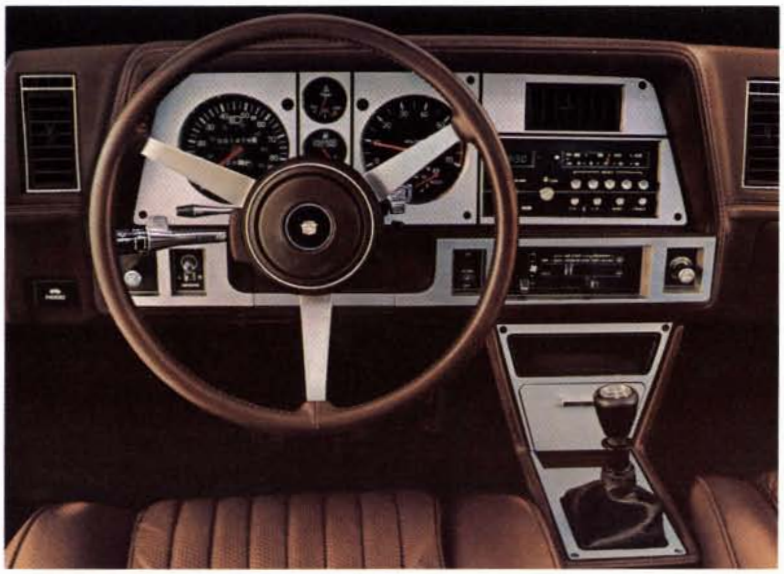
SCIENTIFIC AMERICAN

CHRISTOPHER WREN

$2.00

July 1981



ANNOUNCING BY CADILLAC

Cimarron

Quick-handling. Road-hugging. And fun to drive. This . . . is Cimarron. An efficient new kind of Cadillac. With the traction of front-wheel drive . . . MacPherson strut front suspension . . . and power-assisted rack and pinion steering with responsive 14:1 steering gear ratio. Plus, it has Cadillac refinements such as genuine leather seating areas, body-contoured bucket seats, air conditioning and more. All standard. Test-drive Cimarron by Cadillac. Due to limited initial production, Cimarron is not available at all Cadillac dealers at this time.

42 HWY. EST. 26* EPA EST. MPG

*Use estimated mpg for comparison. Your mileage may differ depending on speed, distance, weather. Actual highway mileage lower. Cadillacs are equipped with GM-built engines produced by various divisions. See your Cadillac dealer for details.

A NEW KIND OF CADILLAC FOR A NEW KIND OF CADILLAC OWNER.

GM

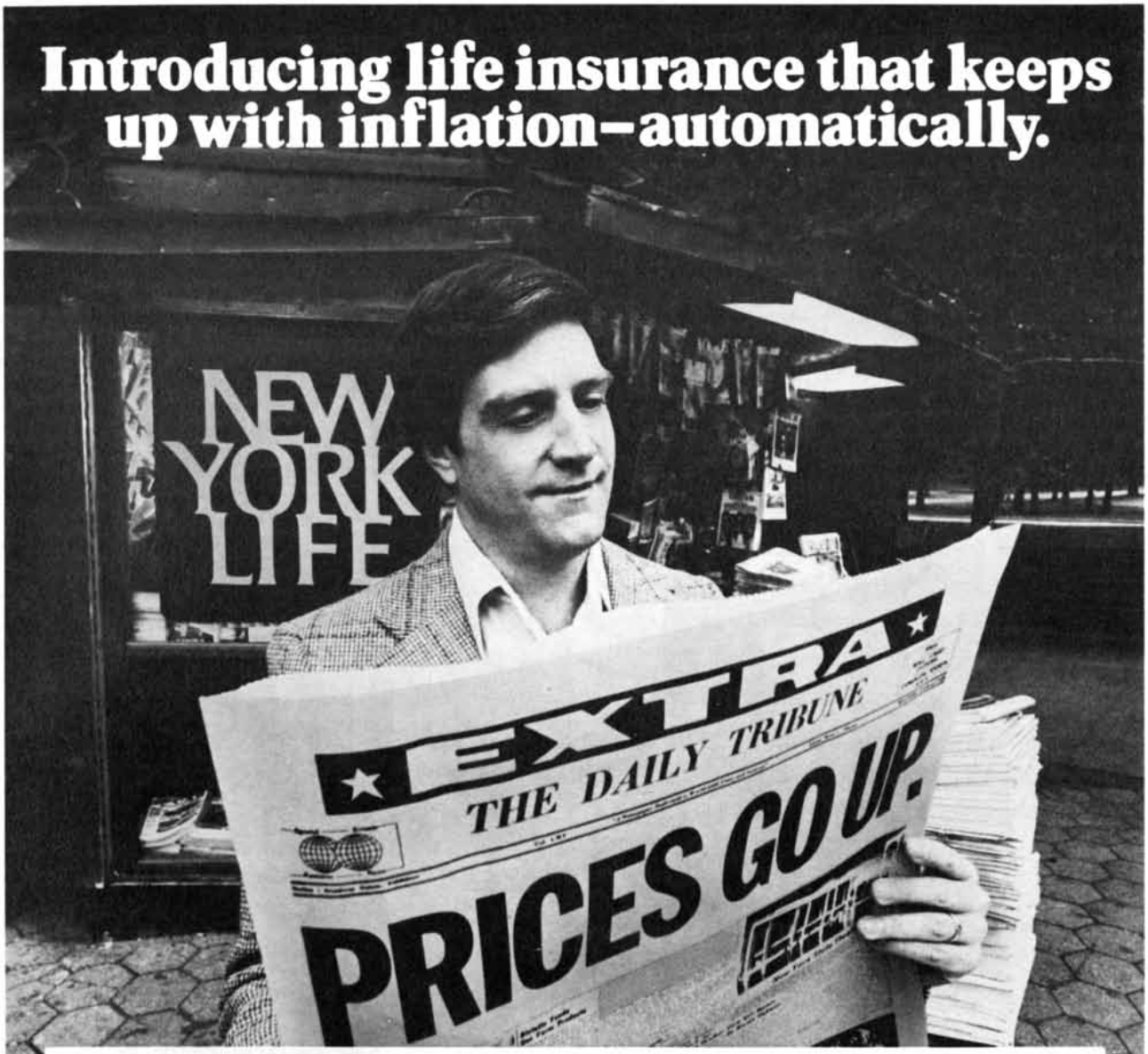
THE MACNEIL-LEHRER REPORT. UNERRINGLY ON TARGET.

Weeknights on PBS
with Charlayne Hunter-Gault

EXXON

Coproduced by WNET/New York and WETA/Washington, D.C., and made possible by grants from the Corporation for Public Broadcasting, Exxon Corporation, AT&T and the Bell System Companies and member stations of PBS.

Introducing life insurance that keeps up with inflation–automatically.



When the cost of living goes up, your life insurance coverage should go up, too. Otherwise your family will no longer have all the protection it needs.

Too much trouble, you say? And too costly? Well, as a New York Life Agent, I've got good news for you. New York Life now offers a brand new option that will take care of increasing your coverage—automatically.

You don't have to ask for the added protection each year. You don't have to take a physical exam or fill out a form. All you have to do is sign up for this option when you buy a popular New York Life cash-value policy.

Then, we'll automatically increase the face value of your policy any year the Consumer Price Index rises, as long as the increase in your policy amounts to at least $500.

The premiums you pay for this increase are less than those for a new policy. So you not only keep up with inflation, but get a little relief from it, too.

Ask me. Your New York Life Agent.

Ask me NEW YORK LIFE

New York Life Insurance Company, 51 Madison Ave., New York, N.Y. 10010. In Canada: 1240-444 St. Mary Avenue, Winnipeg, Manitoba R3C3T1. Life, Group and Health Insurance, Annuities, Pension Plans.

Established 1845 **SCIENTIFIC AMERICAN** July 1981 Volume 245 Number 1

ARTICLES

45 **VALUES AND ATTITUDES OF THE POLISH PEOPLE, by Stefan Nowak**
Recent events in Poland stem from a demand for social institutions in accord with people's values.

54 **THE SALINITY OF RIVERS, by Arthur F. Pillsbury**
Irrigation tends to deposit salts in the soil. How can they be made to continue on to the sea?

66 **THE ATMOSPHERE OF VENUS, by Gerald Schubert and Curt Covey**
Its clouds of sulfuric acid are driven by winds that attain a speed of 360 kilometers per hour.

106 **GENETIC ENGINEERING IN MAMMALIAN CELLS, by W. French Anderson and Elaine G. Diacumakos** One potential of recombinant DNA is curing human genetic disease.

122 **FIBER BUNDLES AND QUANTUM THEORY, by Herbert J. Bernstein and Anthony V. Phillips** Topological ideas illuminate the interactions of elementary particles.

138 **MIMICRY IN THE SEXUAL SIGNALS OF FIREFLIES, by James E. Lloyd**
Some of the females responding to male flashes are members of other species looking for a meal.

148 **BINOCULAR DEPTH INVERSION, by John I. Yellott, Jr.**
When the mask of a face is seen from the back instead of the front, its relief reverses. Why?

160 **THE ARCHITECTURE OF CHRISTOPHER WREN, by Harold Dorn and Robert Mark**
Wren was a man of science, but it seems he did not exploit theoretical mechanics in his buildings.

DEPARTMENTS

8 LETTERS
13 50 AND 100 YEARS AGO
14 THE AUTHORS
18 METAMAGICAL THEMAS
32 BOOKS
76 SCIENCE AND THE CITIZEN
176 THE AMATEUR SCIENTIST
188 BIBLIOGRAPHY

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, U.S., ITS POSSESSIONS AND CANADA; $27 PER YEAR, ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

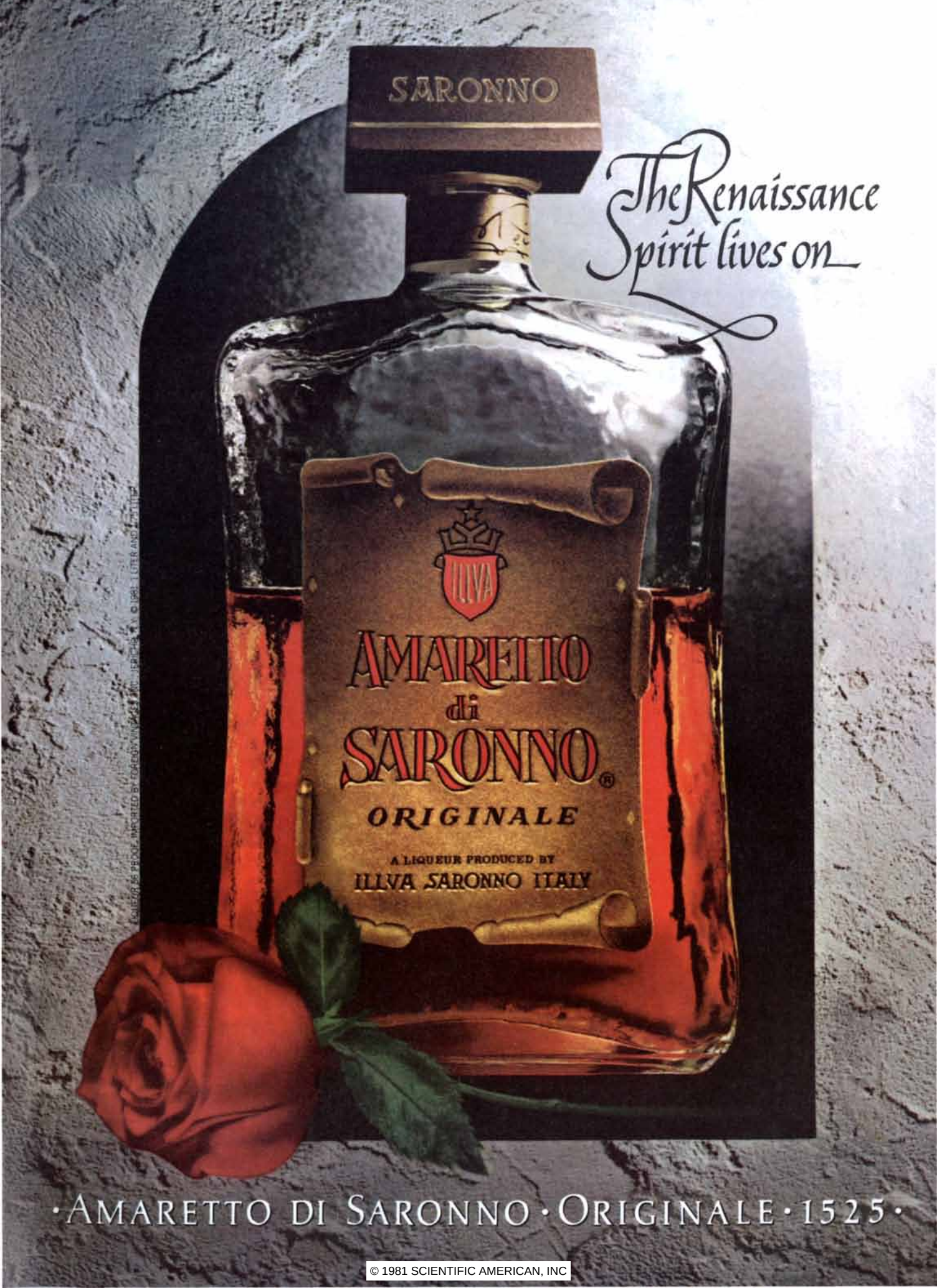
THE COVER

The painting on the cover shows an exterior detail of the Sheldonian Theatre at the University of Oxford, which was the first building designed by Christopher Wren. He based the design, which was notable for breaking with the Gothic tradition of Oxford, on the Roman theater of Marcellus. The Roman prototype had the shape of a horseshoe and was open at the top, covered only by a canvas awning. Wren's design, done in 1662, was a polygonal approximation to a horseshoe and had a permanent roof. His contemporaries were impressed by the ceiling of the theater because it was unsupported by columns over a span of 70 feet. Whatever the beauty and the engineering quality of Wren's ceiling design, it owes nothing to theoretical analysis (see "The Architecture of Christopher Wren," by Harold Dorn and Robert Mark, page 160). At the peak of the Scientific Revolution in Europe, Wren was Savilian professor of astronomy at Oxford and president of the Royal Society. Nevertheless, there is no definite connection between his science and his architecture.

THE ILLUSTRATIONS

Cover painting by Marvin Mattelson

Page	Source
20–27	Ilil Arbel
46–53	Albert E. Miller
54	EROS Data Center, National Aeronautics and Space Administration
56	Andrew Tomko
58–59	Ilil Arbel
60–61	Charlie Siegel, Water and Power Resources Service, Department of the Interior
62–64	Andrew Tomko
67	Larry D. Travis and Anthony Del Genio, Goddard Institute for Space Studies
68–72	Ian Worpole
73	Larry D. Travis and Anthony Del Genio, Goddard Institute for Space Studies
74	Michael J. S. Belton, Kitt Peak National Observatory
107	Elaine G. Diacumakos, Rockefeller University
108–121	Bunji Tagawa
122	Ralph Morse
123–136	George V. Kelvin
138	James E. Lloyd, University of Florida
140–145	Albert E. Miller
149	Ralph Morse
150–151	Walken Graphics
152	Walken Graphics (*top*); Bela Julesz, Bell Laboratories (*bottom left*); Ilil Arbel (*bottom right*)
157	John I. Yellott, Jr., University of California at Irvine
158–159	Walken Graphics
161	Thomas Photos, Oxford
162	John Moss, Black Star
163	Alan D. Iselin (*top and middle*); Allen Hess (*bottom*)
164	Ralph Morse
165	Valerio Simini
166	Ralph Morse
167	John Moss, Black Star
168	Robert Vuyosevich and Alan D. Iselin (*top*), Robert Mark (*bottom*)
170	Valerio Simini and Alan D. Iselin
172	John Moss, Black Star
173	Alan D. Iselin
177–180	David G. Stork
182–187	Michael Goodman

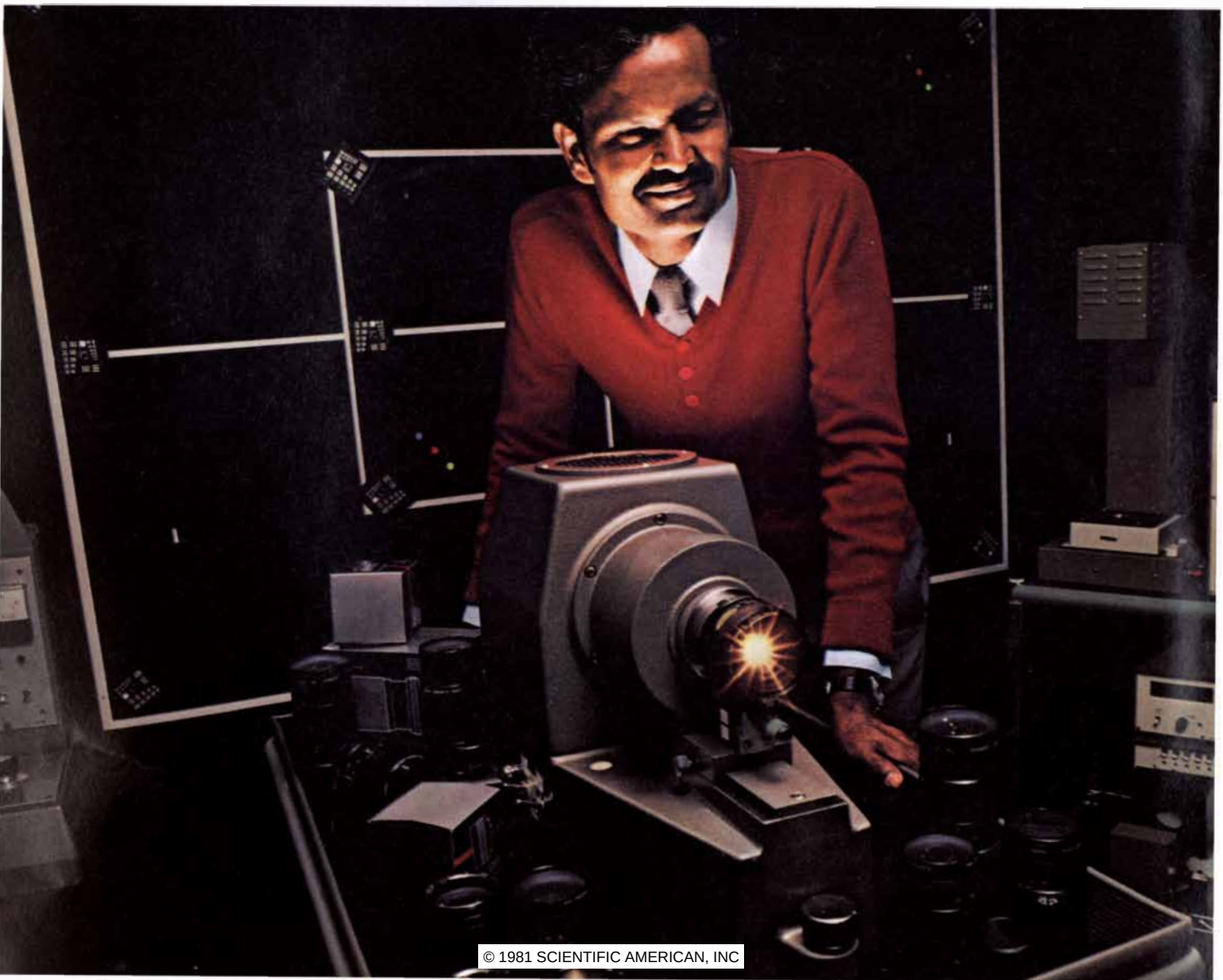


Reddy Chirra improves his vision with an Apple.

Reddy is an optical engineer who's used to working for big companies and using big mainframes.

But when he started his own consulting business, he soon learned how costly mainframe time can be. So he bought himself a 48K Apple II Personal Computer.

And, like thousands of other engineers



and scientists, quickly learned the pleasures of cutting down on shared time and having his own tamper-proof data base.

His Apple can handle formulas with up to 80 variables and test parameters on 250 different optical glasses.

He can even use BASIC, FORTRAN, Pascal and Assembly languages.

And Apple's HI-RES graphics come in handy for design.

Reddy looked at other microcomputers, but chose Apple for its in-depth documentation, reliability and expandability.

You can get up to 64K RAM in an Apple II. Up to 128K RAM in our new Apple III. And there's a whole family of compatible peripherals, including an IEEE-488 bus for laboratory instrument control.

Visit your authorized Apple dealer to find out how far an Apple can go with scientific/technical applications.

It'll change the way you see things.

The personal computer. apple

For the dealer nearest you, call (800) 538-9696. In California, call (800) 662-9238. Or write: Apple Computer Inc., 10260 Bandley Dr., Cupertino, CA 95014.

LETTERS

Sirs:

Eli Ginzberg and George J. Vojta ["The Service Sector of the U.S. Economy," SCIENTIFIC AMERICAN, March], contend "that human capital, defined as the 'skill, dexterity and knowledge' of the population, has become the critical input that determines the rate of growth of the economy and the well-being of the population." But they measure the growing expenditures for education, research and development and health care, not the resulting rate of growth of the economy. By their measure the U.S. has more "skill, dexterity and knowledge" than Japan, whereas judging by the results it is Japan (with lower expenditures) that has the more of it....

GABRIEL ZAID

Mexico City

Sirs:

...Ginzberg and Vojta make two points of fundamental importance: that we must continue to support our educational and research facilities and that we must not allocate precious resources to pump up failing industries. But they make no mention of the extent to which the comparative advantage of an industry (even those perceived as "sunset" industries such as steel and automobiles) depends on the efficiencies of production afforded by capital goods that embody innovations of technology. They do not discuss the degree to which the service sector is employed in the development of technology and in the manipulation of technology.

We have learned that technology can be purposefully stimulated by government policy. The Japanese support of the microelectronics industry is a leading recent example. We know that technological innovations tend to become embodied in products and processes that increase efficiency in both the industrial and the service sectors. Reindustrialization, one would hope, would be directed toward greater efficiencies by providing industry with state-of-the-art production techniques and products. This demand should in turn be a stimulus for advancing the state of the art, resulting in a further increase in productivity, an increase in the comparative advantage of U.S. industry, the increased availability (to all sectors) of technology in both producer and consumer goods, and the increase of available goods and services that we characterize as an increase in G.N.P.

S. DONALD GONSON

Cambridge, Mass.

Sirs:

Mr. Zaid raises a good question: How does one distinguish between the investment and consumption aspects of education and other outlays for human capital? One cannot do so scientifically, only judgmentally. His enthusiasm for Japan would be tempered if he read the MITI report for the 1980's, in which the Japanese raise serious questions as to their ability to lead at the cutting edge of research and development.

Mr. Gonson's explanation for the troubles of the U.S. automobile and steel industries as residing primarily in the weakness of the capital goods they employ is not without merit but appears farfetched. Detroit misjudged the market, and steel sought to maintain a leadership role without adequate reinvestment. In the face of such errors why pick on the backwardness of the capital-goods sector?

With respect to government assistance in Japan for the microelectronics industry one should remember that even the Japanese government has some hits and some misses.

ELI GINZBERG

GEORGE J. VOJTA

Graduate School of Business
Columbia University
New York

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 60 cents.

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

Name

New Address

Old Address

Introducing Gold Reserve. By Bacardi of course.

Now there's something special for rum drinkers who want something more, something luxurious.

That something is Bacardi Gold Reserve rum, the premium sipping rum from Bacardi. It's so smooth, mellow and rich, you'll only want to sip it neat or on the rocks.

After all, doesn't a rum drinker deserve a taste of luxury as much as a Scotch or Canadian drinker?

BACARDI® Gold Reserve rum. Our premier spirit.

