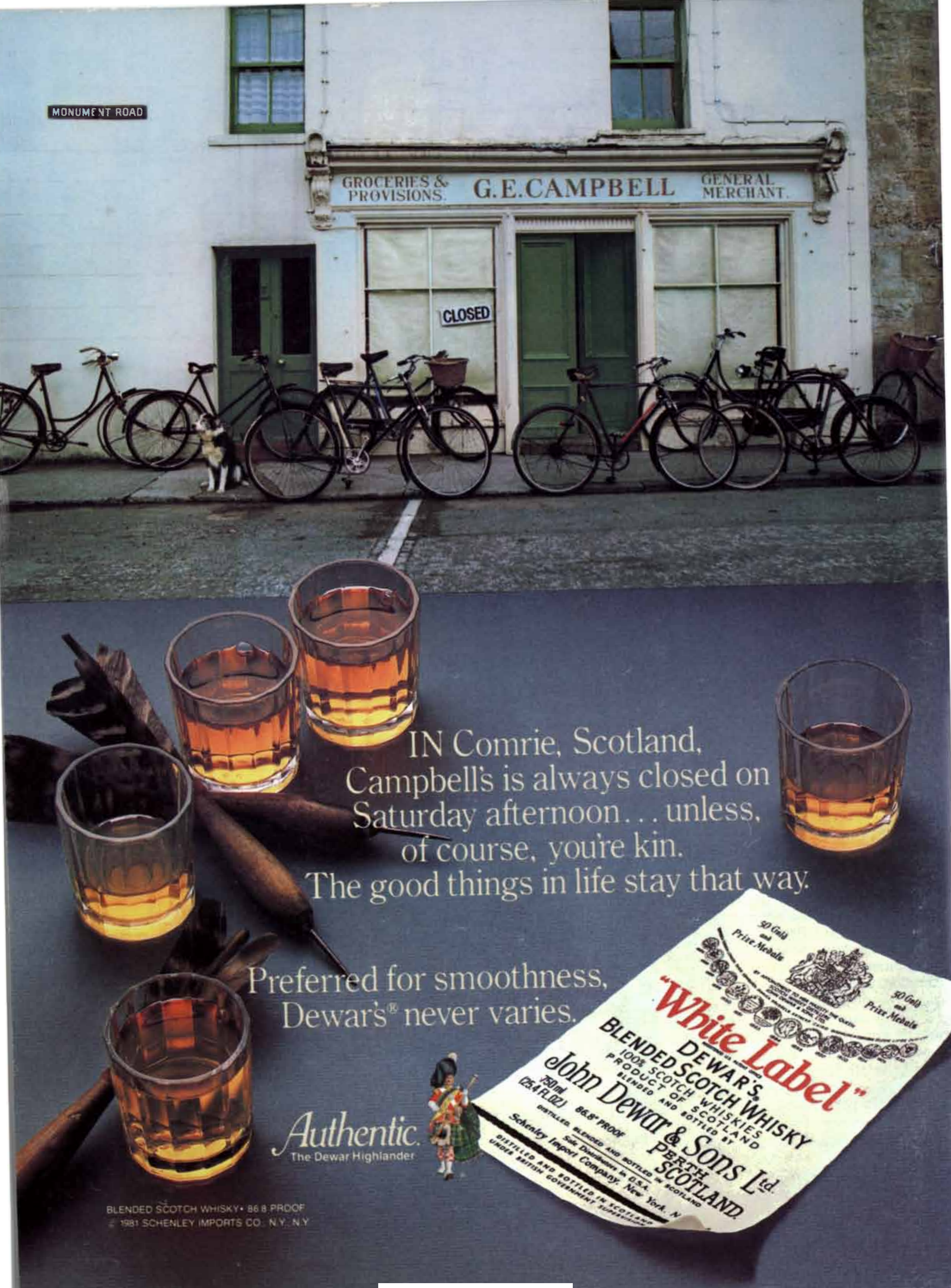
## HP can be your personal computer partner, too!

A growing family of personal computers is available through HP's network of qualified dealers, or through HP's computer sales organization.

For example:

**The new HP-87,** with enhanced BASIC, up to 544K bytes of user memory, and dynamic graphics capabilities, can handle large computing tasks and present the data in easily understood charts and graphs. Extensive software offerings are available from HP, as well as a CP/M® system module which opens the HP-87 to an extensive repertoire of software solutions. A wide range of peripherals gives your personal computing needs unlimited expression: a new color plotter; printers; and versatile mass storage including a new Winchester hard disc drive and flexible disc drives. Altogether, the HP-87 is a new, high-performance, highly affordable personal computer system.




MONUMENT ROAD
GROCERIES & PROVISIONS
G.E.CAMPBELL
GENERAL MERCHANT
CLOSED
IN Comrie, Scotland, Campbell's is always closed on Saturday afternoon. . . unless, of course, you're kin. The good things in life stay that way.
Preferred for smoothness, Dewar's® never varies.
Authentic.
The Dewar Highlander
"White Label"
DEWAR'S
BLENDED SCOTCH WHISKY
100% SCOTCH WHISKIES
PRODUCT OF SCOTLAND
BLENDED AND BOTTLED BY
John Dewar & Sons Ltd.
PERTH SCOTLAND
750ml (25.4 FL.OZ.)
86.8° PROOF
Schenley Import Company, New York, N.
BLENDED SCOTCH WHISKY • 86.8 PROOF
© 1981 SCHENLEY IMPORTS CO. N.Y., N.Y.