Consumer Orientation

No. 12 in a Series of Technical Papers

12

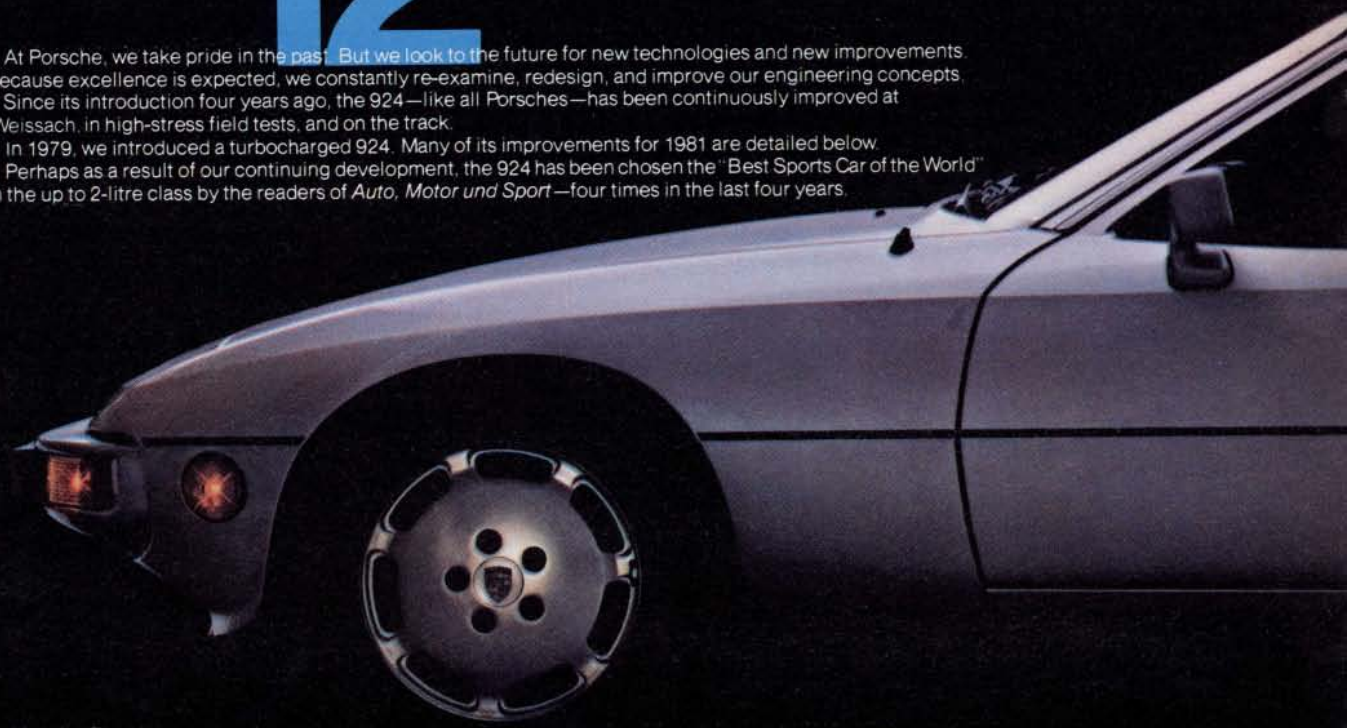
Subject:
Design, Improvement, and Production of the Porsche 924/924 Turbo. Model Years 1977 through 1981.

At Porsche, we take pride in the past. But we look to the future for new technologies and new improvements. Because excellence is expected, we constantly re-examine, redesign, and improve our engineering concepts.

Since its introduction four years ago, the 924—like all Porsches—has been continuously improved at Weissach, in high-stress field tests, and on the track.

In 1979, we introduced a turbocharged 924. Many of its improvements for 1981 are detailed below.

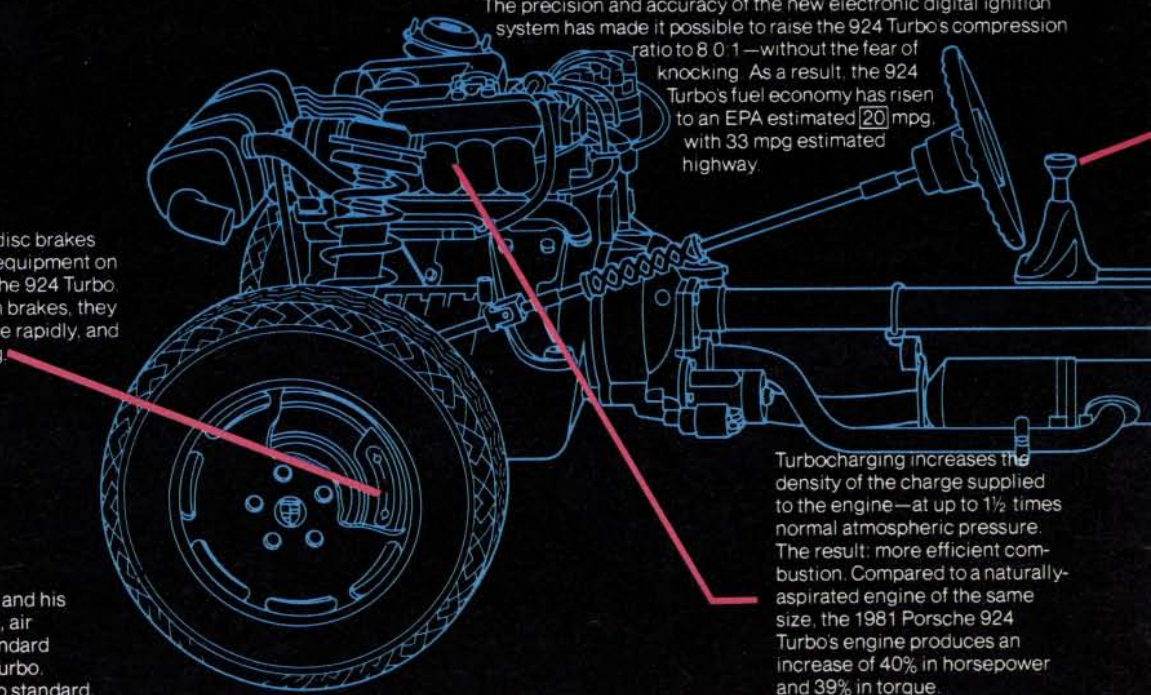
Perhaps as a result of our continuing development, the 924 has been chosen the "Best Sports Car of the World" in the up to 2-litre class by the readers of *Auto, Motor und Sport*—four times in the last four years.



New halogen headlamps provide wider and brighter light—an increase of approximately 30% over last year. In addition, the diffusion pattern of each lens provides greater visibility, extending the driver's field of vision.

The 1981 Porsche 924 Turbo introduces a new electronic digital ignition system with automatic idle stabilization. Now, regardless of engine load, engine rpm is maintained (±50 rpm). In addition, engine output has been increased from 143 to 154 hp because of greater ignition accuracy (every 1.8° of crank rotation is controlled) and more efficient combustion chamber design. The new electronic ignition is set permanently. No maintenance is required because there is no mechanical wear and tear.

The precision and accuracy of the new electronic digital ignition system has made it possible to raise the 924 Turbo's compression ratio to 8.0:1—without the fear of knocking. As a result, the 924 Turbo's fuel economy has risen to an EPA estimated 20 mpg, with 33 mpg estimated highway.



Internally-vented disc brakes are now standard equipment on all four wheels of the 924 Turbo. Compared to drum brakes, they dissipate heat more rapidly, and thus reduce fading.

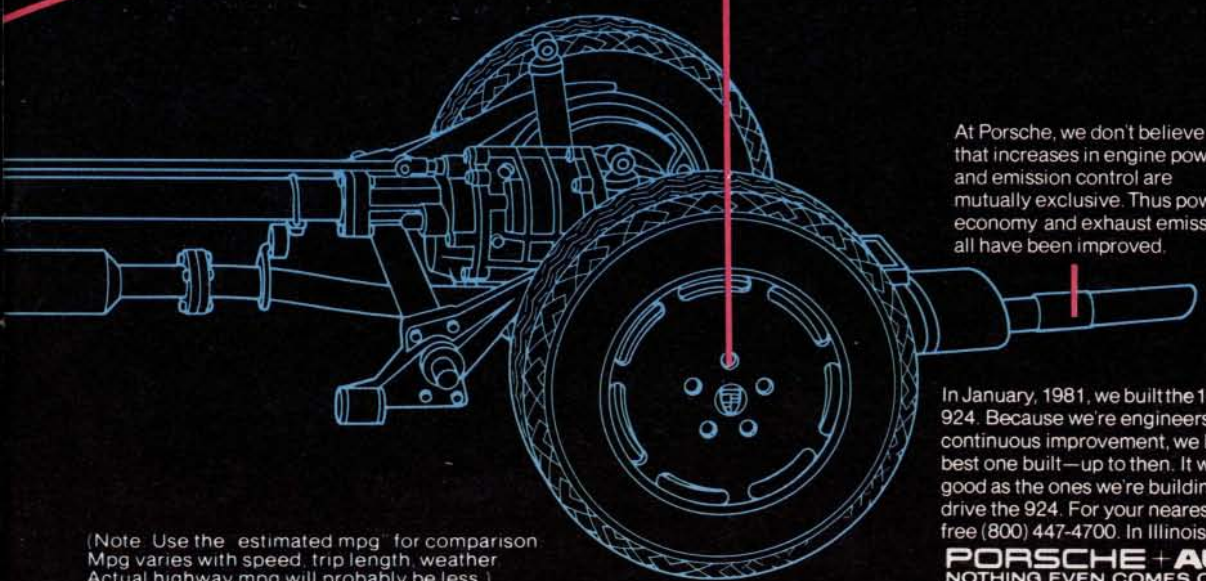
Turbocharging increases the density of the charge supplied to the engine—at up to 1½ times normal atmospheric pressure. The result: more efficient combustion. Compared to a naturally-aspirated engine of the same size, the 1981 Porsche 924 Turbo's engine produces an increase of 40% in horsepower and 39% in torque.

To help keep the driver and his passenger comfortable, air conditioning is now standard equipment in the 924 Turbo. Power windows are also standard.

Porsche 924/924 Turbo

For more convenience in stop-and-go city driving, first gear—which is seldom used on the track—has been moved into the H-pattern:

New five-bolt 6JX15 light alloy wheels maintain strength yet reduce unsprung weight. Fitted with high-speed 185/70VR15 radial ply tires, they improve both ride and tire traction.



At Porsche, we don't believe that increases in engine power and emission control are mutually exclusive. Thus power, economy and exhaust emissions all have been improved.

In January, 1981, we built the 100,000th Porsche 924. Because we're engineers, interested in continuous improvement, we know it was the best one built—up to then. It was almost as good as the ones we're building today. Test drive the 924. For your nearest dealer, call toll-free (800) 447-4700. In Illinois, (800) 322-4400.

PORSCHE + AUDI
NOTHING EVEN COMES CLOSE

(Note: Use the "estimated mpg" for comparison. Mpg varies with speed, trip length, weather. Actual highway mpg will probably be less.)

Follow America's roughest trail to the world's smoothest whisky, Canadian Club.

Find the case of C.C.® we hid along the Lewis and Clark trail.

We retraced Lewis and Clark's historic expedition up the Missouri River into Montana. And where they found their roughest going, we hid a case of the smoothest whisky, Canadian Club.

Where Lewis and Clark had floated the unspoiled river, Canadian Club's rafts followed. We tested our nerves, as the explorers had, on the wild rapids of the Clark's Fork.

As we explored those historic Montana streams, we buried a case of Canadian Club overlooking the very site of one of the expedition's most important sightings. One clue: neither Lewis nor Clark made it.

Discover the taste of the world's finest whisky.

As you search along this historic trail, remember Canadian Club has a proud history of its own. For over 120 years it's been enjoyed by those who seek the very best. Canadian Club is lighter than Scotch, smoother than bourbon, and enjoyable on the rocks, in a sour or Manhattan, or with your favorite mixer.

So come search along the Lewis and Clark trail to discover why it's "The Best In The House", or simply venture down to your favorite bar and say, "Canadian Club, please!"

Canadian Club®

"The Best In The House"® in 87 lands.

6 YEARS OLD. IMPORTED IN BOTTLE FROM CANADA BY HIRAM WALKER IMPORTERS INC., DETROIT MICH. 86.8 PROOF. BLENDED CANADIAN WHISKY © 1981

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JULY, 1931: "If, as appears to be true, the only permanent entity in the universe is light, then Albert A. Michelson, who has died in his 79th year, identified himself so intimately with it that the one can scarcely be thought of without the other. We may attempt to isolate Michelson's major achievements, summing them up broadly as the perfection of the Michelson interferometer for the measurement of the diameter of stars, and the high refinement of the determination of the velocity of light, not to speak of the famous Michelson-Morley experiment, whose unanticipated by-product was the Einstein theory of relativity. But when we have accomplished that much, Michelson's major contributions to science and the world have been little more than touched. Everyone who knows the inner world of physics and its personnel is aware that Michelson's real contribution to science was simply in being Michelson: Michelson the leader, Michelson the teacher, Michelson the inspirer—particularly the inspirer. To this a host of Millikans and Comptons who grew up under his tutelage will testify. He sensed which problems were the big ones and not only directed the energies of his best workers toward them but also kept them there."

"On the cliffs of St. Margaret's Bay near Dover, England, the International Telephone and Telegraph Laboratories of Hendon, England, in co-operation with the Laboratories of Le Matériel Téléphonique of Paris, recently gave a successful demonstration of a new ultra-short-wave radio telephone and telegraph circuit between Dover and Calais across the English Channel. In this demonstration oscillations of wavelengths as short as 10 centimeters, designated 'micro-rays,' were used for the first time to provide a high-quality two-way radio-telephone circuit. When compared with radiations of the more usual wavelengths, 'micro-rays' present many striking features. For example, their extremely short wavelength permits the use of electro-optical devices more usually associated with light, such as reflectors or refractors. A very important use will be for television transmission. The present difficulty with regard to television is the very large frequency range required for satisfactory definition of the object transmitted. It should now be possible to allocate as wide a band as is necessary for television without causing any ether congestion."

"The world is not as familiar with the name of Willem de Sitter, the Dutch astronomer-cosmologist, as it is with that of Albert Einstein. Readers of science often encounter the term 'the Einstein universe,' 'the de Sitter universe' and so on. Such expressions refer to concepts of the size, shape and general nature, also the finiteness or infiniteness, of the totality of things. Of course, no existing telescope can penetrate the entire extent of the universe, and therefore these concepts are based partly on inference. In de Sitter's universe, which is finite, space is curved or bent, not so much because of the presence of matter, as in Einstein's universe, but inherently. It is an unstable universe, expanding or contracting. Recent research by E. P. Hubble of the Mount Wilson Observatory, indicating an expansion, favors de Sitter's concept and has caused Einstein to revise his own."

"As business and industry begin gradually to adjust themselves to present conditions and to plan for the future, the controversy over the possibility of wage reductions is becoming more serious. President Hoover has consistently fought against such a step, although many executives, bankers in particular, have as strenuously favored it. It is going to be no easy matter to adjust wages to a lower level even if in time that is shown to be the only feasible thing to do. Nevertheless, others besides the workingman have felt the depression, and on a stupendous scale. If, therefore, the restoration of economic balance can be effected in no other way, it may finally be necessary to reduce wages, in certain lines at least."

JULY, 1881: "Judge Lowell of the United States Circuit Court in Boston rendered an important decision on the 27th ult., in which he virtually confirms to the American Bell Telephone Company the exclusive right of talking over a wire by electricity. If his decision stands, what a marvelous honor belongs to Alexander Graham Bell! He is declared to be the original and first discoverer of the far-reaching art of speech transmission by electricity. It has been heretofore supposed by electrical laymen that Bell's devices are simply improvements upon something previously done in the same line by others, such as Oersted, Reiss and Gray. But a trial has been had, and the hopes of hundreds of telephonic inventors have been laid low in the dust. It may equally be said of Bell's telephone that although it is a good receiver, it is a poor transmitter—so poor that its use has been almost abandoned in favor of superior instruments such as the Blake or the Edison. If we had to rely only on the Bell instruments, the telephone would be a nuisance, and the wide-spread use of speaking telegraphy now enjoyed could never have been realized."

"Sir William Thomson confirms the favorable results of his previous experiments with the Faure battery. He says: 'If the processes are pushed on too fast, there is necessarily a great loss of energy. If the processes are carried on too slowly, there is inevitably some loss through local action, the spongy lead becoming oxidized and the peroxide losing some of its oxygen viciously, that is to say, without doing the proper proportion of electric work in the circuit. I have seen enough, however, to make me feel very confident that in any mode of working the accumulator will not be uselessly slow and the loss from local action will be very small. I think it most probable that, at rates of working that would be perfectly convenient for the ordinary use of fixed accumulators in connection with electric lighting and the electric transmission of power for driving machinery, the loss of energy in charging the accumulator and taking out the charge again will be less than 10 per cent of the whole that is spent in charging the accumulator. It seems probable that a tramcar arranged to take in, say, 850 pounds of freshly charged accumulators, on leaving headquarters for an hour's run, may be driven more economically by the electric energy operating through a dynamo-electric machine than by horses.'"

"Dr. W. F. Gintl discusses the question of whether, according to the experiments of Crookes, the assumption of an especial fourth state of matter is necessary, or whether the facts may be satisfactorily explained without such hypothesis. He shows that the latter alternative is possible with the aid of a mechanical theory of electricity: Particles become separated from the surface of the substance of the negative pole, they are repelled and they move away from the pole with a speed resulting from the antagonistic forces in a parallel and rectilinear direction, preserving their speed and their initial path so long as they do not meet with obstacles that influence their movement. At a certain density of the gases present in the exhausted space these particles, in consequence of the impact of gaseous molecules more or less opposed to their direction of movement, lose their velocity after traveling a short distance and soon come to rest. The luminous phenomena of the Geissler tube Dr. Gintl supposes to be produced by the intense blows the gaseous molecules receive from the polar particles flying through the apparatus."

THE AUTHORS

STEFAN NOWAK ("Values and Attitudes of the Polish People") is a sociologist on the faculty of the University of Warsaw. He is currently on leave as a visiting professor at Columbia University. A native of Warsaw, Nowak was educated at the university there, earning his M.A. in the sociology of literature in 1951. He taught a course in the history of Renaissance philosophy at the university for several years before returning to the study of sociology, obtaining a Ph.D. in the methodology and philosophy of the social sciences in 1958. Nowak has made a number of visits to academic institutions outside Poland, beginning with a year as a postdoctoral fellow at Columbia's Bureau of Applied Social Research in 1958–59. On other occasions he has been a visiting professor at universities in the U.S., Sweden, Finland and Norway. In recent years two of his books have been translated into English: *Essays in the Methodology of Social and Behavioral Theories* (1976) and *Methodology of Sociological Research* (1977). Nowak is currently president of the Polish Sociological Association and a member of the executive committee of the International Sociological Association.

ARTHUR F. PILLSBURY ("The Salinity of Rivers") is a retired engineer who was affiliated for most of his career with the University of California. He has two degrees from Stanford University: a B.A. in 1928 and a C.E. in 1930. He was employed as an irrigation engineer at both the Berkeley and the Riverside campus of the university system until 1939, when he joined the U.C.L.A. faculty, serving for many years as chairman of the department of irrigation and soil science. At the time of his retirement in 1972 Pillsbury was professor of engineering and director of the Water Resources Center at U.C.L.A.

GERALD SCHUBERT and CURT COVEY ("The Atmosphere of Venus") are in the department of geophysics and planetary physics of the University of California at Los Angeles. Schubert, who has the rank of professor, was graduated from Cornell University in 1961 with degrees in engineering physics and aeronautical engineering. He then joined the Navy, serving for four years on the faculty of the Naval Nuclear Power School at Mare Island in California, while attending the University of California at Berkeley, where he obtained a Ph.D. in aeronautical sciences in 1964. After a postdoctoral year in the department of applied mathematics and theoretical physics at the University of Cambridge he returned to the U.S. to join the U.C.L.A. faculty in 1966. Covey, a graduate student at U.C.L.A., majored in biology as an undergraduate at the Massachusetts Institute of Technology. He already has two advanced degrees, both from the University of California at Santa Barbara: an M.A. in biology and an M.A. in physics.

W. FRENCH ANDERSON and ELAINE G. DIACUMAKOS ("Genetic Engineering in Mammalian Cells") are cell biologists with a common interest in the new technology of gene splicing. Anderson is chief of the Laboratory of Molecular Hematology at the National Heart, Lung, and Blood Institute. He is also adjunct professor of genetics at George Washington University. A graduate of Harvard College, he has an M.A. from the University of Cambridge and an M.D. from the Harvard Medical School. Anderson has been associated with the National Institutes of Health since 1965, when as a newly commissioned officer of the U.S. Public Health Service he was assigned to the National Heart Institute's Laboratory of Biochemical Genetics. Diacumakos is head of the laboratory of cytobiology at Rockefeller University. She did her undergraduate work at the University of Maryland at College Park, obtaining a B.S. in 1951, and her graduate work at New York University, where she got an M.S. in 1955 and a Ph.D. in 1958. She was associated with the Sloan-Kettering Institute for Cancer Research in 1962 when she was awarded a special fellowship of the Public Health Service to work in the laboratory of Edward L. Tatum at Rockefeller. Diacumakos has been head of the cytobiology laboratory there since 1976.

HERBERT J. BERNSTEIN and ANTHONY V. PHILLIPS ("Fiber Bundles and Quantum Theory") are respectively a theoretical physicist and a mathematician. Bernstein, who teaches at Hampshire College, studied at Columbia University (B.S., 1963) and the University of California at San Diego (Ph.D., 1967). Before joining the faculty at Hampshire in 1971 he was a visiting member of the Institute for Advanced Study, an instructor at Southeastern Massachusetts University, an investigator at the Cambridge Electron Accelerator and a visiting assistant professor at the Institute for Theoretical Physics of the University of Louvain in Belgium. In 1977–78 Bernstein served as technical director of Volunteers in Technical Assistance (VITA) in Washington, D.C. Phillips, who is professor of mathematics at the State University of New York at Stony Brook, was graduated from the Massachusetts Institute of Technology in 1960 and went on to earn his Ph.D.

HOW TO GET A BANKER TO PLAY BALL WITHOUT BATTING AN EYE:

It's a lot easier than you might think.

All you have to do is team up with the professionals at A Full Service Bank,® and be sure you are taking advantage of the many services they have to offer.

To really get a good customer relationship going, reassure them that they will handle your checking business, your savings business, loan business, direct deposits, whatever you need in the way of financial services.

They'll be pleased. Impressed. And they'll work hard to accommodate you in every way they can.

That's how you get to first base with your banker, because...

GOOD CUSTOMERS COUNT AT A FULL SERVICE BANK.

A FULL SERVICE BANK®

from Princeton University in 1966. He has worked at a variety of institutions, including the University of California at Berkeley, the Steklov Institute of Mathematics in Moscow, the Institute for Advanced Study, the Pontifical Catholic University in Rio de Janeiro and the University of Paris. Phillips' article "Turning a Surface Inside Out" appeared in the May, 1966, issue of SCIENTIFIC AMERICAN.

JAMES E. LLOYD ("Mimicry in the Sexual Signals of Fireflies") is professor of entomology and nematology at the University of Florida. A graduate of the State University of New York at Fredonia, he went on to obtain an M.A. in biology at the University of Michigan in 1962 and a Ph.D. in entomology from Cornell University in 1966. Lloyd's field work on fireflies is done largely in the tropical rain forests of Southeast Asia, New Guinea and Central America.

JOHN I. YELLOTT, JR. ("Binocular Depth Inversion"), is professor of psychology at the University of California at Irvine. All his degrees are from Stanford University; the last, a Ph.D., was awarded in 1966. He taught for several years at the University of Minnesota, leaving in 1970 to spend a year as a visiting member of the Institute for Advanced Study. He then moved to the school of social sciences at Irvine, where he has been ever since. Yellott's research work has focused primarily on applications of mathematics to problems in learning and perception.

HAROLD DORN and ROBERT MARK ("The Architecture of Christopher Wren") have collaborated on several projects since they first met in the late 1960's at Princeton University, where Dorn was a graduate student in the history of science and Mark taught civil engineering. Dorn, who is now professor of the history of science and head of the humanities program at the Stevens Institute of Technology, has a B.A. in philosophy and a B.C.E. in civil engineering, both from the College of the City of New York, an M.S. in mathematics from the Courant Institute of New York University and an M.A. and a Ph.D. in history from Princeton. Mark, who in recent years has become increasingly interested in the technology of historic buildings, now holds a joint appointment as professor of both civil engineering and architecture at Princeton. He majored in engineering as an undergraduate at C.C.N.Y. and worked for several years as an engineer before moving to Princeton in 1957. Mark was the author of an earlier article in SCIENTIFIC AMERICAN: "The Structural Analysis of Gothic Cathedrals" (November, 1972). His book, *Experiments in Gothic Structure,* is scheduled to be published by the MIT Press later this year.

An invitation to serious scientific readers...

Take any set for only $2.95

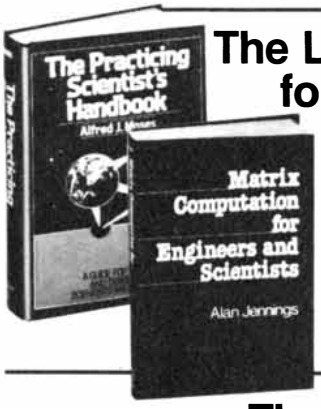
You simply agree to buy 3 more books—at handsome discounts—within the next 12 months.

Van Nostrand's Scientific Encyclopedia.

A 2,370-page classic, enlarged and updated. Hailed by the New York Times as "an amazing book for both the general and scientific reader." 7,200 articles and 2,500 photographs, drawings and charts.

A $69.50 value.

The Library for Practicing Scientists...



Alfred J. Moses' **The Practicing Scientist's Handbook**. 1,290 pages. A vast compendium of tabular data on materials properties. Publisher's Price $54.50.

Alan Jennings' **Matrix Computation for Engineers and Scientists**. An essential guide explaining the most common matrix techniques and how to apply them. Publisher's Price $35.75.

Two volumes, a $90.25 value.

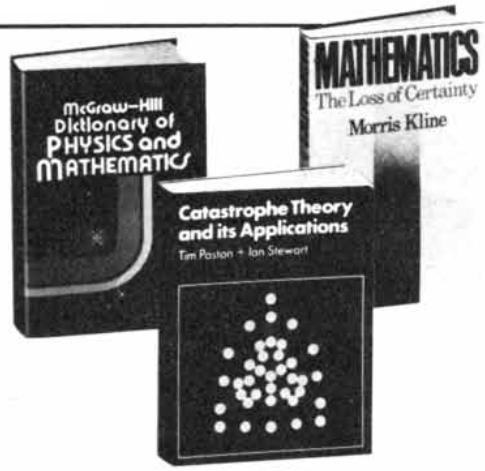
The Mathematics Library...

The McGraw-Hill Dictionary of Physics and Mathematics. The basic physics and mathematics vocabulary, plus the latest, most specialized terminology. Over 1,000 pages. More than 700 drawings, graphs and tables. Publisher's Price $22.50.

Tim Poston and Ian Stewart's **Catastrophe Theory and its Applications**. This brilliant exposition explains the theory in "unstuffy language" *(Physics Bulletin).* "Delightful, beautifully illustrated . . . informative and up-to-date" *(New Scientist).* Publisher's Price $24.95.

Morris Kline's **Mathematics: The Loss of Certainty**. Engrossing, compelling exploration of today's world of mathematics. Publisher's Price $19.95.

Three volumes, a $67.40 value.



The Astronomy Library...

Simon Mitton, editor, **The Cambridge Encyclopedia of Astronomy.** A one-volume comprehensive survey of the universe. Over 500 illustrations. Publisher's Price $40.00.

Paolo Maffei's **Monsters in the Sky.** Comets, quasars, star forges and much more in this investigation of deep sky mysteries. Filled with photos, charts, maps, drawings. Publisher's Price $15.00.

Michael Rowan-Robinson's **Cosmic Landscape.** Six voyages along the paths of light photons reveal worlds invisible to the human eye. Publisher's Price $12.95.

Three volumes, a $67.95 value.

The Library of Science is the one book club expressly formed for dedicated scientists and sophisticated amateurs who have a keen interest in the entire world of scientific discovery.

Join The Library of Science for a trial period and we'll ship your choice of Van Nostrand's Scientific Encyclopedia or any of the other sets shown for just $2.95.

Reply card removed? Please write **The Library of Science,** Department 2-AH8, Riverside, NJ 08075 for membership information and application.

METAMAGICAL THEMAS

Pitfalls of the uncertainty principle and paradoxes of quantum mechanics

by Douglas R. Hofstadter

You have dropped a coin into a crevice in the upholstery of a fancy old chair. You want to retrieve the coin, and so you push your hand gingerly into the crevice and try to grab it. But the very act of pushing your hand in widens the crevice, and the coin slips farther down. You can feel that if you push any more, the coin will be lost forever in the innards of the chair. What to do? This commonplace little drama illustrates a feeling everyone knows: that striving for something can have the effect of reducing that thing's availability.

A good friend is visiting you from far away, and before she returns home you want to capture her enchanting laugh on film. She, however, is terribly camera-shy. The instant you bring out the camera she freezes; spontaneity is lost, and so is any chance of recording that laugh. The act of trying to capture the elusive phenomenon completely destroys it.

Examples such as these are sometimes attributed to the notorious uncertainty principle. That principle of quantum mechanics was first enunciated by Werner Heisenberg in about 1927. Careless paraphrases since then, however, have eroded and obscured the true meaning of the principle in the popular mind. Here I should like to clarify matters by discussing the genuine uncertainty principle and its impostors.

Let me first state an imitation version clearly, so that you know what I am attacking.

The pseudo uncertainty principle states: The observer always interferes with the phenomenon under observation. It tends to be cited heavily in particular domains, often where the phenomenon involves a reciprocal observer—someone who can observe back! But even in such cases this pseudo principle is too simplistic. It rests on a misunderstanding of how experimentation proceeds and how science explains. The main thing to keep in mind is that science is about classes of events, not particular instances. Science explains through abstractions that underlie a potentially unlimited number of concrete phenomena.

Consider the following example. I recently heard about a woman who, in speaking of a friend of hers, said: "At that time she always date shranks." She meant that her friend used to go out only with psychiatrists. The tense marker somehow got shifted from the verb "date" to the noun "shrink," which was then conjugated as if it were functioning as a verb: "dated shrinks" became "date shranks." It would be fascinating to know exactly what was going on in the woman's brain as she made this bizarre transformation. We should like to know exactly where things went awry—down the wrong track. What went where, and why?

But this was a one-shot phenomenon; it will probably never be repeated. A scientific explanation of the details is unlikely to be forthcoming. Instead we have to abstract some general phenomenon we think is the essential component of this particular event. We have to be able to imagine other events in the same general class. We have to be able to imagine some way of provoking them or of detecting them when they happen, so that we can study the patterns. Perhaps the appropriate level of abstraction is "grammatical errors in the speech of woman *W*." Or perhaps it is "shifts of tense markers from verbs to nouns." Depending on how we abstract the event we shall have to plan a particular course of experimentation.

In the case of the camera-shy friend her laugh is presumably a repeatable phenomenon; in missing it once you have not missed it forever. And with enough patience and ingenuity you could capture that laugh with a telephoto lens mounted somewhere in the distance and a remote-control shutter release held in your hand. You could then snap the picture without her knowing it.

In the case of the coin in the upholstery you could, with some effort, make a special tool to retrieve it with. In fact, in *any* such everyday case—even with phenomena involving reciprocal observers—by investing sufficient effort and ingenuity in a revised version you *can* isolate the phenomenon, render it impervious to the fact that you are observing it. You will never get a perfect replay of some specific event, but as long as it is a general phenomenon and not a one-shot event you are interested in you can always reduce the effect of the observer (yourself) to as close to nil as you want. To be sure, the budget might have to be a trillion dollars, but here we are discussing principle, not pragmatics.

The point needs repeating because many people think the uncertainty principle actually applies to everyday phenomena. Nothing could be further from the truth. What, then, is Heisenberg's principle about?

To explain it we have to go back to one of Albert Einstein's three fundamental papers of the year 1905, the paper in which he postulated that light is made up of the discrete entities called photons. It was in this paper that the window onto the mysterious world of quantum mechanics was first opened. Two centuries of careful experimentation and observation had demonstrated unequivocally that visible light acts like a wave with an exceedingly short wavelength (some 10^{-4} centimeter). Light waves had been observed interfering with themselves, canceling or reinforcing themselves. Such behavior is analogous to phenomena seen on lakes or other bodies of water, such as the momentary canceling of one part of a speedboat's wake by another part reflected off a jetty, or the shimmering patterns created on a still lake by the crisscrossing circular ripples emanating from the successive bounces of a skipped rock.

In some ways light waves are simpler than water waves. Whereas water waves of different wavelengths travel at different speeds, all light waves travel at one speed: c, or 3×10^{10} centimeters per second. In water, waves of long wavelength travel faster than waves of short wavelength. Thus a single circular ripple, as it expands, breaks up into its various components. The outer edge, traveling fastest, consists of long-wavelength components; the inner edge consists of short-wavelength components. Hence gradually the ripple flattens out, or gets dispersed. Water is thus said to be a dispersive medium, whereas the medium that light waves travel through is nondispersive. What is that medium? The remarkable fact is that light needs no medium, or, if you prefer, the medium of light is a vacuum. The vacuum is nondispersive. But how peculiar it is for waves to be able to continue waving even when there is nothing to wave!

This anomaly persistently puzzled Einstein, and in 1905 his fertile mind came across two fundamental elements of the resolution. One element was the theory of relativity and the other was the counterintuitive idea of particlelike quanta out of which light waves would somehow be constituted. Where, how-

OUR NEW CAMERA VS. NOTHING ELSE.

You might say the new Minolta XG-M is in a class by itself.

Because no other 35mm SLR in its class has its combination of creative features, yet is so easy to use.

All you do is point, focus and shoot.

The automatic XG-M does everything else to give you clear, sharp, beautiful pictures. It even has electronic features to prevent mistakes. And advanced features like manual metering for advanced photographers.

To make things tougher on the competition, we gave the XG-M something you only find on more expensive cameras. The option of professional motor drive, to let you take pictures at an astonishing 3.5 frames per second.

So you can capture fast action. Like a diver's twisting entry into the water. Or a racer's moment of triumph at the finish line.

And when you hold this camera you'll know it's extraordinary, inside and out. Its built-in textured grip was sculptured to fit your hand more comfortably.

And its advanced design represents an exciting new achievement in making fine cameras. Something we've been doing for over 50 years.

The XG-M accepts over 45 interchangeable, computer-designed Minolta lenses. As well as the Minolta system of SLR accessories.

The new Minolta XG-M. There's simply nothing else like it.

WAIT 'TIL YOU SEE HOW GOOD YOU CAN BE.

MINOLTA

For more information write Minolta Corporation, 101 Williams Drive, Ramsey, N.J 07446
Or see your Minolta dealer In Canada: Minolta, Ontario, L4W 1A4

Product appearance and/or specifications are subject to change without notice

ever, did this curious flash of insight come from?

The classical theory of light as an electromagnetic wave had left a mystery about the way light of various colors, or wavelengths, is emitted from a "black body." The term is misleading; it merely means any object that absorbs light of all frequencies and does not reflect light of any frequency. As a black body heats up it begins to glow, first dull red, then bright red, then orange, eventually bluish and even white. (Think of the glowing burner on an electric stove.) The unexplained problem was to determine how much light of each wavelength is put out by a black body at a given temperature. In the water-wave analogy this would correspond roughly to predicting how deep the leading, central and trailing parts of a ripple created by a falling stone would be as a function of, say, the momentum of the stone as it hit the water surface.

The actual black-body spectrum at many temperatures had been carefully measured by experimental physicists, and the characteristic shape of the curve of intensity v. wavelength was familiar. At both very long and very short wavelengths the intensity died away toward zero, and at an intermediate value, determined by the temperature, the intensity was at a maximum. This disagreed sharply with the prediction of classical physics concerning the intensity of the various colors. Classical physics gave the prediction that at very short wavelengths, no matter what the temperature, the intensity would approach infinity. In modern terminology this amounts to saying that every object, even an ice cube, is constantly radiating lethal gamma rays at arbitrarily great intensities. This is obviously preposterous. Up to 1900, however, no one had any idea how to patch up the classical theory.

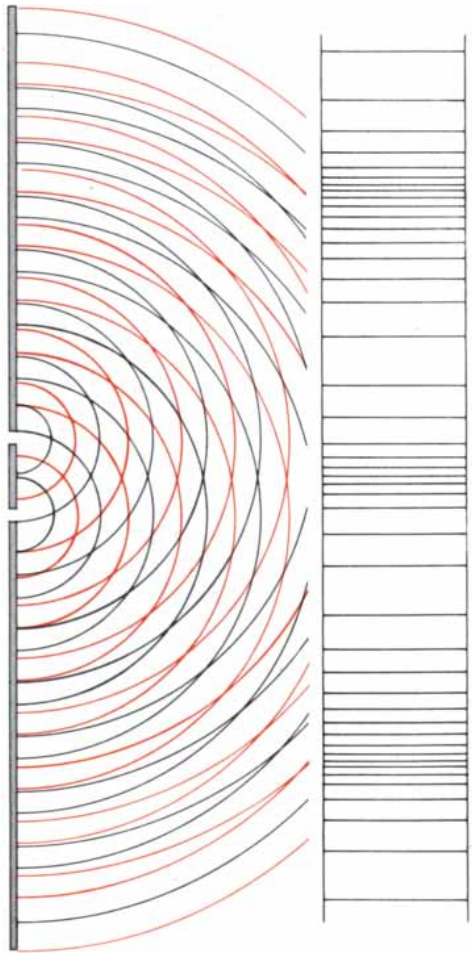
In that year Max Planck discovered a kind of hybrid formula that looked like a mathematical splicing of two different components, one pertaining to long wavelengths and the other to short wavelengths. At the longer wavelengths the formula agreed with the classical prediction and also with the measured data. At the shorter wavelengths Planck's formula diverged from the classical prediction but stayed in agreement with the data. In short, Planck's equation seemed right on the money for all wavelengths and temperatures—but it had not been derived from first principles. It was a lucky guess, although much more than luck was involved, since Planck's intuition had guided him like a bloodhound to the formula.

Planck himself was particularly baffled by the fact that he had had to throw a strange quantity he called "the elementary quantum of action, h," into his formula. What h meant was not clear. It was just a constant that, with a suitable value, would make the formula exactly reproduce the observed spectrum. It seemed therefore to be a universal constant of nature.

But what in the world was it doing in this equation? What did it mean? Einstein was the first to postulate the reason behind the appearance of Planck's constant h in the equation. Einstein began with the concept that the energy content of light waves is deposited in tiny lumps—photons—whose size has to do with h and their wavelength. For example, if the light is red, its photons carry about 3×10^{-12} erg of energy. Green photons carry about 4×10^{-12} erg. AM radio-wave photons carry somewhere between 3×10^{-21} and 9×10^{-21} erg (depending on what station you are tuned in to). The amount of energy per photon was postulated to be invariant, given its color (that is, its wavelength).

In the water-wave analogy you can try to visualize ripples that, when they reach the shore, vanish and are replaced by frogs that hop up the bank where the water would have lapped. The longer the wavelength of the ripple, the smaller the frog that jumps out. Conversely, when fine ripples with a very short wavelength reach the shore, they turn into thundering monster frogs that knock trees down and send boulders crashing into the lake (an analogue to the photoelectric effect, in which electrons of sufficient energy knock electrons out of a metal surface). Einstein's interpretation of Planck's formula implied that the frog's energy—or rather a photon's energy—and its wavelength must be inversely proportional. The equation linking them is $E = hc/\lambda$, where E is the photon's energy, h is Planck's newly discovered constant, c is the speed of light and λ is the photon's wavelength. Here E and λ are the only variables. This mixing of wave and particle viewpoints was one of the most baffling aspects of quantum mechanics, and it has continued to plague the intuition of physicists ever since, although mathematically it was greatly cleared up by the blossoming of the field in the 1920's and 1930's.

The next step en route to Heisenberg's uncertainty principle came in 1924, when Prince Louis Victor de Broglie was reflecting on the mysterious particlelike nature of light waves. He asked himself: Why should only light waves be particlelike? Why is it just a one-way street? Why not the reverse? Might not particles also have wavelike properties? De Broglie's intuition was more or less as follows: If you want to generalize



A single electron can pass through two slits and create an interference pattern **(right)**

TEN REASONS WHY BRONCO IS THE NEW FRONTIER IN 4WD.

1
BEST GAS MILEAGE
The efficient Bronco Six leads all six-cylinder utility vehicles in mileage ratings: (18) EPA est. MPG, 24 est. hwy.*

2
THE ONLY INDEPENDENT FRONT SUSPENSION
Tough Twin-Traction Beam is the only IFS in an American-built sport utility. Front wheels work independently for a stable ride plus off-road control.

3
LONGEST RANGE
No other sport utility equals Bronco for range: (576) estimated miles, 768 est. highway miles with optional 32-gal. single fuel tank.*

4
HIGHEST TORQUE SIX
Ford's husky 300 Six turns out more torque than competitive Sixes.

5
4-SPEED OVERDRIVE
Only Ford offers this new option in a 4WD vehicle. It's one reason for Bronco's dramatic 20% boost in fuel economy for '81.

6
COMFORT FOR AS MANY AS SIX
Bronco seats up to six with front and rear bench options. It's roomy and comfortable inside, trim and maneuverable outside.

9
EFFORT-LESS SHIFTS
Efficient 2-speed transfer case is designed for smooth, low-effort 2WD/4WD shifts.

7
AUTOMATIC LOCKING HUBS
New optional hubs let front wheels run free on the highway, lock in automatically when you shift into 4WD.

10
FREEDOM OF CHOICE
Bronco gives you a lot to choose from: 19 exterior colors, for example.

Built Ford Tough

8
Designed for the needs of the 80's, Bronco is built on a tough Ford Truck chassis.

*With standard 4.9L (300 CID) Six and optional 4-speed overdrive. Use for comparison. Your mileage and range may differ depending on speed, distance and weather. Actual highway mileage and range will probably be lower than estimated. California ratings less. V-8 required.

FORD

FORD DIVISION Ford

Einstein's equation so that it holds for particles other than photons, you have to get rid of the one direct reference in it to light, namely c. After all, most particles do not travel at the speed of light. Hence de Broglie thought about how to recast the equation in a c-less form.

It proved to be not too hard, because by then it was known that photons have both energy E and momentum p, and that they are related by the equation $E = pc$. If you combine the two equations, you can cancel out the c, and the result is $p = h/\lambda$. This is still Einstein's equation; it contains no new physics. De Broglie's conceptual bravery was to propose—without any experimental evidence for it—that the equation should be universal. It should apply to all matter, not only photons but also electrons, protons, atoms, billiard balls, people—even dogs! Thus Fido would have a wavelength whose value would depend on how fast he was running.

What would this mean physically? What can a running dog's wavelength mean? Well, if you calculate it, you will find that Fido's wavelength is far shorter than the radius of a proton, yet Fido finds himself considerably bigger than a proton. If Fido were very, very small, small enough for his wavelength and his size to be comparable, his wavelength would make him diffract around objects the way water waves do. But since Fido is macroscopic, his having a microscopic wavelength is all but irrelevant.

For electrons it is entirely another matter. They *are* smaller than their own wavelength. (In fact, as far as anyone knows electrons are perfect point particles, with zero radius.) Soon after de Broglie made his suggestion, experiment and theory thoroughly confirmed his notion. Electron waves were soon being diffracted in laboratories around the world, just like light waves.

Now there arises a puzzle. Are electrons spread out in space the way waves must be, or are they localized? If they are truly points, how can they be diffracted? If they are truly waves, where is their electric charge carried?

Experiments have shown that even a *single* electron can be diffracted. Richard P. Feynman in his book *The Character of Physical Law* (The MIT Press, 1967) describes it beautifully. In an idealized experiment one electron is released in the direction of a barrier with two slits in it. On the far side of the barrier is a detecting screen. The electron follows some trajectory and hits the screen somewhere. One such event simply results in one dot being made on the screen. Suppose we repeat the experiment many times, each time releasing just one electron. We get a buildup of dots on the screen. Intuition tells us clearly to expect the dots to be clustered directly behind each of the two slits, their distribution tailing off with distance. In other words, we would expect to find two clusters of dots and no other kind of distribution.

But if the de Broglie wavelength of the electron is close to the distance between the slits, the pattern on the screen after thousands of arrivals will look very different. It will be a complex regular structure characteristic of waves interfering with each other. Actually it will reproduce the intensity pattern created by a wave that splits itself into two parts, which interfere with each other on the far side of the barrier. It must be inferred that each electron, as it flew in its trajectory from the source to the screen, somehow "sensed" *both* slits and *interfered with itself* in the manner of a wave and yet deposited itself like a frog, that is, at a point, on the screen without a trace of its schizophrenia.

The dilemma is, then, that electrons act as if they are both spread out *and* localized—as if they were both waves *and* particles. This kind of wishy-washiness is inconceivable in the macroscopic realm. We have no trouble distinguishing between, say, oranges and ripples in a pond. (If you think you might be in a tight spot someday where you would need to tell quickly if something is an orange or a ripple, clip out the following wallet-sized orange-ripple distinguisher: An orange is solid, tangible and above all always *somewhere*. A ripple has no mass, is intangible and is spread out.)

If you are hungry and want to know where your orange is, you just look around, and when you sense some orange-flavored photons entering your eyes, you will have found it. Those photons bounced off the orange and into your eyes. But suppose the orange somehow grew smaller and smaller. After it got down to the size of a mitochondrion in a living cell its diameter would be about the wavelength of orange light. Then it would diffract light, and you would not be able to find it so easily. If it grew even smaller, something terrible would begin to happen. The individual photons that were hitting it would, with their momentum and energy, begin to jostle it around. The particlelike quality of photons would start to enter the picture. Indeed, an orange the size of an electron would probably be very hard to find. Therefore if you were starved for an orange, you would do better to look around for a bigger one.

Unfortunately the same cannot be said of electrons. You cannot find a bigger one. To find an electron you cannot do anything but bombard it with other particles or with photons. Since particles and photons have both particlelike and wavelike aspects, either bombardment will lead to similar consequences. If you want to locate a particle precisely, you need waves whose wavelength is about the size of that particle (or shorter). To understand this intuitively, think of the way water waves would be affected by a floating piece of wood. If they have a very long wavelength, they will not even "notice" the wood. Only if their wavelength gets down to the size of the object will they begin to be affected by it.

Hence in order to find our electron we need photons of very short wavelength. Wavelength, however, is inversely proportional to momentum. That is the deadly part of de Broglie's equation. You pay for your short wavelength by having a lot of momentum. And so as you try to diffract waves ever so gently off your particle, hoping not to move it, you will not be able to do so without transmitting momentum to it. Either you are gentle (long-wavelength photons) and do not see the electron well, or you are violent (short-wavelength photons) and throw the electron completely off its course.

Heisenberg made a careful study of this perversity that follows from de Broglie's equation and discovered, to the bewilderment of epistemology lovers the world over, that to know the position of a particle perfectly is to give up any hope of knowing its momentum, and that to know the momentum is to give up any hope of knowing its position. And to know either one imprecisely would be to impose limits on the precision with which you could know the other. A recently observed graffito illustrates this idea: "Heisenberg may have been here." The principle can be summarized even in an inequality, which Heisenberg deduced. If you are trying to determine the location of the particle, there will be an uncertainty, conventionally denoted Δx. There will also be an uncertainty in the value of the momentum, denoted Δp. Heisenberg's uncertainty principle is the inequality $\Delta x \cdot \Delta p \geq h/2\pi$.

There are a couple of things to point out here. First, note the presence of h, Planck's mysterious constant. This tells you that the effect is a consequence of the wave-particle duality of matter (and of photons) and has nothing to do with the notion of an *observer* disturbing the thing under observation. Second, note that even with this epistemological restriction, arbitrarily accurate measurement of either position or momentum is possible; it is just that you cannot get both measurements.

In short, it is a complete misinterpretation of Heisenberg's uncertainty principle to suppose it applies to macroscopic observers who are making macroscopic measurements. For example, it does not follow from Heisenberg that psychologists studying the phenomena of human cognition are somehow limited in principle by the fact that they are observing conscious human beings, who are capable of the same kind of observation. What they are limited by is their knowledge of the human brain and their ingenuity (and perhaps these days by their funding).

If you wanted to know more about

Grand Prizes: 15 round trips for 2 to Ireland on Aer Lingus.

Now, you can make that fantastic vacation you've always dreamed about, a reality. Soon, John Jameson Irish Whiskey will be giving 15 lucky grand prize winners two round trip tickets to Ireland on Aer Lingus, Ireland's quality international airline. There, they'll discover the unspoiled beauty and historic landmarks of an exciting era. A time when John Jameson himself started to distill his famous whiskey in the Dublin of 1780. And to help them enjoy their visit to the 18th century, winners also receive $1,000 in expense money.

Plus 200 other prizes

200 runners-up will win an elegant set of six imported crystal "on the rocks" glasses. (The perfect complement for the delicate taste of imported John Jameson.) To enter, just visit your favorite bar, liquor store, or restaurant and look at the front label on a bottle of John Jameson. Then, simply answer the question on the coupon in this ad and mail it in. And put 200 years of Ireland's past in your future.

OFFICIAL RULES

1. On entry form, or 3" x 5" piece of paper, print your name, address and zip code. Then answer the question with information found on the front label of any bottle of Jameson Irish Whiskey. If you don't own a bottle, visit your favorite restaurant or tavern, or go to any participating liquor store and look at a bottle of Jameson. A label may also be obtained by sending a stamped, self-addressed envelope to John Jameson Irish Sweepstakes, P.O. Box 8288, St. Paul, Minn. 55182.
2. Enter as often as you wish, but each entry must be properly completed, addressed and mailed in a separate envelope and received by September 30, 1981 to be eligible. Prize winners will be determined in a witnessed random drawing by Siebel/Mohr Incorporated, an independent judging organization whose decisions are final. Winning entries must also include the correct answer to the question asked on the entry form. No purchase required.
3. Each grand prize winner will receive 2 round trip economy tickets to Ireland on Aer Lingus plus $1,000 cash. It's incumbent upon winners to make their own arrangements for the trip. Winners shall depart from the normal Aer Lingus departure points within the Continental U.S.A. Any additional expenses incurred on the trip will be borne by the winners. Trips must be taken by December 31, 1982. Any necessary taxes must be paid by winners. The 200 second prize winners receive an elegant set of six imported crystal "On the rocks" glasses. Prizes are non-transferrable and non-redeemable.
4. Only one prize per family or household. The odds of winning will be determined by the number of correct entries received. All prizes will be awarded.
5. Sweepstakes open to residents of the Continental U.S.A., Alaska and Hawaii. Employees and their families of Calvert Distillers Company and Aer Lingus, their affiliated and subsidiary companies, liquor wholesalers and retailers, their advertising agencies and judging organization, are not eligible. Sweepstakes void where prohibited or restricted by law. All federal, state and local laws apply.
6. Entrants must be of legal drinking age under the laws of their home state.
7. A list of all major winners can be acquired at the conclusion of the sweepstakes by sending a stamped, self-addressed envelope to John Jameson Irish Sweepstakes, P.O. Box 8283, St. Paul, Minnesota 55182. NO PURCHASE REQUIRED.

JOHN JAMESON IRISH SWEEPSTAKES

I have read the contest rules listed on this page, and would like to enter the John Jameson Irish Sweepstakes. My answer is written below (correct answer appears on the front label of each John Jameson bottle).

What is the motto under the John Jameson coat of arms?

_____

Name _____

Address _____

City _____ State _____ Zip _____

Mail entries to: John Jameson Irish Sweepstakes
P.O. Box 8211
St. Paul, Minnesota 55182

For more information on travel to Ireland, call Aer Lingus toll-free (see local directory).

80 PROOF • CALVERT DISTILLERS CO., N.Y.C

John Jameson. World's largest-selling Irish Whiskey.

What if you chose as a technical

"At Nikon, HP computers save us $750,000 annually on our production line alone."

Nippon Kogaku K.K. (Nikon) is a leading Japanese optical company and maker of world-famous Nikon cameras. Nikon's Camera Division uses 13 HP 1000's in its Computer Aided Manufacturing process.

Shigehide Segawa, General Manager of Camera Production Engineering, says, "The HP 1000 is our capital partner in manufacturing cameras noted for precision and ease of use.

"With integrated circuits automating the camera's operation, our acceptance testing of IC's, adjustment and inspection of circuits during assembly, and final inspection of the end product depend heavily on computer control.

"Our HP 1000's have saved us $750,000 a year in production. But their total contribution to quality and reliability cannot be expressed by a number."

HP can be your business computer partner too!

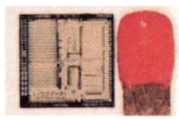
The new, top-of-the-line HP 3000 Series 44 computer—with advanced systems software—makes it easy for novices to enter, process, and retrieve data from up to 96 terminals. Thus, it's a powerful tool for high-volume distributed data processing. And, as a member of the compatible HP 3000 family, the Series 44 uses HP's award winning data base management software, and can be networked for instant information access and resource sharing. Update kits for smaller 3000's are available.

99-percent uptime service guarantee!

This unprecedented guarantee is available under full-service maintenance plans on Series 44's within 100 miles of any of the 43 HP service centers throughout the U.S.

World's most powerful computer CPU chip.

HP has developed a new proprietary chip containing 450,000 transis-



Hewlett-Packard computer partner?

"At Boeing, HP computers are helping to save $1 million a year in time and inventory costs."

Boeing's Manufacturing Research and Development Functional Test Group, in Renton and Everett, Washington, uses several HP 9845 computers to control the mechanical, electrical, and avionic testing of newly manufactured airplanes.

Says Merlin Wiese, Group Manager, "Using HP 9845s, we have developed an automated test system for the Boeing 747, and are now redesigning it for the new Boeing 757 and 767. Before automation, we were spending many days on the manual preflight testing of each new airplane.

"The HP 9845s have enabled us to improve test accuracy, speed test time significantly, and to simulate flight conditions while the airplane is on the ground. This means airplanes spend less time in test and, as a result, labor and inventory costs are reduced substantially."

tors, more than double the number previously considered the technological limit. Shown here beside a paper match at 2X magnification, this central processor chip is an example of the leading edge technology that keeps HP computers among the world's most advanced.

HP's new Microsystem: modular and low in cost.

The HP 1000 Model 5 is the smallest, lowest-priced complete system in HP's family of real-time computers. It is easy to configure for a wide

range of industrial and lab operations, and uses software packages upwardly compatible throughout the HP 1000 line, including networking, data base management, and graphics. Prices start at under $10,000*.

*Domestic U.S. prices only.

A working partnership with HP.

HP offers a free, 75-page catalog of computer products that provide solutions for Original Equipment Manufacturers. For your copy, call Dept. 304A toll free, (800) 547-3400 (except from Alaska and Hawaii). Oregon residents call 758-1010. Or write A.P. Oliverio, Vice-President, Marketing, Hewlett-Packard Co., 1502 Page Mill Rd., Palo Alto, CA 94304.

When performance must be measured by results

HEWLETT PACKARD

00144

Tested under gruelling conditions. Because life isn't a simple Sunday drive.

The Accord is Honda's luxury car. But that doesn't mean it's built to live a life of leisure.

On the contrary, the Accord is constantly tested to the very limits of performance and durability.

SOME SIMPLE ENGINEERING FACTS.

The Accord's standard engineering features are the result of years of exhaustive testing at every stage. From the design through production.

Beginning with its front-wheel drive. Which we've been perfecting for the past ten years. Including its transverse-mounted engine. Four-wheel independent suspension. 5-speed manual transmission. Rack and pinion steering. And dual-diagonal brakes.

But the testing story doesn't stop there.

DURABILITY HELPS MAKE THE ACCORD SIMPLE TO MAINTAIN.

Since no human is as durable as a Honda, mechanical devices are used to perform the toughest durability tests.

Like the continuous opening and closing of doors.

And bouncing up and down on seat cushions.

Other tests are done in refrigerated rooms or heat chambers.

And then repeated in places like Alaska and Death Valley.

Of course, the chances are you will never drive your Honda in a climate that extreme.

Or push it to those limits of performance.

But if you do, you'll enjoy the luxury of knowing it can take it.

You'll notice the results of this ongoing testing every time you drive your Honda.

But if you ever decide to sell it, you'll notice them especially.

RESALE VALUE. ANOTHER SIMPLE REASON TO CONSIDER THE HONDA ACCORD.

According to the January 1981 *N.A.D.A. Official Used Car Guide*, the first 1976 Honda Accords have an average retail value of over 90% of their original suggested retail price.

And that's something few other cars can claim.

One more simple reason why when so many Honda owners finally trade in their Hondas, they buy another Honda.

HONDA

We make it simple.

grammatical anomalies in the speech of woman *W,* there are in principle many ways you could go about it without making her self-conscious. To choose a few extravagant examples, for just a few thousand dollars you could install a hidden microphone in her home and monitor her conversations. For a few hundred thousand dollars you could have tiny radio transmitters made and secretly sewn into all her lapels. For perhaps a few million dollars you might be able to convince her she needed minor surgery of some kind, and then while she was anesthetized you could have harmless electrodes implanted in her brain to monitor her speech areas without her knowing it. If you thought that such physical interference with her brain might disturb her grammatical habits, then you might have to wait for a while longer until we figure out how neural activity can be examined remotely. These possibilities are clearly ridiculous, but the point is that *in principle* we can study macroscopic phenomena with an arbitrary degree of precision.

To recapitulate, the uncertainty principle does not state that the observer always interferes with the observed but rather states that at a very fine grain size the wave-particle duality of the probing object becomes relevant. The uncertainty principle is not an epistemological law about observation but a simple consequence of the fact that Planck's constant is not zero.

The uncertainty principle is not an axiom of quantum physics; it is a deduced principle, just as Einstein's most famous equation $E = mc^2$ was deduced from the more fundamental equations of relativity. Both are useful because they are pithy. The uncertainty principle is often applied by physicists as a rule of thumb. If you want to estimate the approximate momentum a neutron will have when it is emitted by a nucleus decaying from an excited state, a seat-of-the-pants estimate is given by $p = h/x$, where x is on the order of the dimensions of the confining nucleus. You can think of the confinement within the nucleus as making the *position* uncertainty very small, so that the neutron is bouncing around inside its "cage" with a compensatingly large *momentum* uncertainty. When the neutron escapes, a rough estimate of the momentum it will have is given by the uncertainty value.

When you examine the foundations of quantum mechanics, it becomes clear that the uncertainty principle is more than an epistemological restriction on human observers; it is a reflection of uncertainties in nature itself. Quantum-mechanical reality does not correspond to macroscopic reality. It is not just that we *cannot know* a particle's position and momentum simultaneously; the particle does not even *have* definite position and momentum simultaneously!

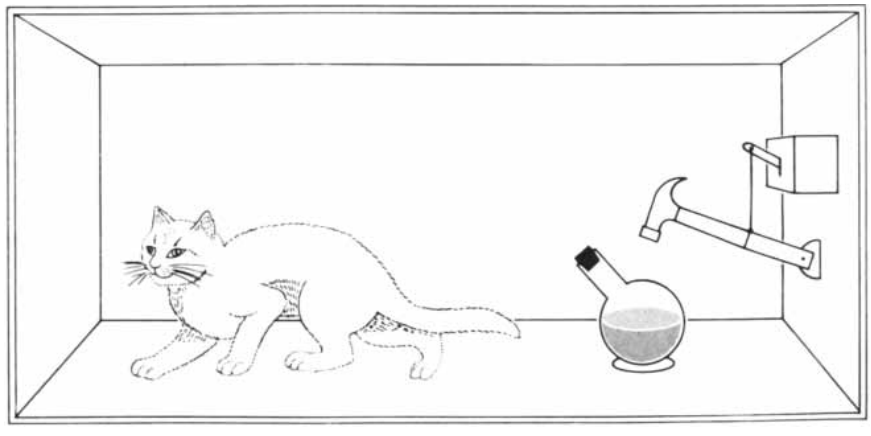
A particle is represented in quantum mechanics by a wave function describing the probabilities that the particle is here, there or somewhere else; that the particle is heading east, west, north or south, and so on. For each point in space there is what is called a probability amplitude of finding the particle there, and this number is given by the wave function. Alternatively one can read the wave function through different "mathematical glasses" and obtain a probability amplitude for each possible value of momentum. All the facts about the particle are wrapped up in its wave function. In more modern terminology the term "state" is often used instead of wave function.

In classical physics quantities such as x and p—position and momentum—directly enter the equations governing a particle's behavior. The values of x and p are definite at any one moment, and they change according to the forces that are acting on the particle. With such equations of motion physicists can plot in advance the positions and momenta of particles in simple, stable systems with incredible accuracy. An example is the motions of the planets, which even the ancients learned to predict with considerable accuracy. A more contemporary example is provided in computer space games, where spacecraft and planets are affected by a star's gravity and can go into orbit right before your eyes, swinging about in perfect ellipses on a screen. The underlying equations of such motion are differential equations, and one obvious property they have (we take it for granted) is that the motions they describe are smooth. Planets and spacecraft do not jump out of their orbit. There are no sudden discontinuities in their motion.

In quantum mechanics x and p do not enter into the equations of motion as they do in classical mechanics. Instead it is the wave function that evolves in time according to a differential equation: Schrödinger's equation (in nonrelativistic quantum mechanics). As time progresses the values of the wave function ripple through space as a water wave ripples on the surface of a lake. This would seem to imply that quantum phenomena, like nonquantum ones, proceed smoothly and with no jumps. In one sense that is right. A familiar example is the smooth precession of a spinning charged particle in a magnetic field. It is a kind of electromagnetic analogue to the precession of a spinning gyroscope in a gravitational field. The parameters that characterize the state of the spinning particle do indeed change smoothly, without any jumps.

However, and it is a big however, there are exceptions to this smooth behavior, and they seem to form just as central a part of quantum theory as the smooth evolution of states does. The exceptions occur in the act of measurement, or the interaction of a quantum system with a macroscopic one. As quantum mechanics is usually cast, it accords a privileged causal status to certain systems known as observers, without specifying what observers are (in particular without specifying whether consciousness is a necessary ingredient of being an observer). To clarify this point it is necessary to present a quick overview of the measurement problem in quantum mechanics, and for that purpose it will be useful to invoke the metaphor of the quantum water faucet.

Imagine a water faucet with two knobs, one labeled "H" and one labeled "C," each of which you can twist continuously. Water comes streaming out of the faucet, but the system has a strange property: the water is always either totally hot or totally cold; there is no in-between. The two states of the water may be called temperature eigenstates. (The prefix eigen- can be translated from the German as "particular." Here it refers to the fact that the temperature has a particular value.) The only way you can tell which eigenstate the water is in is by putting your hand in it. Actually in orthodox quantum mechanics it is trickier than that. It is the act of putting your hand in the water that *throws* the water into one or the other eigenstate. Up until that instant the water



Schrödinger's cat confronts a paradox of quantum mechanics

Cutting back without compromise isn't easy for any company. And a lot of companies that depend on electrical connectors are trying to reduce their usage of precious metals.

That's why they're coming to AMP.

Over the past decade, we've reduced the gold in our connectors by 33%. Yet they're as reliable as ever.

The reason is our advanced technology in plating. And now for the '80s, it's ACCU-PLATE—our own method of selective plating. With extraordinary precision, it uses the least amount of gold, palladium, or plating combination needed for the best performance.

As a result, our customers are not compromising. And they're saving money.

Help for the '80s.

Call us for electrical or electronic connectors or interconnection systems, and we'll respond with innovations and modern technology. What's more, we'll give you the benefit of our research and development—where we invest more than anyone else in the industry so you can spend less making your product.

A better way.

For reliability, for cost-effectiveness, for help to make your company more productive, AMP has the better ways.

Write for our brochure, "AMP Has A Better Way." AMP Incorporated, Harrisburg, PA 17105.

Barcelona • Brussels • Buenos Aires • Frankfurt • Harrisburg • Helsinki • s-Hertogenbosch • London • Luzern • Mexico City
Montreal • Paris • San Juan • Sao Paulo • Stockholm • Sydney • Turin • Toronto • Tokyo • Valley Forge • Vienna

AMP means productivity.™

is said to be in a *superposition* of states.

Depending on the setting of the knobs, the likelihood of your getting cold water will vary. Of course, if you open only the "H" valve, you will always get hot water, and if you open only the "C" valve, you will get cold water for sure. If you open both valves, however, you will create a superposition of states. By trying the same setting repeatedly you can measure the probability of getting cold water with that setting. Then you can change the setting and try again. There will be some crossover point where hot and cold are equally likely. It will then be like flipping a coin. (This quantum water faucet is sadly reminiscent of many a bathroom shower.) You can eventually build up enough data to draw a graph of the probability of cold water as a function of the knobs' settings.

Quantum phenomena are like that. Physicists can twiddle knobs and put systems into superpositions of states, analogous to the superpositions of the hot-cold system. As long as no *measurement* is made of the system, the physicist cannot know which eigenstate the system is in. Indeed, it can be shown that in a fundamental sense the system itself does not "know" which eigenstate it is in, and that it only decides (at random) at the moment the observer's hand is put in to "test the water," so to speak. Up to the moment of observation the system acts as if it were *not* in an eigenstate. For all practical purposes, for all theoretical purposes and in fact for *all* purposes the system is not in an eigenstate.

You can imagine doing a lot of experiments on the water coming out of a quantum water faucet to determine whether the water is actually hot or actually cold without sticking your hand in it. (We are assuming, of course, that there are no telltale clues to the temperature of the water, such as steam rising from it.) For example, you could run your washing machine on water from the quantum faucet. Still you will not know whether your wool sweater has shrunk until the moment you open the machine (a measurement made by a conscious observer). You could make some tea with water from the faucet. Still you will not know whether you have hot tea until you taste it (again a measurement made by a conscious observer). The critical point here is that the sweater and the tea, not having conscious-observer status themselves, have to play along and, just as the water did, enter superpositions of states: shrunk and nonshrunk, hot tea and cold tea.

That may sound as if it has nothing to do with physics per se but merely with ancient philosophical conundrums such as: Does a tree in a forest make a sound when it falls if there is no one present to hear it? The quantum-mechanical twist on such riddles is that there are observational consequences of the reality of the superpositions, consequences that are diametrically opposed to those that would follow if a seemingly mixed state were in reality always a true eigenstate, merely hiding its identity from observers until the moment of measurement. In crude terms, a stream of maybe-hot-or-maybe-cold water would act differently from a stream of water that is actually hot or actually cold, because the alternatives "interfere" with each other. This would become manifest only after a large number of sweater-washings and tea-makings, just as in the two-slit experiment it takes a large number of electron-landings to reveal the interference pattern of the alternative trajectories. Interested readers should consult the exposition of this difference in either Feynman's *The Character of Physical Law* or, for an account with more detail, *The Feynman Lectures on Physics,* by Richard P. Feynman, Robert B. Leighton and Matthew Sands (Addison-Wesley Publishing Co., 1973–75).

The plight of Schrödinger's cat (named for Heisenberg's fellow quantum-mechanical pioneer Erwin Schrödinger) carries this idea further: even a cat could be in a quantum-mechanical superposition of states until a human observer intervened. The tale of the unfortunate cat goes as follows. A box is prepared for a cat's occupancy. Inside the box there is a small sample of radium. Also in the box is a detector of radiation that will detect any decays of radium nuclei in the sample. The sample has been chosen so that there is a 50–50 probability that within any period of an hour one decay will occur. On the occurrence of such a decay a circuit will close, tripping a switch that will break a beaker filled with a deadly liquid, spilling the liquid onto the floor of the box and killing the cat.

The cat is placed in the box, the lid is secured and an hour ticks away. At the end of the hour a human observer approaches the box and opens the lid to see what has happened. According to one extreme view of quantum mechanics (and the reader should bear in mind that it is not the usual view), only *at that moment* will the system be forced to jump into one of the two possible eigenstates—cat alive and cat dead—that are represented together as a superposition in the wave function of the system. (Note that it is necessary for the randomness to be of a clearly quantum-mechanical origin: the decay of the radium nucleus. This thought experiment would not pack any punch if there were a spinning roulette wheel in the box instead of a radium sample.)

One might object and say, "Wait a minute! Isn't a live cat as much of a conscious observer as a human being is?" It probably is, but note that this cat is possibly a dead cat, and therefore certainly not a conscious observer. We have in effect created in Schrödinger's cat a superposition of two eigenstates one of

which has observer status and the other of which lacks it. Now what do we do?

To many physicists the distinction between systems with observer status and those without it has seemed artificial and even repugnant. Moreover, the idea that the intervention of an observer causes a "collapse of the wave function"—a sudden jump into one randomly chosen pure eigenstate—introduces caprice into the ultimate laws of nature. Einstein's lifelong belief was that God does not play dice ("Der Herrgott würfelt nicht").

A radical attempt to save both continuity and determinism in quantum mechanics is known as the many-worlds interpretation, first proposed in 1957 by Hugh Everett III. According to this bizarre theory, no system ever jumps discontinuously into an eigenstate. What happens is that the superposition evolves smoothly with its various branches unfolding in parallel. Whenever it is necessary, the state sprouts further branches that carry the various new alternatives. For example, there are two branches in the case of Schrödinger's cat, and they develop in parallel. "Well," you say, "what happens to the cat? Does *it* feel itself to be alive or does it feel itself to be dead?" One must wonder. Everett would answer: "It depends on which branch you look at. On one branch the cat feels itself to be alive and on the other there is no cat to feel anything." With intuition beginning to rebel, you then ask: "Well, what about a few moments before the cat on the fatal branch died? How did the cat feel *then?* Surely the cat can't feel two ways at once. Which of the two branches contains the genuine cat?"

The problem becomes even more intense as you realize the implications of this theory as it applies to you here and now. For every quantum-mechanical branch point in your life (and there have been billions on billions) you have split into two or more you's riding along parallel but disconnected branches of one gigantic universal wave function. (By this term is meant the enormous wave function representing all the particles in all the parallel universes.) At the critical place in Everett's article where this difficulty arises he calmly inserts the following footnote:

"At this point we encounter a language difficulty. Whereas before the observation we had a single observer state, afterward there were a number of different states for the observer, all occurring in a superposition. Each of these separate states is a state for an observer, so that we can speak of the different observers described by different states. On the other hand, the same physical system is involved, and from this viewpoint it is the *same* observer, which is in different states for different elements of the superposition (i.e., has had different experiences in the separate elements of the superposition). In this situation we shall use the singular when we wish to emphasize that a single physical system is involved [Everett is referring to the multiply branching universe], and the plural when we wish to emphasize the different experiences for the separate elements of the superposition. (E.g., 'The observer performs an observation of the quantity *A,* after which each of the observers of the resulting superposition has perceived an eigenvalue.')"

All of this is said with a poker face. The problem of how it feels *subjectively* is not dealt with; it is not even swept under the rug. It is probably considered meaningless.

And yet one simply has to wonder: Why, then, do I feel myself to be in just *one* world? Well, according to Everett's view, you do not: you feel all the alternatives simultaneously. It is just *this* you going down *this* branch who does not experience all the alternatives. This is completely shocking. In his story "The Garden of Forking Paths" Jorge Luis Borges describes a fantastic vision of the universe in the following way:

"'The Garden of Forking Paths' is a picture, incomplete yet not false, of the universe as Ts'ui Pên conceived it to be. Differing from Newton and Schopenhauer, [he] did not think of time as absolute and uniform. He believed in an infinite series of times, in a dizzily growing, ever spreading network of diverging, converging and parallel times. This web of time—the strands of which approach one another, bifurcate, intersect, or ignore each other through the centuries—embraces every possibility. We do not exist in most of them. In some you exist and not I, while in others I do, and you do not, and in yet others both of us exist. In this one, in which chance has favored me, you have come to my gate. In another, you, crossing the garden, have found me dead. In yet another, I say these very same words, but am an error, a phantom."

That quotation is featured at the beginning of *The Many-Worlds Interpretation of Quantum Mechanics: A Fundamental Exposition,* edited by Bryce S. DeWitt and Neill Graham (Princeton University Press, 1973). It is rather eerie to find its bizarre theme reproduced at the core of what is supposedly our stablest and least erratic science.

And yet in a way it is not so surprising. There is a clear connection between the imaginary worlds of our minds and alternate worlds evolving in parallel with the one we experience. The proverbial young man picking apart a daisy and muttering "She loves me, she loves me not, she loves me, she loves me not..." is clearly maintaining in his mind (at least) two different worlds based on two different models for his beloved. Or would it be more accurate to say that there is *one* mental model of his beloved that is in a mental analogue of a quantum-mechanical superposition of states?

And when a novelist simultaneously entertains a number of possible ways of extending a story, are the characters not, to speak metaphorically, in a mental superposition of states? If the novel never gets set on paper, perhaps the split characters can continue to evolve their multiple stories in their author's brain. Furthermore, it would even seem strange to ask which one is the *genuine* version. All the worlds are equally genuine.

Perhaps one way to think of the universal wave function is as the mind—or brain, if you prefer—of God, in which all possible branches are being simultaneously entertained. We would be mere subsystems of God's brain, and *these* versions of us are no more privileged or authentic than our galaxy is the only genuine galaxy. God's brain, conceived in this way, evolves smoothly and deterministically, as Einstein always maintained. The physicist Paul Davies, writing on just this subject in his recent book *Other Worlds* (Simon and Schuster, 1980), says: "Our consciousness weaves a route at random along the ever-branching evolutionary pathways of the cosmos, so it is we, rather than God, who are playing dice."

Yet this leaves unanswered the most fundamental riddle that each of us must ask: Why is my unitary feeling of myself propagating down *this* random branch rather than down some other? What law underlies the random choices that make me go down this branch? Why doesn't my feeling of myself go along with the other me's as they split off, following other routes? In a way this is almost like asking: Why do I feel *my* feelings and not someone else's? What attaches *me* to the viewpoint of *this* body evolving down *this* branch of the universe at *this* moment in time? The question is so basic that it almost seems to defy clear formulation in words. And the answer does not seem to be forthcoming from quantum mechanics. In fact, this is exactly the collapse of the wave function reappearing at the far end of the rug under which it was shoved by Everett. It turns into a problem of personal identity, no less perplexing than the problem it replaces.

One can fall even more deeply into the pit of paradox when one realizes that there are branches of this one gigantically branching universal wave function on which there is no Werner Heisenberg, no Max Planck, no Albert Einstein, branches on which there is no evidence for quantum mechanics whatsoever, branches on which there is no uncertainty principle or many-worlds interpretation of quantum mechanics. There are branches on which the Borges story did not get written, branches in which this column did not get written. There is even a branch in which this entire column got written exactly as you see it here, except that it ended with a different word.

YOU DON'T GET TO BE THE BEST SELLING IMPORT IN MAINE, VERMONT AND ALASKA ON LOOKS ALONE.

It isn't just the mountains that are rugged in Maine. The driving conditions and the price of gas are no picnic either.

The same goes for Vermont. And certainly for Alaska.

That's why it's very, very interesting that in all three states the #1 imported car is Subaru.

Why? Our cars are dependable. Easy on gas. And easy to handle.

We make a full line of front wheel drive and *On Demand* 4-wheel drive vehicles that can handle almost anything.

And on top of being very affordable, they've earned a reputation for durability and reliability. Great to have no matter where you live.

In fact, even if you only see snow on television, wouldn't it be great to have a car that's over-qualified?

SUBARU®
INEXPENSIVE. AND BUILT TO STAY THAT WAY.

*1981 EPA ESTIMATES FOR OUR 4WD WAGON. USE ESTIMATED MPG FOR COMPARISONS. YOUR MILEAGE MAY DIFFER DEPENDING ON DRIVING SPEED, WEATHER CONDITIONS AND TRIP LENGTH. ACTUAL HIGHWAY MILEAGE WILL PROBABLY BE LESS.

BOOKS

Arms and Japan, regular patterns, Lucy of Hadar and the psychology of psychics

by Philip Morrison

GIVING UP THE GUN: JAPAN'S REVERSION TO THE SWORD, 1543–1879, by Noel Perrin. Shambhala Publications, Inc., 1920 Thirteenth Street, Boulder, Colo. 80302 ($4.95). In 1543 a Chinese trading vessel anchored in the small harbor of Tanegashima Island, 20 miles south of Kyushu, the southernmost big island of Japan. Three of the tough crew, rovers and adventurers, were Portuguese seamen, the first Europeans known to reach Japan. Two of them had harquebuses; the moment the feudal master of the island saw one of these foreigners bring down a duck, "the gun enters Japanese history." Lord Tokitaka began lessons in shooting, and within the month he had bought the two guns at a princely price in gold, 30 years' fair wages for a workingman. The day of the purchase Tokitaka set his chief swordsmith to copying the weapons.

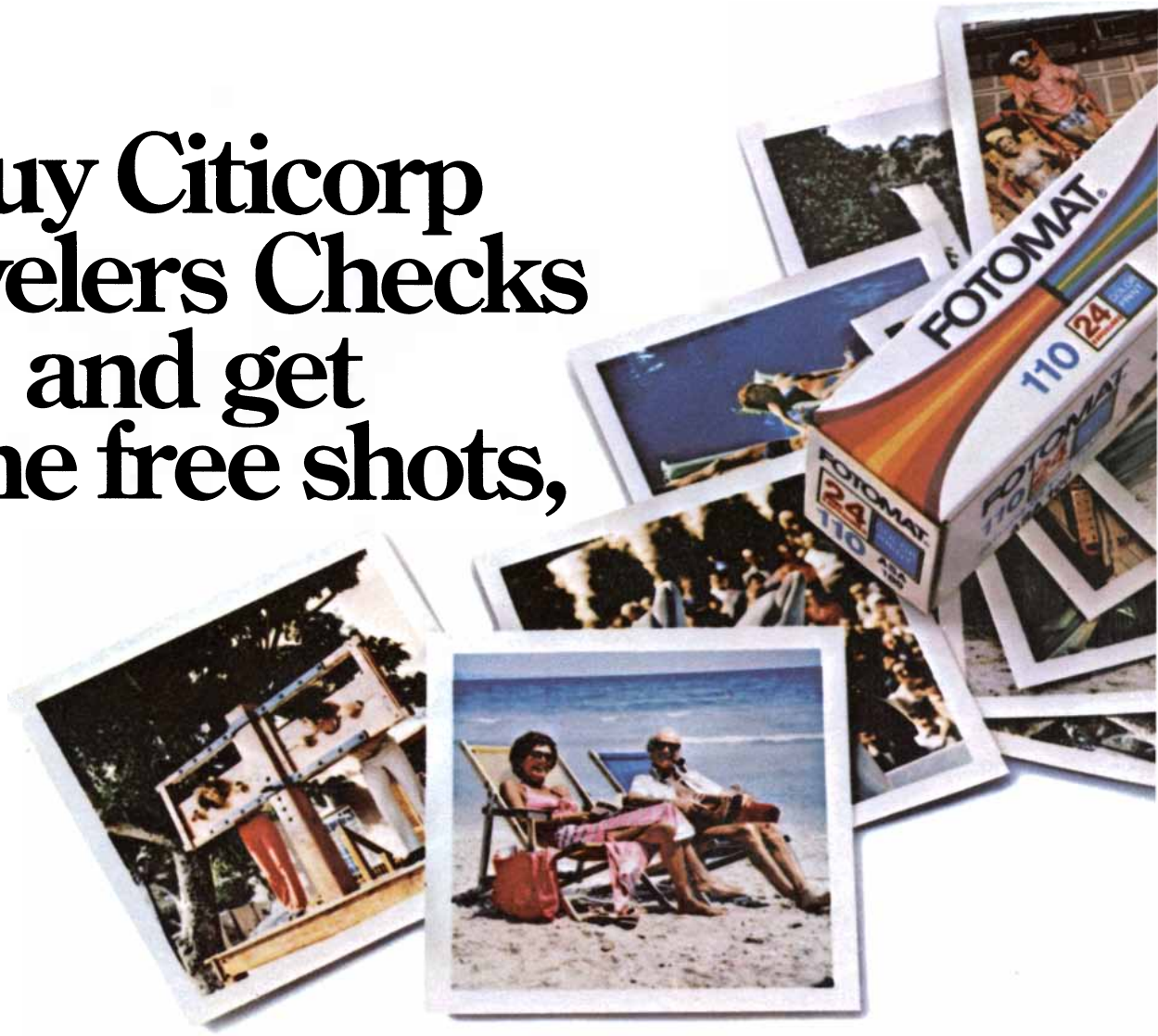
Japan was in the midst of her bloody century of feudal wars. Her craftsmen in iron and steel, her armorers, were perhaps the most skilled in the world; Japan was a leading exporter to all Asia of swords, pikes and armor. The country was populous, having six times the head count of contemporaneous England. The island was fertile soil, and the gun took root. True, the imported matchlocks did not catch on at once; the harvest was slow. Within a generation, however, the matchlock had been naturalized, refined, enlarged in caliber and muzzle energy and even fitted with rainproofing devices of Japanese design (shown here in one of the many delightful prints). The gunsmiths had been both inventive and busy; their improved product was no longer rare.

The great Lord Oda, who was to begin the unification of the nation, had as a young commander praised the spear as "the weapon on which to rely in battle." In 1575, however, he appeared at the terrible Battle of Nagashino, at the head of his army of 38,000 men, 10,000 of whom bore matchlocks. The men were unchivalrously concealed behind breastworks across a river; they held their fire until the time came to shoot "in volleys of a thousand." Bullets ended the noble samurai charge. Oda's gunpowder infantry tactics had been a brilliant success; within a few months after Nagashino two cannon, the first Japanese-made artillery, were test-fired for him, bronze two-pounders. Oda's matchlock corps at Nagashino had enlisted many young farmers, even though specially trained samurai knights stiffened it. "It was a shock to everyone to find out that a farmer with a gun could kill the toughest samurai so readily." Skill had been transferred back to the gunsmith from the soldier and up to his commander.

But in 1853 when Commodore Perry anchored his black ships in Tokyo Bay, the gunports of his flagship showed 64-pounder cannon and some even a little bigger. The shogun's samurai harbor defense manned cannon cast before 1620, six- and eight-pounders. One order of magnitude and more separated the arms of the barbarians from those of the men of honor. Firearms R&D had come to a full stop in Japan before its craftsmen had ever gone to flintlocks, say by 1650. Japan had plenty of flint, plenty of skilled artisans and a number of late Dutch prototypes. By then, however, they had simply lost interest in guns, except for the maintenance of an archaic coast defense. The lesson of Nagashino had been well absorbed. Once the feudal lords were brought securely under the unifying rule of the Tokugawas the newly centralized state began slowly and cleverly to undo the nation's firearms industry. Senior gunsmiths were promoted to samurai and converted to swordmakers. The scattered provincial gunmakers were required to move to either of two centers of the industry, and by 1607 they had all been assembled. They found only a dribble of orders from the central government and a salary paid them even if no guns were made. The land was at peace, the gun trade was reduced to a symbol: a few hundred matchlocks a year for almost a century (with some exceptions) in a country counting half a million sword-bearing warriors. Politics was in command; the technology of war had to obey.

The author, a Dartmouth English professor, has drawn much material

Buy Citicorp Travelers Checks and get some free shots,



and a shot at these free prizes.

Every time you buy $600 worth of Citicorp Travelers Checks, you'll get a coupon good for a free package of Fotomat® color film. Plus, get a chance to win a free trip around the world by entering our "Carefree Vacation Sweepstakes."

If you don't win the Grand Prize, don't despair. Our other prizes are also rather grand. You could still win a Hawaiian vacation, a portable Sony TV/Radio/Clock combination, a John Weitz Luggage Set, or a Kodak Ektralite™ Camera with built-in flash.

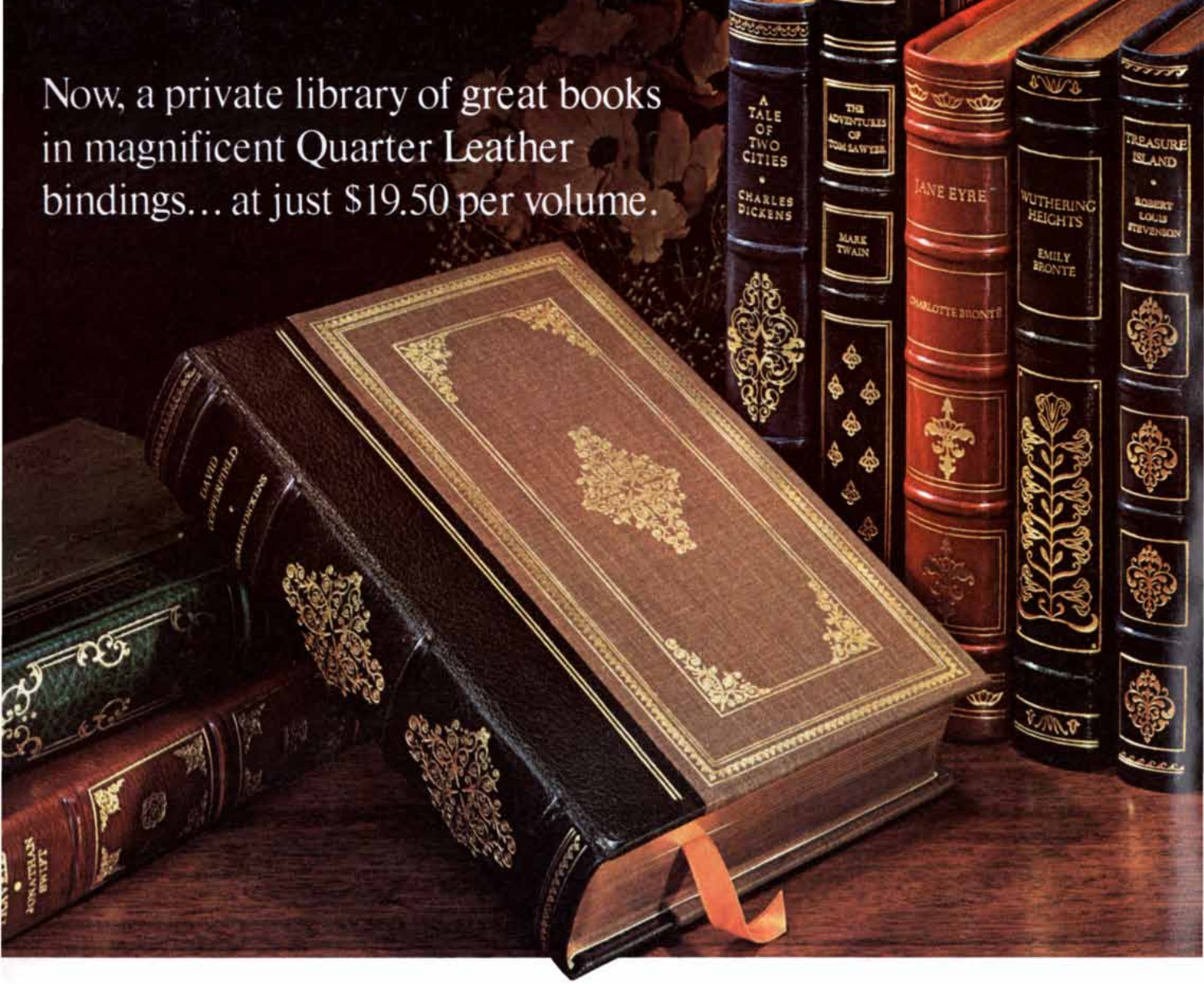
Besides pictures and prizes, Citicorp Travelers Checks give you the perfect way to protect your cash when you travel. Our checks are good in millions of places, and you can get a refund at tens of thousands of locations around the world. You'll be able to find out for yourself if you win our Grand Prize trip. THERE'S NO PURCHASE NECESSARY TO ENTER OUR SWEEPSTAKES—here are the official rules.

OFFICIAL RULES—NO PURCHASE NECESSARY

1. On an official entry blank or on a plain 3" x 5" piece of paper, hand print your name and address, including zip code and the name Citicorp. No purchase is required.
2. Enter as often as you wish, but mail each entry separately to: CITICORP TRAVELERS CHECKS SWEEPSTAKES, P.O. Box 2902, Westbury, New York 11591. Entries must be postmarked by September 30, 1981 and received by October 15, 1981. Entries become the property of Citicorp. None will be returned.
3. All prizes will be awarded. The approximate retail value of the prizes is as follows: Grand Prize—$25,000; Two First Prizes—$7,300 each; 125 Second Prizes—$274.95 each; 170 Third Prizes—$180 each; and 1,000 Fourth Prizes—$79.95 each.
4. Trips must be taken before October 31, 1982. Travel and accommodations are by winner's choice, not to exceed $15,000 (Grand Prize) and $6,300 (First Prize). The Grand Prize winner may elect to receive $25,000.00 in Citicorp Travelers Checks in lieu of the trip. Trip winners who are under eighteen must choose as the person to accompany them a parent or legal guardian. Trip winners must sign a waiver of liability.
5. Winners will be selected in a random drawing conducted on October 21, 1981 by National Judging Institute, Inc., an independent judging organization whose decisions are final. Winners will be notified by mail. One prize per household. No substitution or exchange of prizes except as specified in Rule No. 4 above. The odds of winning any prize will be determined by the number of entries received.
6. All taxes are winners' responsibility. Winners may be asked to execute an affidavit of eligibility and release. The Sweepstakes is void wherever prohibited by law. All Federal, State and local regulations apply.
7. Sweepstakes open to all residents of the United States except where prohibited or restricted. Employees (and their families) of Citicorp, its subsidiaries, advertising and promotional agencies and selling agents are not eligible.
8. For a list of prize winners, send a stamped, self-addressed envelope to: Citicorp Travelers Checks Winners List, P.O. Box 2974, Westbury, New York 11591.

Start your trip with Citicorp Travelers Checks, because you never know where you'll end up. Especially if you win our Sweepstakes. And no matter where you end up, you'll be able to take some free pictures. See your participating Citicorp Travelers Checks selling agent for full details.

CITICORP®
TRAVELERS CHECKS



THE FAMILY LIBRARY OF THE WORLD'S GREAT BOOKS

- *Fifty of the greatest, most enjoyable books ever written, selected by a distinguished board of advisers.*
- *In elegant, enduring bindings of fine leather and fabric, leather accented with 22 karat gold.*
- *With exciting illustrations in every volume.*
- *For only $19.50 per volume—a superb family library to be enjoyed now and treasured as an heirloom in years to come.*

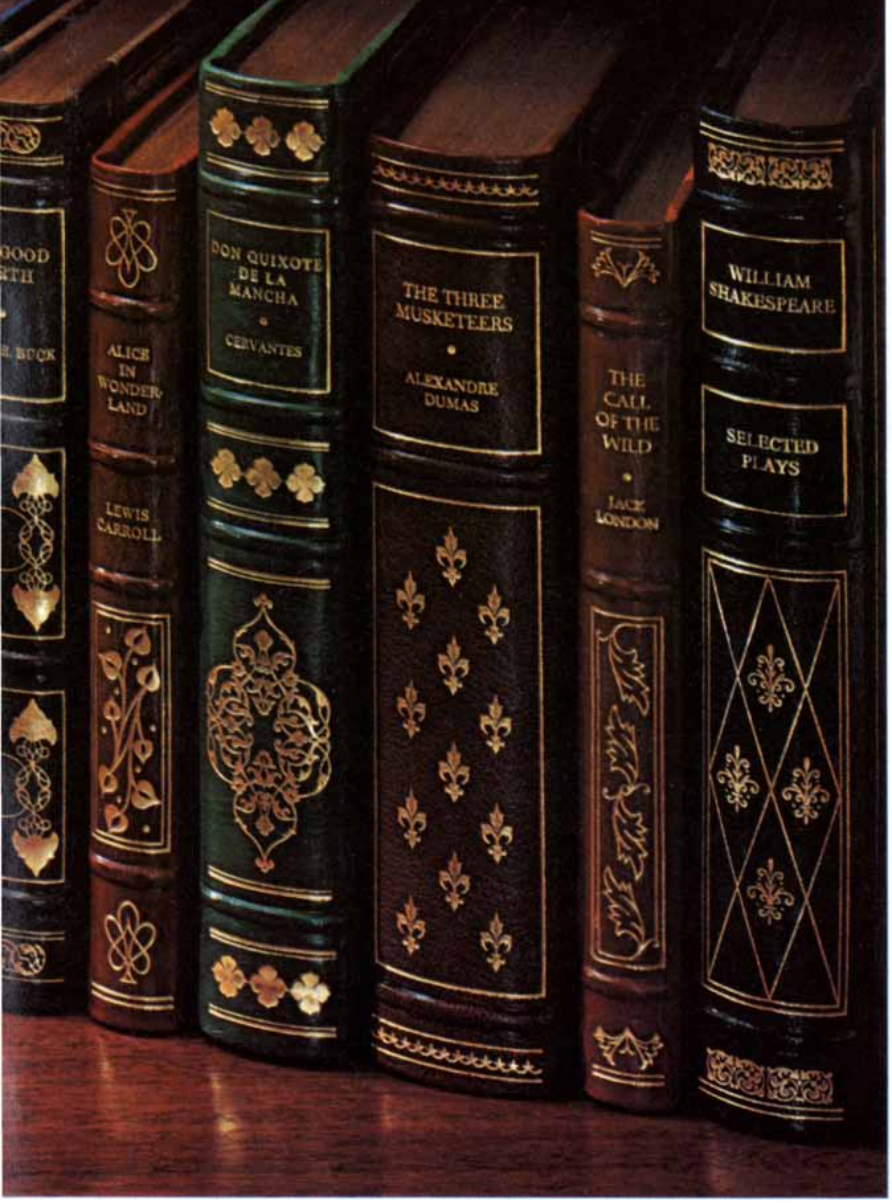
"Books are the treasured wealth of the world," wrote Thoreau—"the fit inheritance of generations." For great books have the power to enrich the mind and spirit. To expand our horizons. To make our lives more interesting. To make *us* more interesting.

Now, The Franklin Library brings you the fifty greatest, most enjoyable, most readable, most memorable classics of world literature in an edition of great and enduring beauty...at an affordable price.

These are the *universal* classics—the very core of the cultural heritage of all mankind—specially printed and bound to enhance the joy of reading, and to grace the finest homes.

The traditional English Quarter Leather binding of every volume in this edition will be a beautifully coordinated combination of leather and fine fabric. The leather will be ornamented with 22 karat gold. The fabric will be embellished with original designs. And each volume will be enriched by elegant endleaves...exciting illustrations...and fine bookpapers.

As a subscriber, the first volume you will receive will be Dickens' *David Copperfield,* one of the greatest novels



ever written. A novel whose fascinating story closely parallels Dickens' own life. "Of all my books," Dickens said, "I like this the best."

And each month thereafter, you will receive one additional book—its pages filled with the world's greatest writing. Each book a delight to read and reread, until its scenes and characters and insights—the *life* and *experience* within its pages—become the stuff of memory ...leaving you and your family forever enriched.

Books of enduring beauty

As you open each volume and turn its pages, you will discover new delights. The decorative endpapers and bound-in ribbon page marker—carefully and tastefully color-coordinated to the book's covers. The page edges, gilded with a special tarnish-free finish, adding protection as well as beauty to each volume. Magnificent illustrations, some of them in full color. And the bookpaper itself, specially milled to last for many generations.

For these luxurious volumes will be crafted to *retain* their beauty. So that this private library will be a lasting heirloom. To be treasured by you and your family now, and by your children and your children's children in years to come as "the fit inheritance of generations."

Imagine the satisfaction of glancing around your living room and seeing this treasury of great literature—the fifty beautiful volumes which make up The Family Library—the rich leather of their distinctive spines enriched by 22 karat gold ornamentation.

Imagine the pleasure of anticipation, as you slide one of the volumes from its place, open it at the bound-in ribbon marker, and begin to read. Enjoying the handsomely printed text, the fine, evocative illustrations.

And imagine, too, seeing your family immerse themselves in these challenging and enduring works of literature—and reaping all the rich rewards which that entails.

Return your application postmarked by July 31, 1981.

The Family Library of the World's Great Books is being offered at this time at the attractively low price of just $19.50 per volume. But it is available at this price *only* to those who subscribe to the complete collection.

The accompanying application assures you of the *guaranteed* price of just $19.50 per volume for the entire Family Library. And you have the right to cancel your subscription at any time upon 30 days' notice, or return any book, for any reason, within 30 days. To be accepted, however, your application should be postmarked no later than July 31, 1981.

THE BOARD OF ADVISERS

The Franklin Library gratefully acknowledges the assistance of the very distinguished Board of Advisers in selecting the fifty great books to be included in The Family Library.

ISAAC ASIMOV
Noted scientist, professor, critic and writer.

PAULINE FREDERICK
Noted correspondent, author and news analyst.

DR. BRUNO BETTELHEIM
Distinguished psychologist and author, perhaps best known for his work with children.

HELEN HAYES
One of America's greatest dramatic actresses, winner of two Academy Awards.

RALPH ELLISON
Distinguished professor and author of the award-winning novel *Invisible Man*.

JEAN KERR
Widely popular writer of perceptive and humorous plays and books.

SUBSCRIPTION APPLICATION

THE FAMILY LIBRARY OF THE WORLD'S GREAT BOOKS

The Franklin Library 281
Franklin Center, Pennsylvania 19091

Please enter my subscription to *The Family Library of the World's Great Books,* consisting of 50 volumes bound in genuine leather combined with fine fabrics. The books will be sent to me at the rate of one each month, and the issue price of $19.50* per book will be guaranteed to me for the entire library. However, I have the right to cancel my subscription at any time on 30 days' written notice, or return any book within 30 days.

I need send no payment now. I will be billed for each volume, individually, in advance of shipment.

**Plus my state sales tax and $2.50 per book for shipping and handling, subject only to postage increases.*

Signature _____
ALL SUBSCRIPTIONS ARE SUBJECT TO ACCEPTANCE.

Mr.
Mrs.
Miss _____
PLEASE PRINT CLEARLY

Address _____

City _____

State, Zip _____

Please mail by July 31, 1981.

Canadian residents will be billed for each volume in advance of shipment at $23.50 plus $2.50 for shipping and handling ($CDN)

How to slim down.
Save energy. Use solar energy.
Jump start your car. Deal with stress.
Remove a stain. Check for breast cancer.
Select a smoke detector. Get better mileage.
Control pests. Cope with arthritis. Get a patent.
Insulate your home. Control your blood pressure.
Rent a home. Get rid of a headache. Spot a con job.
Keep records. Invest.
Budget your money.
Prevent drug abuse.
Garden organically.
Make toys out of junk.
Repair a leaky faucet.
Choose a new carpet.
Restore an old house.
Start a small business.
Learn the metric system.
Jog successfully. Backpack.
Read labels. Avoid sunburn.
Relieve the common cold.
Buy a car. Save money.
Administer first aid.
Donate your body.
Find a job. Retire.
Tune up your car.
Grow tomatoes.

from scholarly Japanese sources available to him in special translation. He has collected a striking set of illustrations, and his brief, readable text is full of helpful and often surprising comparisons with the history of Europe most readers know much better. The reasons behind this parable are many; certainly the most important one is the geopolitical status of the Empire of Japan, so long isolated from foreign entry by the sea and the Divine Wind. Two rebellions serve to accent the account. In 1637 the last effort of outlawed Christian Japan to reclaim freedom was the rising at Shimabara. Both sides then were armed with guns; the outnumbered rebels were killed to the last man. Long afterward, once the black ships had come and gone, a strong faction of the samurai, nearly one in 10 of them, joined in counterrevolution against the modernizing Meiji Emperor. In 1877 it was patrician swordsmen against peasant soldiers in Frenchified uniforms, with rebuilt matchlocks and some new rifles. After the decisive battle the scattered rebels yielded a few small firearms, mainly antique matchlocks, and a mound of captured swords rising "at least 10 feet from the ground." The march of military Japan was once again in melancholy step with the warlike West. Today's Japan, however, has at least proscribed war by nuclear fire.

HANDBOOK OF REGULAR PATTERNS: AN INTRODUCTION TO SYMMETRY IN TWO DIMENSIONS, by Peter S. Stevens. The MIT Press ($37.50). That we need to mind our *p*'s and *q*'s is a profound consequence of the nature of the near-Euclidean space in which we dwell, particularly important when we address paper, wall, pot or cloth. The surfaces we make and study, two-dimensional regions whereon we mark maps of worlds real or fanciful, are often given over to repetitive order. That repetition, as common in organic form as in decorative art, implies subtle constraints, unbreakable law. "There is something in such laws that takes the breath away. They are not discoveries or inventions of the human mind but exist independently of us." So wrote M. C. Escher, whose lifetime of artistic exploration of symmetry in two dimensions bridged decades of pioneer solitude to a briefer time of all-but-ubiquitous publication.

This handsome and substantial volume carefully presents an attractive path to understanding of a branch of applied mathematics. The arguments and notation of the crystallographers are central, under the explicit influence of such scholars as Hermann Weyl and Martin Buerger. Peter S. Stevens is well known for his fine book of some years back on form in nature. He has directed this volume "to the practicing artist and designer," not only as another helpful introduction to group theory in the plane but also as a well-ordered reference handbook of repetitive designs.

Around a framework of definition and example, keyed by a prototype pattern of the obliging asymmetric motif of the paisleylike comma, the author (and his drafting partner) has ordered some 300 pages of patterns. Almost all are redrawn from the widest tradition of decorative art and architecture. A number of Escher's prints are included, and a few significant photographs of the real world as well. The exposition begins with the fundamental operations in the plane, including the rotations and reflections about a point. A 24-fold interlaced Russian star and a mosaic from Roman Ostia exemplify the highest-order symmetries shown; the circle, of course, is the infinitely rotatable form of the Pythagoreans.

The interaction of the symmetry operations, however, is the deepest point made. It is this interaction that merges the mirror-related *p*'s and *q*'s, *b*'s and *d*'s by the strictest of rules, to give rise to exactly seven line groups, the only classes that allow infinite repetition of a line of motifs. About halfway through the text there is a recapitulatory page that musters all seven groups as constructed from the four lowercase letters; there are plainly no more. The rest of the text is devoted to filling the entire plane; here the catalogue is longer and rather more difficult to establish. Seventeen plane groups emerge, illustrated in their turn by hundreds of examples, together with careful pedagogical diagrams that make quite clear such knotty matters as the arbitrariness of the unit cell. On that fact rest the ingenious distortions of Escher's interlocking figure-and-ground play, which the student is invited to imitate.

It is easy to print a border of six-pointed asterisks, implicitly of any length. Their rotational symmetry, however, does not affect the symmetry of the whole; the line cannot be rotated into itself except through one or more complete half-turns. The six-foldedness remains sterile, a local phenomenon. In the same way it is attractive to see the fivefold symmetry of the starry insignia of a full General of the Army, of the apparent orbit of Venus viewed from the earth (perhaps the most unexpected example drawn here), of a cut apple and of a mortuary planned by Frank Lloyd Wright. But over the entire plane fivefold axes cannot inform any pattern. The proof is here, in the elegant manner of W. Barlow, and we are shown the efforts of Albrecht Dürer, Johannes Kepler and the great Moorish designers of the 14th century to evade the law, which rests on the nature of the first few integers themselves. It is this richness of example that makes the volume of interest to the student of mathematics, even though many other tighter treatments exist. Amateurs in particular stand to gain much.

The apparatus provided for this book is excellent. The index is a full one, allowing the location of a pattern by its culture of origin, its motif or its mathematical nature. The great majority of the ornaments are redrawn in the purity of a hard-edge black-and-white diagrammatic style; the work loses a little of the delicious textural noisiness that ought to separate a Mimbres pot from an American appliqué quilt, but perhaps this is in the nature of a clear mathematical approach. One can complain, however, that the precise sources of the illustrations are rarely cited, a slip of editorial judgment in a volume otherwise meticulous, beautiful and instructive, both in the short run and the long. "See how various the forms and how unvarying the principles," wrote Owen Jones in one of the wreaths of aphorisms that ornament the writing as the heap of patterns entrances the eye.

LUCY: THE BEGINNINGS OF HUMANKIND, by Donald C. Johanson and Maitland A. Edey. Simon and Schuster ($16.95). It is about 20 years since we encountered Honest Jim Watson, whose brash and artful story of the finding of the double helix opened for so many readers the sense of the humanity within scientific investigation. In this excellent book, which shares the form of the earlier one to a remarkable degree, we encounter another youthful and brilliant success, Lucky Don Johanson. He agrees with the epithet: "I know I am lucky, and I don't try to hide it." You have to be lucky to succeed so soon in paleoanthropology. Like Watson, Lucky Don has an intellectual counterpoise, acute, skeptical, rigorously bound to the facts. "When I met Don, I was suspicious of him," recalls Tim White, now of Berkeley. "I'm suspicious of any anthropologist who wears Gucci loafers and Yves Saint Laurent pants. He had to prove to me that he had the same attitude toward careful fieldwork that I had. Well, he did prove it."

The correspondence with the double helix goes still deeper. Offstage, like Linus Pauling, half heroes, half rivals, are the Leakeys, an entire family of them. Just as Pauling first read the helix out of the protein X rays, so did Louis Leakey first open up the new fossil sites of ancient humankind north along the Rift. It was F. Clark Howell—Don's Ph.D. supervisor—who first followed Leakey's tip north to Omo in Ethiopia. It was hard to enter that country across a lonely, nervous border. The sites became really accessible only by the good fortune that in 1965 the Emperor Haile Selassie

came on a state visit to Kenya. His host, President Kenyatta, summoned Leakey, an old friend, to outline the "fossil wealth of Kenya and Tanzania." "'How is it,' asked Selassie, 'that there are no fossils in my country?' 'There are,' said Leakey.'" The emperor is gone now, his regime swept away, but tact, energy, good fortune and the common interest of human beings in our descent has kept open the Ethiopian door. In early 1980 Lucy herself and the entire collection of hominid fossils from Hadar ("more than 350 priceless bones") were returned after five years of study in Cleveland to their permanent home in the National Museum at Addis Ababa.

This tale too demonstrates unexpected prerequisites for scientific work. Watson needed an adaptation to his Cambridge milieu rather different from Johanson's, out in the badlands of Hadar. Where Jim felt it necessary to take subjective account of the au pair girls and the other "Cambridge popsies," Don faced quite different problems: dealing with broken Land-Rover differentials, getting a reliable supply of goat meat for camp meals, mustering impromptu diplomacy to negotiate with "a file of eight sullen men carrying submachine guns" who walked into camp one morning. The ukases of granting agencies, publication snarls and the sharp rivalry of competitors unite the two accounts neatly. If Watson and Crick were not exactly generous competitors in the race for the helix, the strife among paleoanthropologists startles an innocent reader. Johanson tells us of an unnamed American professor, disappointed in his effort to get a per diem allowance from Johanson's grant for more days than his actual time in the field, who then wrote the Ethiopian authorities a letter calling Johanson incompetent and falsely accusing him of giving over his Ethiopian fossil take to the Kenyans! Cambridge and London crystallographers never played that rough.

There is more to this lively volume than the human process: there is a genuine introduction to content. *The Double Helix* offered little of that; X-ray analysis and molecular model building are easy to allude to but difficult to elucidate at the level of the unprepared reader. *Lucy* includes a few gossipy, informative chapters on the pioneers of the study of fossil man, from the Piltdown fraud to Leakey's luck, and a clear first-person account of C. Owen Lovejoy's conjecture that changing sexual behavior (the human female, alone among primates, is sexually receptive more or less at all times) was the earliest step toward hominization. Tim White remains unconvinced: "Sexual behavior leaves no fossils." The argument is interesting but serves to point up how little is secure in a field where all the bones from the best site yet examined occupy one corner of a tabletop.

The best part of the exposition in this book—we may thank the experienced science writer who is Johanson's coauthor—is the fresh account of radioisotope dating. It is not that the mass spectrometer, say, is made very clear; there is not even a diagram of particle orbits. The argon-dating concept and the arithmetic are easy enough. What is convincing is the view of the technique as the man who does the dating sees it. The chief task is to bring home a reliable sample. ("The condition of his samples is as important as the samples themselves.") Volcanic ash is easily contaminated for the dater by even a tiny inclusion of blown-in dust. Scrupulous care, much detailed knowledge and attention to the minute geologic circumstances of the individual deposit are necessities for reliable dates. They never come easy. James Aronson had to change by nearly a million years his date for the key ba-

salt layer that fixed Lucy's age. Newer microtechniques are on the way that will assign every mineral grain to the particular volcanic eruption that flung it out; such deeper analysis will one day support the evolutionary conjectures with irrefutable evidence.

The result so far is striking. Lucy, about 3.5 million years old, is only the first and most complete find of an entire collection of kindred creatures from Hadar, where the well-stratified dry gulches seemed "to ooze fossils." Those walkers were less than four feet tall, fully bipedal, small-brained, without any tools that we can find. The first stone tools appear at Hadar a million years later, the oldest in the world, made not by Lucy's kind but by her descendants. Those mean *Homo habilis* to the present team, as they were named by Louis Leakey at Olduvai more than a decade ago. A second line of descent, however, leads from Lucy to the robust australopithecines. For a million years these bigger creatures walked side by side with *habilis,* but they left no direct posterity. We stem from *habilis* by way of *erectus* to *sapiens,* our own species only one or two tenths of a million years old.

The graphic evidence here is worth special mention. We see, for example, a set of hominid skeleton drawings with the bones actually known distinguished from those merely extrapolated. The anatomical illustrator Jay Matternes draws reconstructions of *Australopithecus afarensis* (Lucy's species) as in life and provides an interesting five-page justification of his decision, based on analogy as well as on the skeletal remains he works with. An air of openness blows through these pages, as much in the findings as in the deliberately intimate biography; the result is both absorbing and informing, a first-rate book.

THE PSYCHOLOGY OF THE PSYCHIC, by David Marks and Richard Kammann. Foreword by Martin Gardner. Prometheus Books, Buffalo, N.Y. 14215 ($15.95). "Skeptical books," writes Martin Gardner, "rarely sell." This straightforward study by two psychologists at the University of Otago in New Zealand should attract at least "a small but wiser minority."

In the mid-1970's the occult wonders of the media world, the Amazing Kreskin and the superpsychic Uri Geller, carried their powerful claims to New Zealand. About half of this brief and direct volume is a careful account, in the requisite step-by-step detail, of what the activities of these sorcerers were like. We attend a performance by Kreskin, 14 distinct events of mind reading, ending with Kreskin finding his own hidden paycheck for the performance. Here a reviewer needs a delicate touch. Kreskin, although he "encourages the association of the language of parapsychology and the occult with his name," has remained an entertainer. He has never asserted an alliance with galactic rulers or any other patent unworldliness. It has long been the practice of the profession of magicians to honor the ingenious methods of such a man by avoiding public explication. The authors, who came honestly by their conclusions through their own experience, do not withhold what they found. They give a plausible account of five distinct methods this engaging personage of the theater used for his effects. "A mix of tricks makes a bundle of sticks," stronger than any single method can be. This adage of magical stagecraft applies well to a Kreskin performance, dazzling in its impact on an audience all too ready to believe. These paragraphs will disclose no more.

Uri Geller is quite another story. He does not hesitate to invoke the wizards on the spaceship *Spectra* as the patrons of his transcendent powers. Once again

we read a detailed account of his visit and performances in New Zealand and an explication of the effects. Uri can be seen to peek at the blackboard past his shielding hand, and under or around a blindfold. "Yes, folks, it is really that simple." The authors saw it. The starting of defunct watches is now banal as a Geller phenomenon, particularly on television. It has astonished more than one national audience. The dry facts are here. Geller, jewelers, untrained students and people instructed over the telephone by the authors all have a similar success rate: about half of their old watches will go. Any watch held for a few minutes is warmed by the hand; a little shaking is likely to start it ticking once its oil is thinned by the rise in temperature. Geller and student controls alike can often reproduce drawings sealed within an envelope, but if the paper has been folded, the form they reproduce is apt to exhibit the folded pattern, far from the original intention. Envelopes are simply not opaque.

So it goes. The authors undertook remote-viewing experiments: the telepathic identification by someone who stays home of a random scene visited by someone else. This was the topic of a 1977 volume by the Stanford Research Institute team of Russell Targ and Harold Puthoff. The Otago results seem much the same, but they are plainly dependent on wishful thinking and vague measures. "The fact is that any description can be made to match any target." Unwitting cues, biases, ambiguity and selection are heavy claimants on belief.

Personality readings, the predictions of Nostradamus and the strange coincidences admired by Arthur Koestler are other topics treated in the cool, direct, almost ingenuous style of the authors. Not much credence is left. It is wonderful how much Nostradamus' prophetic vision improves once *la dame* is opportunistically translated as "cathedral." A table gives some results of number guessing, demonstrating population stereotypes employed by the magicians. Asked to choose a number from 1 to 10 but not to choose 3, 63 people out of 197 chose 7. It is as hard to be random as it is to cheat; real effort can achieve either. One Roman epigraph says it all quite clearly: *Populus vult decipi.* People want to be deceived.

THE GEOGRAPHY OF FAMINE, by William A. Dando. Halsted Press, a division of John Wiley & Sons, Inc. ($27.95). North Dakota, at the center of the continent of North America, is a land devoted to the bounty of the wheat. It nonetheless seems appropriate that this modest book, an introductory overview of famine seen in time and space, should originate with a professor of geography at the state university there. Men and women have been hungry in hard seasons ever since our species appeared on the earth, but famine is a characteristic of settled agriculture; the diverse dependence of the fewer, more mobile people of hunter-gatherer cultures on their land did not show the acute peaks of epidemic starvation we call famine.

Professor Dando struggles hard for a definition of his topic. His method, which he uses throughout the book, is to cite and consider a wide range of earlier opinion, in this case two dozen authors all the way from Thomas Malthus to the Food and Agriculture Organization and a late Russian study. He then offers a version of his own: "A protracted total shortage of food in a restricted geographical area, causing widespread disease and death from starvation." A drought is not a famine, pressure of population is not a famine, even a crop failure is not in itself a famine; if we take evidence from history, famine is "a social problem, a manmade unnatural phenomenon."

The text proceeds in four parts. First there is an engaging and compact summary of the history and geography of the human bond with the earth through our crops and our husbandry: the 10 staple crops, the centers of domestication, the factors of yield. Then we read of famines of the past in general, the chapter rather too briefly proposing a famine typology. Dando's data include 800 famines over 6,000 years of history; the world map is marked with the regions and periods he finds characteristic. The largest of these is also the most recent: the Asian zone (chiefly Russia, the Indian subcontinent and China), with the 19th century the time of worst tragedy.

The third part is the richest in original information, mainly in carefully documented lists of famines. It is a case study of three markedly differing regions. England and the U.K. harbored famine from Roman times up until the disgraceful Great Hunger visited on the Irish in the years before 1850. It seems probable that this event and a parallel disaster or two in British India represent the most significant impact of economic theory on society so far, short perhaps of revolution. In both cases there is evidence to show that important officials held it unwise and indeed immoral to fix prices, prohibit exports, import grain or distribute food to the starving lest they be "pauperized." The arguments of Adam Smith had been too persuasive within the Civil Service.

The second case study is that of the subcontinent of India, broadly defined. The tale is grim, with a change of nature in the 19th century. Before British rule there was famine somewhere in India every 40 years or so. The famines were crop failures in the face of the added burdens of war or beyond the reach of alleviating transport for one reason or another. All walks of life suffered save whoever could flee. Only about one famine in nine is traceable to strictly physical causes; the bulk were social in origin. Once the British ruled India, famines were more widespread but were socially limited to those who lived in poverty.

Famine in Russia was endemic. As that state expanded, region after region was visited by famine, again not generally following local drought. The famines seem to move southward over the centuries. "The suffering ... was in large measure manmade, i.e., the failure of man to accord relief." The last major patrol of this horseman across the Ukraine took place somewhat obscurely in 1946–47, the aftermath of war and invasion.

In Part IV, Dando addresses still larger issues: the balance of population and food, the policies and strategies of the future. Once again he marshals data, arguments and stances. The risk of famine from natural hazards alone is confined today mainly to isolated continental interiors far from transport and with low reserves. The world is no lifeboat: its countries are interdependent, of flexible carrying capacity and subject to a common fate. Nor is the world a battlefield; the triage of the casualties can at best be a prod.

There is a modicum of hope. In wide India "poverty and hunger exist, famines have been contained. The peoples ... have declared war on large-scale famine and won." Large imports by the states have forestalled hunger in the past decades in the U.S.S.R., in China and on the Indian subcontinent. The time of worldwide test, however, lies ahead; policy, not abstract estimates of carrying capacity, will fix who will eat and who will starve. The vast military investment of our period is not a good augury.

A valuable introduction to a central issue, this book should find many serious students. The editors have nodded quite a few times. Professor Dando, like many another specialist, has made some errors by citing not very credible sources in topics outside his main interest. There is no inscription along the Nile that dates to 4247 B.C., even if the magazines said so. Nor is one quite persuaded by reading, from an anonymous publicity brochure of the Bengal government in 1933, that Shah Jahan, the ruler who caused to be built the elegant domes of the Taj Mahal in mourning tribute to his queen, reigned in 1629–30 over a famine land, where "the number of the dead exceeded all computation.... Instead of seed men ate each other." Irony is certainly the rhetorical device best suited to economic history.

The facts of life.

Swimming's a barrel of fun.

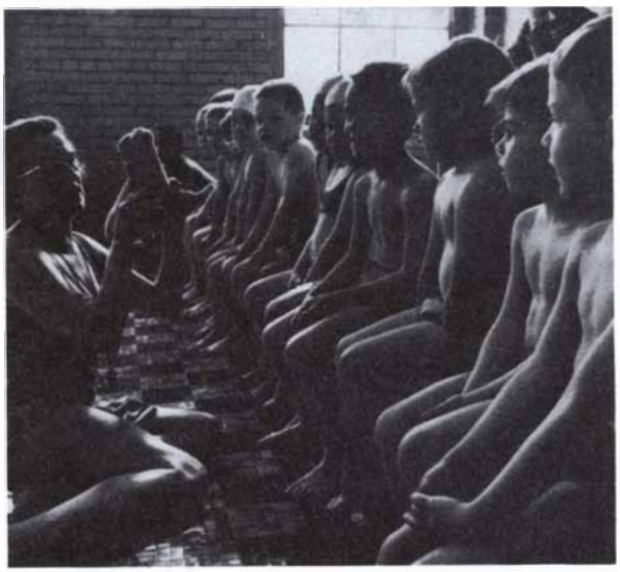
It can also be a barrel of danger, if you don't know the fundamentals—if you don't know how to take care of yourself in the water.

That's why Red Cross swimming, lifesaving and water safety courses are offered all over the country.

We want *everybody*—even little kids—to know The Facts of Life in the water. Since we began to teach swimming, we've issued a grand total of 55,216,000 swimming certificates. Millions upon millions of people have learned to swim with us. Or learned to swim better.

If you've ever wondered where Red Cross money goes, consider just this one fact: nearly every lifeguard serving in America is Red Cross trained and accredited.

Isn't Red Cross a good idea? Belong.



Red Cross.
The Good Neighbor.

FOLONARI®

"Folonari Soave. Clean, crisp...one of the best buys."

Terry Robards

Wine Critic, *THE NEW YORK TIMES* (9/24/80)

Dry, imported Folonari Soave is a totally delicious white wine. Its outstandingly clean, crisp, brilliant taste makes it a favorite — wherever and whenever it's served. More than a "best buy"—Folonari Soave is indeed a find.

Imported By "21" Brands Inc., N.Y., N.Y. © 1981

Dodge Mirada CMX

Dodge Mirada CMX. Chrysler engineered it to such high standards of performance, it was judged superior overall to the legendary Thunderbird—by 34 out of 50 T-Bird owners.

Great automobiles are not made of legends but of engineering.

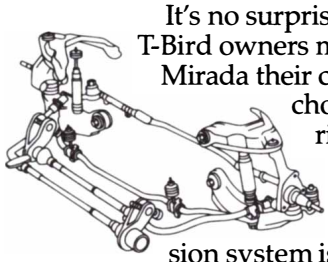
Mirada is the proud achievement of Chrysler engineers starting with concepts geared to today's world. New designs. New technology. And a determination to give people a personal car that spares nothing to return genuine pleasure to driving. Yet still delivers responsible fuel economy.

How well they succeeded shows in Mirada's rating of 18 EPA est. mpg.,* unsurpassed by a comparable Thunderbird; and 24 est. hwy. And in the results of an independently run test.

Mirada preferred to T-Bird by more than 2 to 1.

Fifty Thunderbird owners** judged a Mirada CMX against a comparably equipped T-Bird. They judged them for such qualities as riding comfort, driving ease, power, convenience, styling and value. Result: these Thunderbird owners preferred the Mirada overall by the overwhelming margin of 34 to 16.

Engineered to reward you with a superior ride and better control.

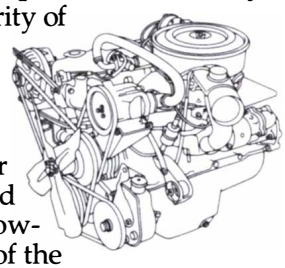


It's no surprise that T-Bird owners made Mirada their clear choice for riding comfort. The suspension system is unique. Transverse mounted torsion bar front springs—located ahead of the front wheels and rubber mounted — isolate the shock of bumps and pot holes from the car. You'll appreciate how little fatigue and tension you feel after a long day's drive.

Concern with human physiology is as much a part of Chrysler engineering as is the geometry of a car's suspension. Mirada's driver oriented interior provides bucket seats contoured to hold you comfortably for hour after hour behind the wheel.

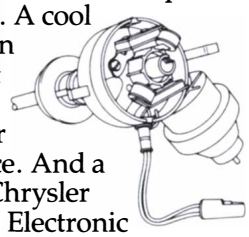
Mirada's overall driving ease and cornering ability, as well, were rated over Thunderbird's. Power assisted recirculating ball steering gives you sensitive control. You never lose the intimate "feel" of the road that epitomizes the well-mannered road car.

Engine technology that delivers more torque—with economy.



A majority of Thunderbird owners chose Mirada for pickup and passing power. Much of the credit must go to Chrysler's standard Slant-Six engine, a proven design with countless refinements in its lifetime.

Some of the latest: Hydraulic valve lifters that need no periodic adjustment. A cool air induction system that improves hot weather performance. And a landmark Chrysler innovation, Electronic Ignition, in which there are no points to wear or need replacing. And the consistently high voltage that results improves performance and reduces spark plug failures.



Chrysler engineers designed Mirada to take you with undiminished power and unimpaired fuel efficiency for 30,000 miles between tuneups.

Jury's choice: an instrument panel engineered for the driver.

T-Bird owners were outspoken in their preference for the efficiency and appearance of Mirada's interior. Behind the sporty 3-spoke steering wheel is your information center, tuned to the convenience of the driver. Gauges, lights, switches set in brushed aluminum applique: trip odometer, brake system warning light, speedometer.

Mirada's lowest price model—$8,083†—comes with equipment most Americans want.

Mirada's base sticker price includes as standard: Auto. trans. Power steering. Power brakes. Tinted glass. AM radio. WSW glass belted radials. Anti-corrosion protection.

The Mirada CMX shown, priced at $9,277,† also includes: Color-keyed cabriolet vinyl roof. 15" steel belted radials. Forged aluminum road wheels. Cornering lights. Sill moldings and wheel lip moldings.

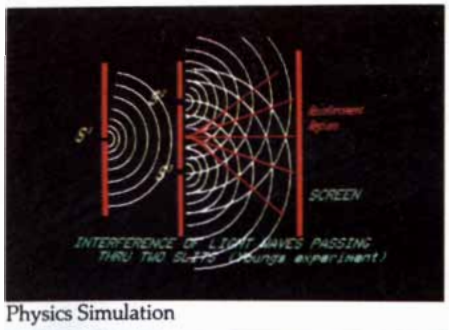
Other options:†† Air cond., $625. With auto. temp. control, $678. Auto. speed control, $149. Power bucket seat, $187. Choice of 6 sound systems, and a full complement of power assists.

Mirada, the personal car engineered without compromise. Buy or lease yours at a Dodge Dealer.

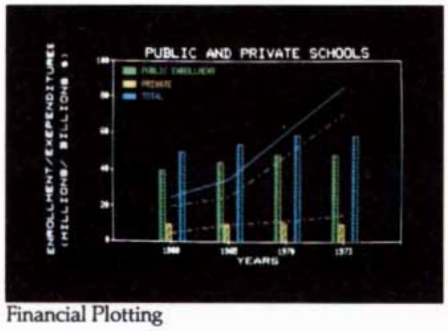
Dodge

The New Chrysler Corporation

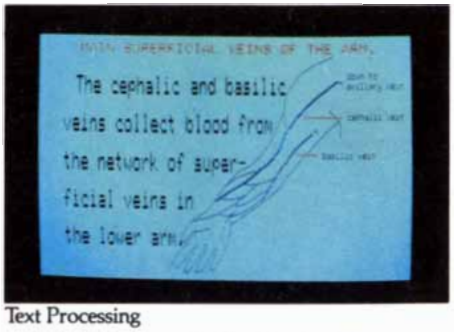
*Use EPA est. mpg. for comparison. Your mileage may vary depending on speed, trip length and weather. Actual hwy. mileage and CA. estimates probably lower. Comparison with Thunderbird based on both cars with standard 6-cylinder engines. **Randomly recruited in Los Angeles area.
†Sticker price excluding taxes and destination charges. ††These options may require the purchase of additional optional equipment.



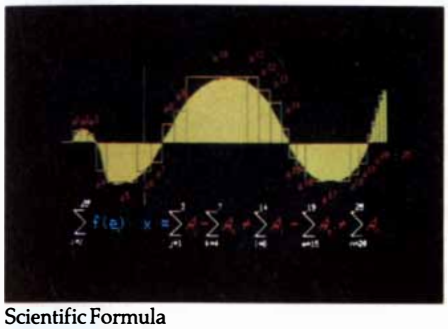
Physics Simulation



Financial Plotting

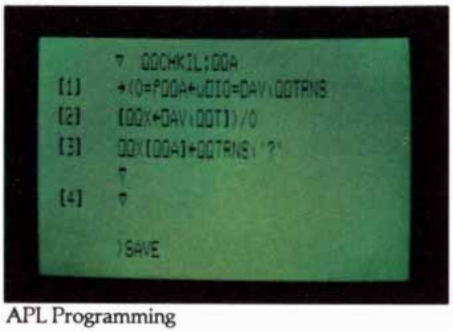


Text Processing

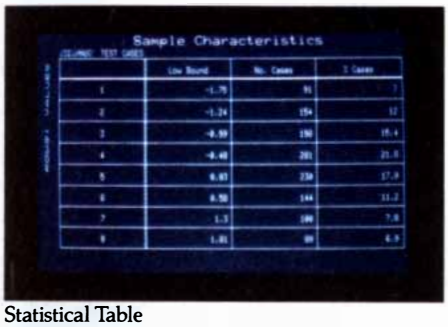


Scientific Formula

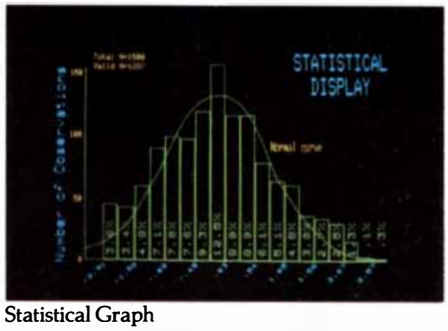
Special Needs Education



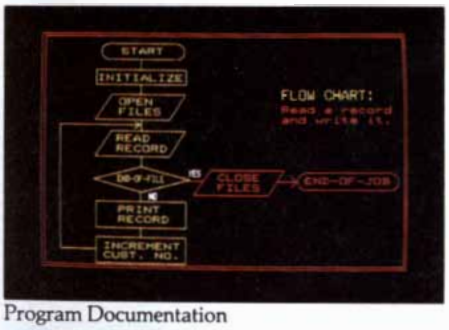
APL Programming



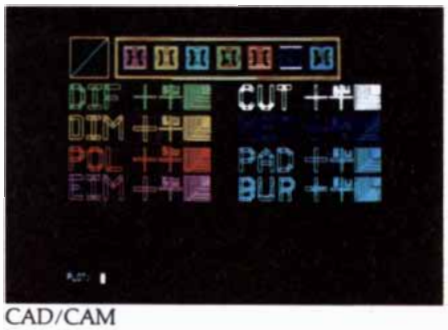
Statistical Table



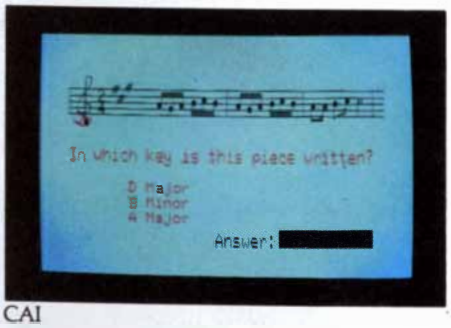
Statistical Graph



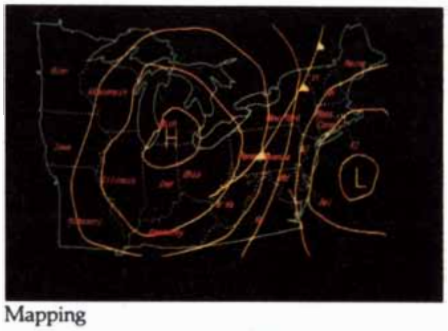
Program Documentation



CAD/CAM

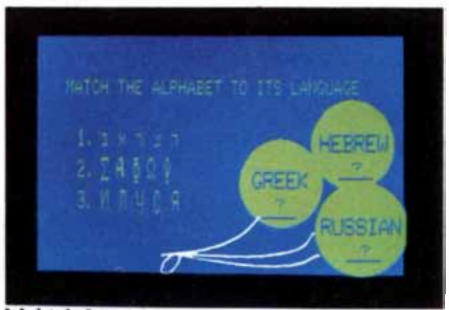


CAI



Mapping

Data Entry



Multiple Languages

Established 1845 SCIENTIFIC AMERICAN July 1981 Volume 245 Number 1

Values and Attitudes of the Polish People

Recent events in Poland, illuminated by 25 years of social research, are shown to stem not from a change in values but from a demand for social institutions more in accord with values consistently held

by Stefan Nowak

The prolonged social and political crisis in Poland has held world attention for almost a year. It began last July with a wave of strikes brought on by a sudden, unannounced increase in meat prices. Such strikes had occurred before in Poland, notably in 1970 and 1976, and for the same reason: a sudden increase in food prices. The strikes of last July and August, however, took a different turn. There were no street demonstrations and no riots or confrontations with the police. The strikers occupied the factories and conducted their demonstrations almost entirely on the factory premises. They were better organized than before, and this time they formulated more general demands. The demands did begin with higher wages to compensate for the higher prices. By the latter part of August, however, when the wave of strikes reached the Baltic shipyards in Gdańsk, Gdynia and Szczecin, the strikers were demanding social equality and social justice, the relaxation of constraints on speech and on the press, and freedom of organization for working people, expressed most forcefully by their demand for government recognition of independent trade unions.

The further course of events is well known. The government signed agreements with the workers recognizing the Independent Free Trade Unions, now called collectively Solidarity. In this confederation some 10 million people were organized within a few months. This was a genuine social avalanche.

Needless to say, such developments constitute a challenge for the sociologist. Even if one is not able to predict them in advance, one must try to understand them after they have happened. Here I propose to consider these events in light of a question: To what degree did the demands formulated by the striking workers (which were supported by the great majority of Poles organized in Solidarity or declaring their approval in public-opinion polls) express values existing in Polish society before the strikes began? If evidence for the earlier existence of such values can be found, further questions can be asked about the sources of the values and about the social mechanisms that shaped them.

It happens that there are ample sociological data bearing on such questions. Sociology has a long tradition in Poland, going back to the end of the 19th century, and Polish sociologists conducted many empirical studies before World War II and immediately after it. From 1951 to 1956, a time that we in Poland call our "Stalinist" period, the work was interrupted in the turmoil and repression attending the establishment of the new political and social order. When we were able to resume our studies, we concentrated on several problem areas, including the transformation of attitudes and values in Polish society.

The conclusions and generalizations of this article are based on more than 150 surveys conducted by Polish sociologists since 1956, among them six large studies by my colleagues and me. Many of the surveys put the same questions to similar samples of respondents over periods of as long as 20 years, allowing one to say something about the stability or the mutability of the values under study. If work of this kind can throw light on historical processes, there are plenty of data available.

Because the term "value" is used in the social sciences with various meanings, I shall first explain how I am going to use it here. By the values of a person I mean the standards that define for him how people should behave, what actions or events merit approval or condemnation and what pattern of relations should prevail among people, groups and institutions. By comparison with such standards the person makes evaluations and holds things, events or courses of action to be good or bad, just or unjust, proper or improper, desirable or undesirable. By his values, in other words, an individual defines what kind of person he wants to be and what kind of world he wants to live in, and he sets standards by which to judge or evaluate both himself and the world.

The person or thing or event subject to such an evaluative judgment can be called the object of a value. Values may be understood to prescribe the "shape" or nature of an object in qualitative terms. Values also sometimes specify in a quasi-quantitative way the "amount" or the intensity of a property an object should possess. Some values may be perceived or felt as normative, and so any deviation from them evokes guilt or condemnation. Values of this kind are often associated with a person's religion, moral philosophy or social ideology. Other values have the character of private wishes and preferences; objects that fail to measure up to such standards induce frustration or dislike rather than guilt or condemnation.

When an object is compared with its standard, the sign and the strength of the resulting evaluation depend on the dis-

Which of the following features should a good social system have (+) or not have (−)? Choose no more than five positive features and five negative ones.		"GENERATION GAP" STUDY				WARSAW STUDENTS (1978)
		WARSAW (1973)		KIELCE (1972)		
		PARENTS	YOUTH	PARENTS	YOUTH	
a. Equality of life opportunities regardless of social origin	+	77	79	71	84	83
	0					
	−	1	2	1	1	1
b. Assurance of a proper standard of living for all citizens	+	61	58	62	62	54
	0					
	−	1	2	1	2	1
c. Freedom of speech; conditions under which differing views can be freely expressed	+	46	57	39	59	59
	0					
	−	2	3	3	3	2
d. Influence of all citizens on the way the society is governed	+	38	49	35	47	56
	0					
	−	7	6	5	11	7
e. Economic efficiency	+	43	44	34	35	50
	0					
	−	1	0	1	1	1
f. Approximately equal incomes for all citizens	+	20	25	32	22	12
	0					
	−	28	20	17	15	35
g. Obedience of citizens to the decisions of authorities	+	22	11	32	14	8
	0					
	−	5	14	4	12	13
h. Nationalized industry	+	29	33	28	46	37
	0					
	−	5	7	2	4	3
i. A large measure of independence for various experts and specialists	+	30	22	24	15	14
	0					
	−	4	11	6	21	12
j. Strong apparatus of central power deciding all important issues	+	15	12	19	18	8
	0					
	−	28	36	22	28	41
k. Equality of life opportunities regardless of the views or opinions held by an individual	+	25	35	19	27	42
	0					
	−	5	4	4	7	3
l. Conditions allowing the exercise of influence and initiative by all citizens	+	23	23	18	11	32
	0					
	−	3	12	5	22	7
m. Fairly strongly differentiated incomes, depending on an individual's qualifications	+	21	13	17	10	17
	0					
	−	22	40	29	50	32
n. Broad development of various forms of self-government	+	9	8	10	5	15
	0					
	−	9	22	8	32	10
o. Limitation of freedom of action for adversaries of the political system	+	5	8	6	9	3
	0					
	−	30	33	24	34	44
p. Absence of greater differentiation of views and opinions among citizens	+	3	6	4	6	5
	0					
	−	24	38	16	30	40

tance between the standard and the perceived property of the object. Toward most of the objects around them people address an entire set of evaluation standards; these constitute a multidimensional evaluation space in which the standard addressed to each aspect of the object places it at a certain distance on one of the several evaluative dimensions. In the total evaluation of the object, however, people accord more importance to some standards than to others, with relative importance acting as a psychological multiplier of the distance of the object from a given standard. A set of standards addressed to a single object or to a class of objects therefore tends to be a more or less consistent hierarchical subsystem in the entire value system of the individual.

The hierarchical clusters of values addressed to different classes of objects are themselves usually hierarchically ordered. Their order is determined primarily by the perceived relative importance of the objects to which they are addressed. For example, I may agree that democratic structure of political institutions is important for any country, but the psychological multiplier applied to this standard is much larger when I am thinking about my own country than it is when I am thinking about, say, Imperial Rome. Health is an important standard in evaluating the condition of any human being, but it has compelling force when the object in question is my own child. The stratification of objects by their relative importance is the basic dimension of our value systems.

Values may motivate action if evaluation of an object leads to the conclusion that action is needed to sustain the proper shape of the object, to improve it or to eliminate it from the field of our experience. Action, however, calls for considerations of another kind. In order to decide to act, we must believe there is some probability of success, some chance of moving the object closer in its evaluation space to the state corresponding to our likings or normative standards. To a large extent the perceived probability necessarily depends on external constraints imposed on our freedom of action by the social system in which we live and by our place in it.

When the society or the condition we live in makes improbable the success of an action indicated by our values, the values may become analogous to recessive genes, which exist but do not act. The values are embraced, but they fail to motivate corresponding behavior. It may well be asked whether such "behavioral recessive" values, describing merely what people might like to have or what they are inclined to dream about, are a legitimate topic of sociological investigation. I would venture the opinion that they present significant questions for research. They do exist in people's minds, are important to them and qualify their satisfaction with their own lives and with the world around them. Moreover, they are worthy of study because, as with latent genes under suitable conditions, latent values can become powerful factors in human behavior, acting on the scale of the individual or of the society. Contemporary Poland seems to present itself as a case in point.

When my colleagues and I undertook our first surveys of attitudes and values, in 1957 and 1958, we produced an astonishing finding with respect to the internal structure of the observed system of values. What we found can be called an "almost random statistical aggregate of values and attitudes."

In the first place, when we measured on a quantitative or ranking scale the degree to which people held a value, we quite often found that the distribution of the scores in the population tended to approach what is called in statistics the normal distribution. If a bimodal distribution can be taken as indicating polarization of views and values in a society, the normal distribution indicates that the more extreme a given view is, the less common it is.

Our data showed, further, only weak correlation between values and the demographic characteristics of the respondents. (Distinction by education and its correlates provides the most visible exception to this generalization.) For the most part the set of values found in a given group, even if the group differs on a given issue from other groups, must be taken as a modification of the generally prevailing pattern rather than a special feature characterizing the group. It is possible, in sum, to speak about a system of values in Poland rather than about systems.

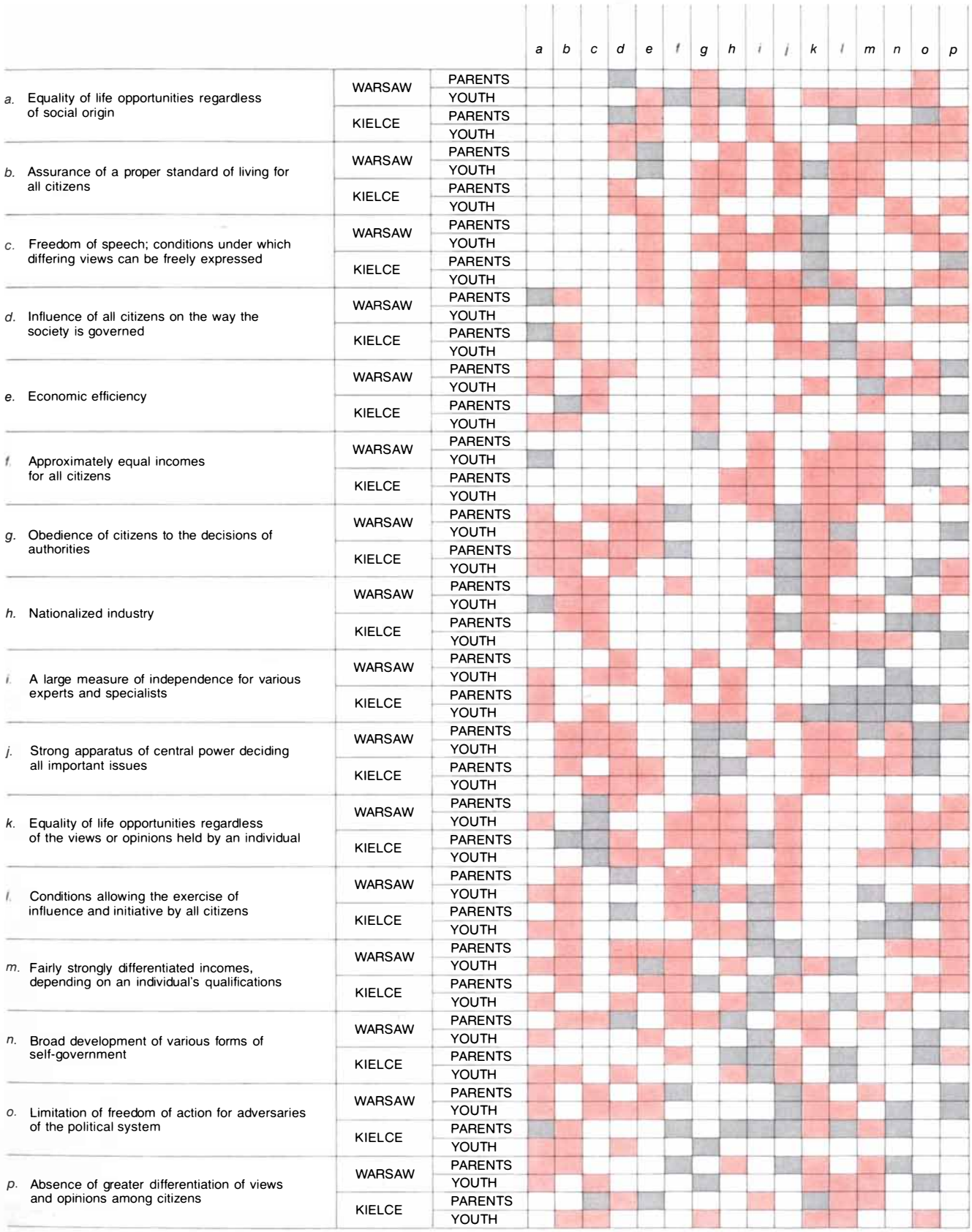
Finally, we often found weak or zero correlation among the various aspects and dimensions in the value system even of an individual respondent. Thus the respondents hold sets of values with little internal consistency; there is no ideological syndrome of views. This has an important social consequence: if two Poles, on finding that their views differ strongly on one issue, continue their conversation long enough, they will discover they have similar views on many other issues. Rarely would their encounter array them in confrontation over differing *Weltanschauungen*.

The finding of such unstructured aggregates gives one the sense that the respondents had somehow absorbed each value and attitude separately and at random. The impression that we were dealing with the almost random absorption of freely circulating values was strengthened by studies of the values of parents and children that we did in the early 1970's. We could not find a generation gap, nor could we find significant correlation between the values of parents and those of their children (with the clear exception of religion, where intrafamily correlations were strong). In people's absorption of values their personal and group identity seemed to play at most the role of weak filters, not of independent factors shaping the values.

This does not seem to be the usual state of affairs. Although we do not have survey results and comparable systematic data for Polish society before World War II, we can make inferences from the programs of political parties and from electoral statistics. From such information we can infer that the social and political views in the society covered the entire spectrum from the extreme left to the extreme right. We can also suppose the views held by people corresponded in a syndromatic way, although very likely with less internal consistency, to the programs put forward by their political parties. Furthermore, we may accept what is a rather banal assumption of political sociology, namely that various social classes and groups were distinguished by different sets of social and political ideas. This was probably true also for the period immediately after the war.

To explain the transformation from such a hypothetical value system to the one we found in the late 1950's it is necessary to bear in mind the salient features of the country's social and political transformation in the intervening years. Aiming at the construction of a socialist society, the new political leadership nationalized almost the entire economy and eliminated the classes of owners of large and small factories and, later on, even those of small shops. Land reform eliminated large private estates, and the first steps toward collectivization went forward until in 1956 they were halted and reversed. As a result social mobility, both horizontal and vertical, increased to an unprecedented degree. It was horizontal because changes in national borders relocated millions of people from east to west. It was vertical because of heavy losses of intelligentsia during the war and, after the war, be-

SYSTEM OF VALUES prevalent in Polish society has been documented over the past 25 years by means of questionnaires and interviews. Here attributes of a social system are rated as desirable (+) or undesirable (−) by high school students and their parents in the cities of Kielce and Warsaw and by students at the institutions of higher education in Warsaw. One notable conclusion drawn from the data is that there are not deep or systematic divisions in Polish society over issues of this kind; for example, there is little evidence of a "generation gap." Hence it is reasonable to speak of a single system of values in Poland rather than a multiple system.



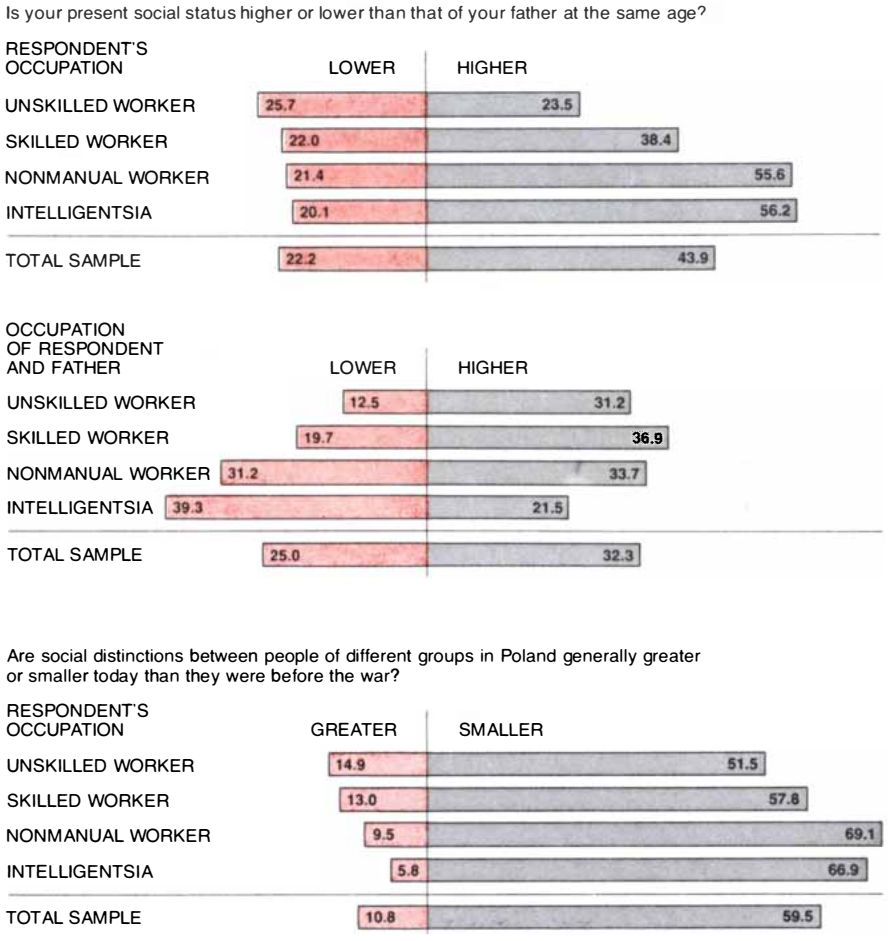
ALMOST RANDOM STATISTICAL AGGREGATE of values is suggested by the weakness of associations between responses in the generation-gap study. A consistent pattern of positive associations (*gray squares*) or negative associations (*colored squares*) would indicate that many respondents subscribe to a comprehensive ideology in which several values are closely correlated. For example, consider the case of a respondent who is known to have chosen item *c* (freedom of speech); the matrix of correlations indicates there is an enhanced probability that he will also choose item *k* (equality of opportunity regardless of opinions) and a reduced probability that he will choose item *e* (economic efficiency). Correlations that extend to all groups in the study are infrequent, however, and more than half of the possible pairs of reponses show no significant correlation (*open squares*). The study was done by the author in cooperation with A. Banaszkiewicz, M. Gadomska, J. Karpiński, K. Kosela, E. Kolbowska, M. Marody, A. Sułek, T. Szawiel, M. Styczeń and E. Tkaczyk.

cause of industrialization, urbanization and the growth of bureaucracy.

All these developments created larger numbers of places in the higher layers of the new social structure, and so upward mobility had clear dominance over downward mobility, particularly in the early years. People's perceptions strongly reflected this historical process. The majority were convinced that differences between social strata were smaller than they were before the war, and most people felt their social position was higher than the position of their father in prewar Poland. The feeling of social promotion was reported not only by those whose occupational status was higher than that of their parents but also by those who had not moved up the social ladder. This testified to the upward movement of entire social strata.

The new authorities steered the upward movement, taking into account the social origins of candidates for promotion and preferentially advancing people from lower classes. The authorities also gave weight to political loyalty as a criterion for promotion in the new power structure. By these processes the society was stirred from bottom to top.

To all of this, in explanation of the homogenization of values and attitudes in Poland, must be added the political transformation of the society. First came the elimination of all parties and organizations that did not identify themselves with the "Democratic Bloc" headed by the Polish Workers Party. Then, after 1948 and the merger of the socialist and communist parties to form the Polish United Workers Party, came the tightening of the system of political control over the behavior and minds of citizens. Nonconforming ideologies or philosophies (with the exception of Catholicism, which remained confined to the churches) were eliminated from public life. The new regime, with its monopoly on mass communications, undertook the indoctrination of the people with Marxist ideology.



SOCIAL MOBILITY in postwar Poland is perceived by most respondents as having improved their status. A plurality of the respondents in four occupational categories consider their social standing to be higher than that of their fathers (*top chart*). Even among respondents who hold jobs in the same category as their fathers, most report an improvement in status (*middle chart*); the only group that disagrees with this evaluation is the intelligentsia, defined for purposes of the survey as professionals and others with higher education. Complementing the widespread sense of upward mobility is an even stronger impression that the social distance between groups has become smaller (*bottom chart*). The survey was first done by the author in 1961 with a sample of about 2,000 urban men. In 1975 the survey was repeated by Elizbieta Otawoka.

Americans employ the metaphor of the melting pot to characterize the mixing and exchange of values from various ethnic sources into one homogeneous pattern. For the Stalinist period of Polish history a more appropriate metaphor is that of the grinding mill. From that mill, I think, came the "almost random statistical aggregate of values and attitudes" we found after 1956.

With respect to the content of people's values, as distinguished from the formal, statistical structure of the value system, our surveys in the late 1950's and early 1960's showed that "the experiment in social learning on a national scale" conducted by the new regime had succeeded to a certain degree. The great changes in the social and economic organization of the society—the nationalization of industry, land reform, economic planning, the abolition of the prewar class structure—were accepted by the people. In reflection of the strong propagation of egalitarian ideology in those formative years the people also embraced the idea of equality of life opportunity for all citizens and even the idea of preferential opportunity for the underdog. Their egalitarianism was more moderate when it came to the distribution of income; the majority accepted some inequality. What the people held to be a morally acceptable spread between the top and the bottom, however, would seem fairly radical by the standards of Western Europe. Indeed, the majority held that there was not enough equality in the social stratification of socialist Poland. Most people also shared in the fairly strong conviction that society as a whole and the state in particular are responsible for the equalization of life opportunities and for the development of the potential of all citizens as well as for the satisfaction of people's basic needs.

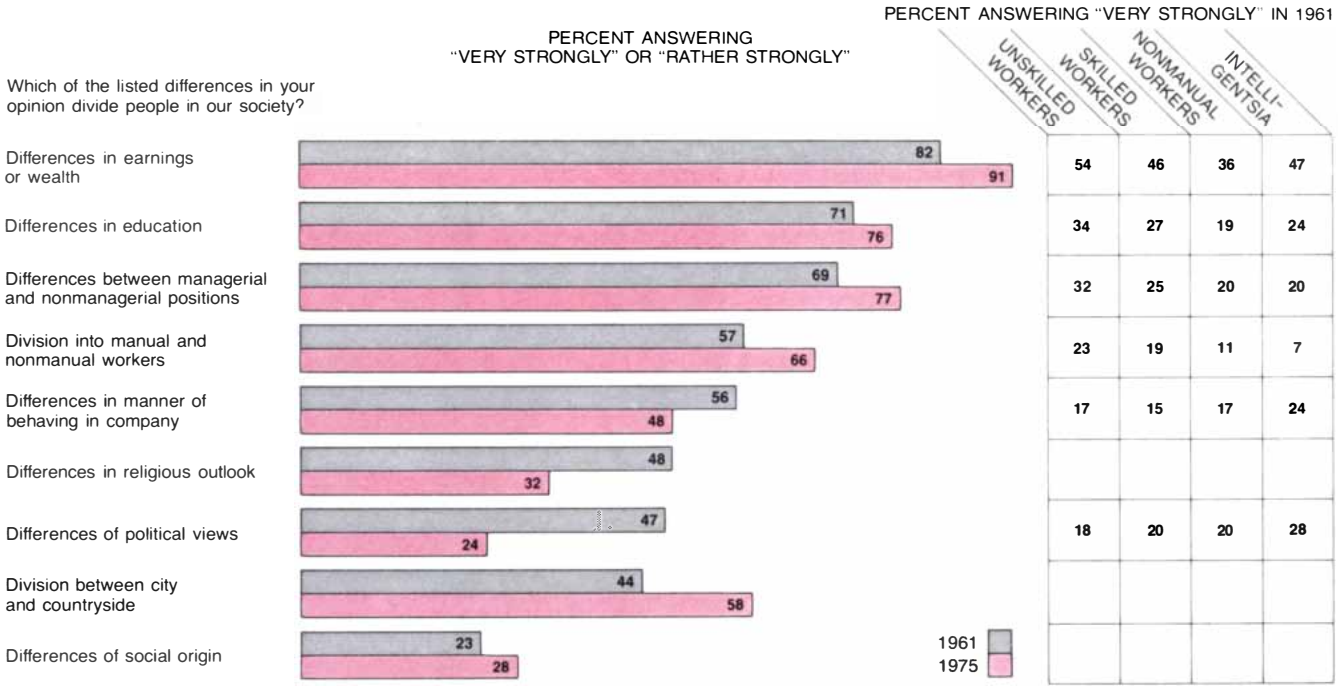
These were the ideas people associated with "socialism" in the first years after 1956, when Wladyslaw Gomulka came to power. Along with the new party leadership, people would stress that they were in favor of "the Polish road to socialism." This meant democratizing the political system, removing the most drastic limitations on citizens' rights and the major sources of their fear, promoting freedom of speech and expression, increasing the influence of people on the government and, after the Yugoslavian model, encouraging the participation of workers' councils in the management of factories. "The Polish road to socialism" was also suffused with the patriotic feelings people attach to nation as distinguished from state.

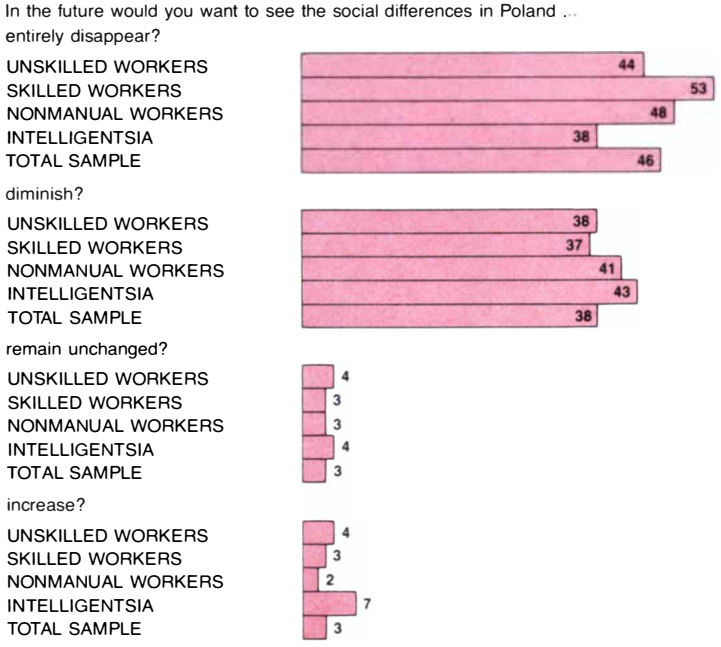
In supporting this kind of socialism the majority of the people did not identify their ideas with Marxism; only 2 percent of the university students we surveyed in Warsaw in 1958 declared themselves "definite Marxists" and only 11 percent called themselves "moderate

Marxists." At the same time our studies showed the persistence of religious sentiments. Gomulka was confronted with evidence of this a few days after he came to power. Thousands of people assembled before the building of the Central Committee of the Party to demand the immediate release of Cardinal Wyszynski, who had been held in seclusion since 1953. On the other hand, the traditional religiosity of the Poles, in which the stress is on ritual rather than on spiritual experience or moral issues, coexists quite comfortably in the prevailing random pattern of association with most other values. Being religious scarcely changes people's acceptance of a nationalized economy, of social equality or of "some form of socialism."

In general one can say that the value system we found in the late 1950's had formed in the interaction of values propagated by the new system and values that persisted somewhere deep in people's minds. The synthesis began to exert pressure on the course of events on the national level in about 1956.

Over the next 25 years Poland had an eventful history. The austerity imposed by Gomulka and the lavish investment program of the administration of Edward Gierek terminated, in 1970 and 1980 respectively, in economic crises that brought workers out on strike. Poland had student protests and demonstrations in 1968 and a wave of political protest by students and intellectuals in the 1970's. There were periods of rela-





RANKING OF DIVISIVE INFLUENCES in Polish society provides evidence that certain egalitarian ideals have long been common but have not been satisfied. In 1961 differences in earnings or wealth were cited by more than 80 percent of the respondents as a divisive influence; by 1975 the proportion had increased to more than 90 percent. Other factors that bear directly on social standing, such as education and kind of work, were also selected more frequently in the later survey. Meanwhile differences in personal behavior, in religious outlook and in political views diminished in importance. Responses to another question in the 1961 survey indicate that most people think social distinctions are too great, and a sizable number express the view that such distinctions should disappear entirely.

tive political freedom when new leadership was coming into power, and later there was a tightening of political control. After 1976 came a new phenomenon: political-opposition groups acting openly and a periodical and book press operating outside of censorship.

The social structure meanwhile went through a marked change toward polarization of the economic conditions in which people lived. There came an increase in the number of those who, in an egalitarian socialist society, were "more equal than others." Particularly in the last years of the Gierek regime, a newly developing style of conspicuous consumption advertised the presence in the society of people who could afford it.

In view of all this history, what changes did our later studies find in the value system of our countrymen? The answer is simple: Almost none at all! In surveying comparable samples of the population a few years or many years later, whenever we repeated a set of questions dealing with social values, we obtained substantially the same distribution of responses. Because the relation between values and group characteristics remained weak, even the character of the compared samples did not matter much; the distribution of responses to questions about values was almost identical in most studies.

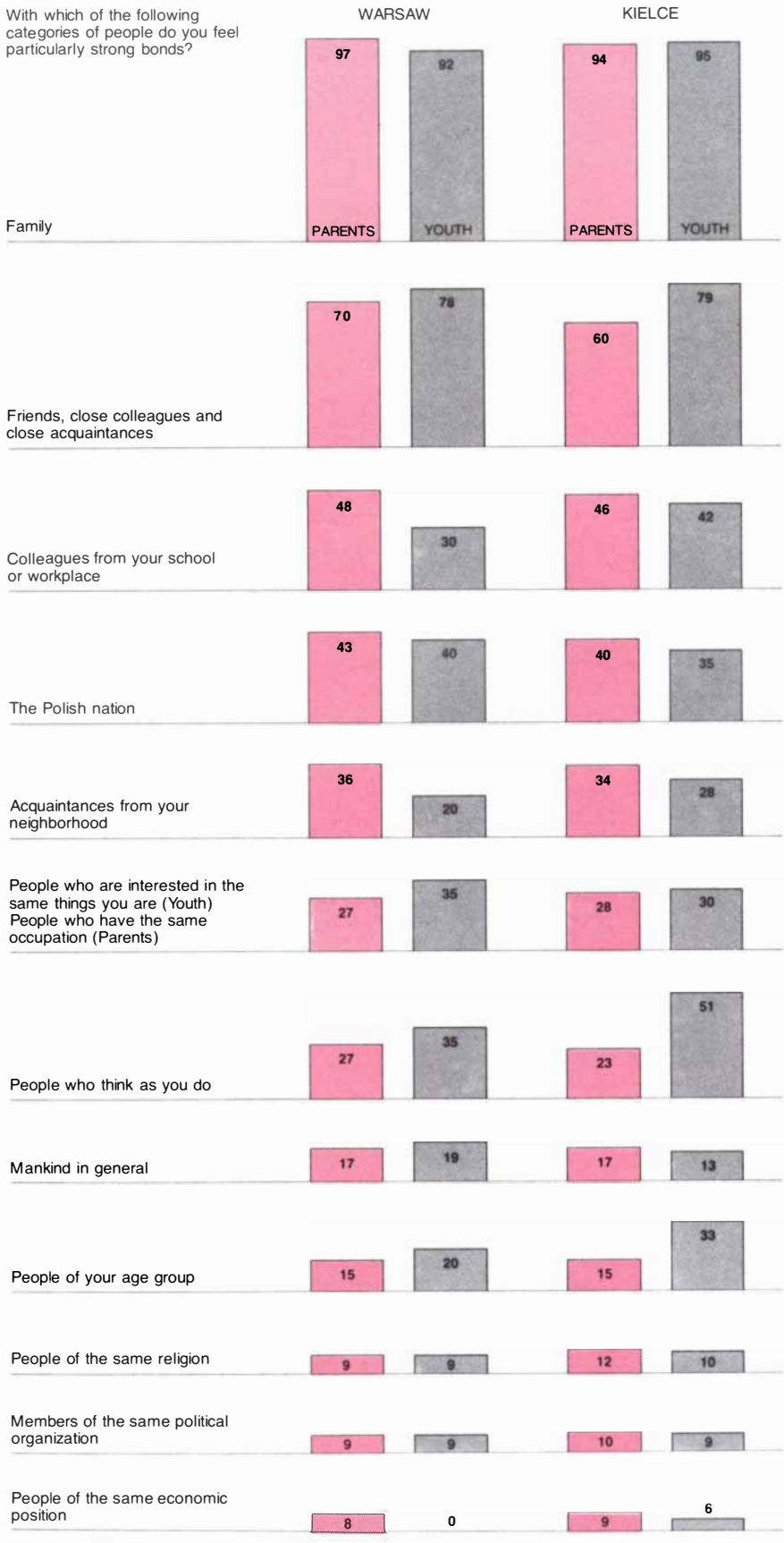
In 1978 we repeated the study we had made in 1958 of the social and political values and attitudes of a random sample of students at the university level in Warsaw. Comparison of the responses to the same questions 20 years apart shows almost no significant changes. Ideologically the 1978 students were almost identical with their 1958 counterparts, some of whom by then held positions in the government, the economy, the culture and the science of Poland.

ENSEMBLE OF VALUES held by the Polish people cannot be understood as having stemmed from a single ideological source. These data are for students in Warsaw in 1958 (*gray*) and 1978 (*color*). Most students favor "some form of socialism" and support policies associated with socialism, such as limitation of the highest incomes. On the other hand, few identify themselves as Marxists and many are religious. Private enterprise is considered appropriate in some areas of the national economy but not in the more essential ones. Opinions on most such matters have remained stable. The set of questions has been presented to Warsaw students three times: in 1958 by Zófia Józefowicz, the author and Anna Pawełczyńska; in 1961 by the author, and in 1978 by the author with Grzegorz Lindenberg, Paweł Karpowicz and Krzysztof Nowak.

Impressed by the stability of the value system of our countrymen over those two decades, we were moved to look at the basic dimension of all value systems: the hierarchical order in which people rank the objects of their values. Given that Poles are like other people, it may be assumed that the self stands first in this order. Our studies showed that the family ranks next highest with practically everyone. In all social groups and generations this institution can be regarded as a kind of extension of the self. Next after the family in our surveys came friends, from close friends to the extended circle of classmates at school and colleagues at the work place. Surprisingly, we found that the next most important group to a Pole is the Polish nation as a whole. This is in line with the stereotype of Polish patriotism, which proves to be more nearly valid than might have been expected.

The social structure of Polish society in the subjective vision of its members would therefore appear to be a "federation" of primary groups united in a national community. In actuality, of course, the objective social structure of Poland is as complex as that of other industrial societies. It has a multilevel system of national administration, quasi-corporate industrial enterprises, political organizations, professional associations, official state trade unions and so on. None of these institutions, however, seems to evoke on a mass scale the kind of identification that comes out most plainly in interviews when the respondent says "we." Instead many of the institutions set up feelings expressed by the word "they," even among people who work in them or belong to them. Our studies thus revealed a kind of social vacuum between the level of primary groups and that of the nation.

Further, our studies showed quite strong changes in people's evaluation of



HIERARCHY OF SOCIAL TIES represents a basic dimension of any value system. Family comes first in Poland, as it does elsewhere. Apart from family, respondents in the generation-gap study of 1972–74 reported strong bonds to friends and colleagues and to the Polish nation as a whole. There was a notable lack of cohesive social institutions at a scale between that of the circle of friends and that of the nation. This gap seems to have been filled in the past year by the spontaneous development of the Independent Free Trade Unions, called Solidarity.

objects to which their otherwise stable value systems were "addressed," particularly changes in their satisfaction with the conditions of their own lives and with the performance of the country's leadership. Their dissatisfaction has been aggravated by shortages of consumer goods in the past few years. Combined with a housing shortage that has kept many from having a flat of their own, these circumstances have brought people to an emotionally charged, negative evaluation of their life situation. The contrasting example of conspicuous consumption by those who can afford such consumption has given poignance to the frustration of the others, inviting their condemnation of the situation as being contrary to socialist standards of social equality. Studies we did in 1961 and 1978 showed that disparities in income and wealth were perceived by a great majority as the main factor dividing people and as a primary source of social tension.

In this connection, however, we found that people's attitudes toward economic equality were strongly related to their own economic position. Those who had a high income and privileges had a much greater understanding of the necessity for differentiation in income and privileges than those who did not. This explains our finding, mentioned earlier, that educated people are less egalitarian in their views, because income in Polish society is increasingly related to education. Educated and well-off people are able to reconcile this attitude with their "prosocialistic" views by saying that the realization of a more egalitarian society is the business of future generations. Some rejected this ideal even for the future, but they did not see any contradiction between that attitude and their acceptance of socialism. On the other hand, the same people showed a strong attachment to various democratic values. Thus the higher up the social ladder the respondents were, the more democratic and the less egalitarian they were.

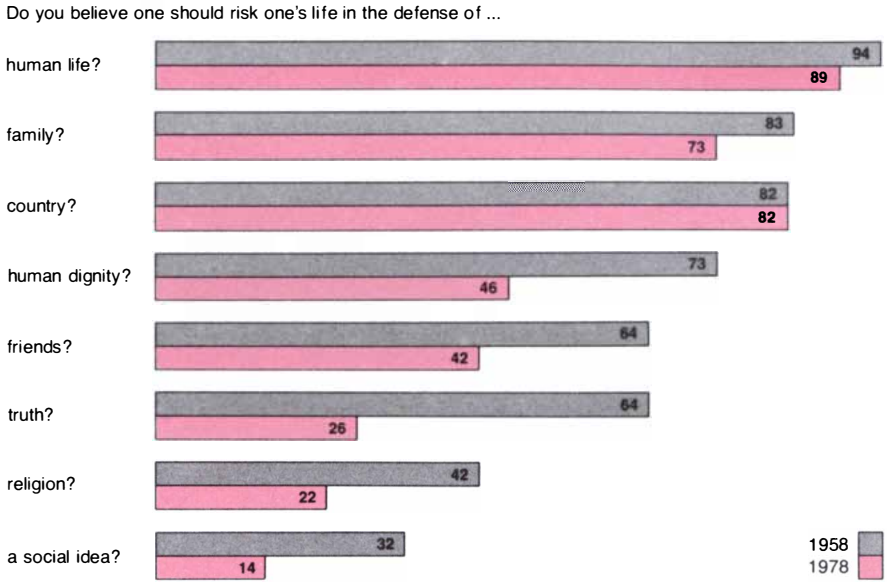
One important democratic value, freedom of speech, proved to be more highly prized among workers, and particularly among highly skilled workers, than it was even among people with a university education. At the same time the highly skilled workers showed more egalitarian attitudes with respect to differentiation in income and privileges. I think this explains to some extent why skilled workers are now the main social force in Poland, demanding social and economic justice and freedom of speech at the same time.

Some other aspects of sociopsychological reality also changed between the 1950's and the 1970's. A mood of increasing irritation that one could almost feel in the air became widespread in the population, particularly in the last few years of the Gierek regime. The classic

coupling of frustration and aggression could be seen in people's punitiveness and intolerance of social deviance. Our studies also revealed diminished interest in social issues in general and increased interest in private and personal affairs, a "reprivatization" of the people's ideology attended by apathy and open cynicism. Cars and summer houses became the engrossing topics of conversation among intellectuals.

A striking expression of this drift in people's attitudes emerged from our surveys of university students in Warsaw. In 1958 one question elicited from them the opinion that an individual should be willing to risk his life for various causes: for human life, family, friends, truth, human dignity, motherland, religion and, to a lesser degree, social ideology. The willingness ranged from 94 percent (human life) to 32 percent (social ideology) and averaged 67 percent. After 20 years the percentages had gone down sharply, and the decline was most marked for values such as truth, human dignity, religion and social ideology. The percentage willing to risk all for human life and for family had also declined, but one value held its ground: the motherland, at 82 percent in both surveys.

Apart from this affirmation of patriotism, the apathy otherwise observed in Polish society was evident in this survey, and indeed it affected a large area of the value system of the students. With respect to the subjective importance of most extrapersonal values, the students of 1978 were a completely different population from those of 1958. Yet with respect to the kind of society they would like to have they were almost identical with their compatriots of 20 years earlier. It was simply that their aspirations and desires in that realm had become less intense and less important to them. Here, in the late 1970's, the Polish people were evincing political irritation and social apathy at the same time. The combination turned out a few years later to be explosive.



FRUSTRATION OF SOCIAL IDEALS could account for the changing pattern of responses to a question included in the 1958 and 1978 surveys of the values of university students in Warsaw. The proportion of the students who approved of the notion that one should risk one's life for various causes declined sharply over the 20-year period in all instances but one: defense of country. Since other survey results indicate little change in the values held by the students, the diminished willingness to risk all can be interpreted as a result of the students' frustration at the failure of the society to measure up to their values. Such frustration appears to have been an important driving force in the recent restructuring, or "reintegration," of Polish society.

The speed with which 10 million Poles assembled in the Solidarity movement testifies to the strength of the needs the movement serves. People perceive the new free trade unions as necessary instruments for the implementation of important social and individual goals. To regard these unions as instrumental associations only, however, would greatly disparage their true nature; they are, at least now, communities with their own self-fulfilling emotional value. The speed of their development testifies also to the degree of frustration, in the social vacuum of the preceding period, of the need felt by so many people to belong and to feel the support of others.

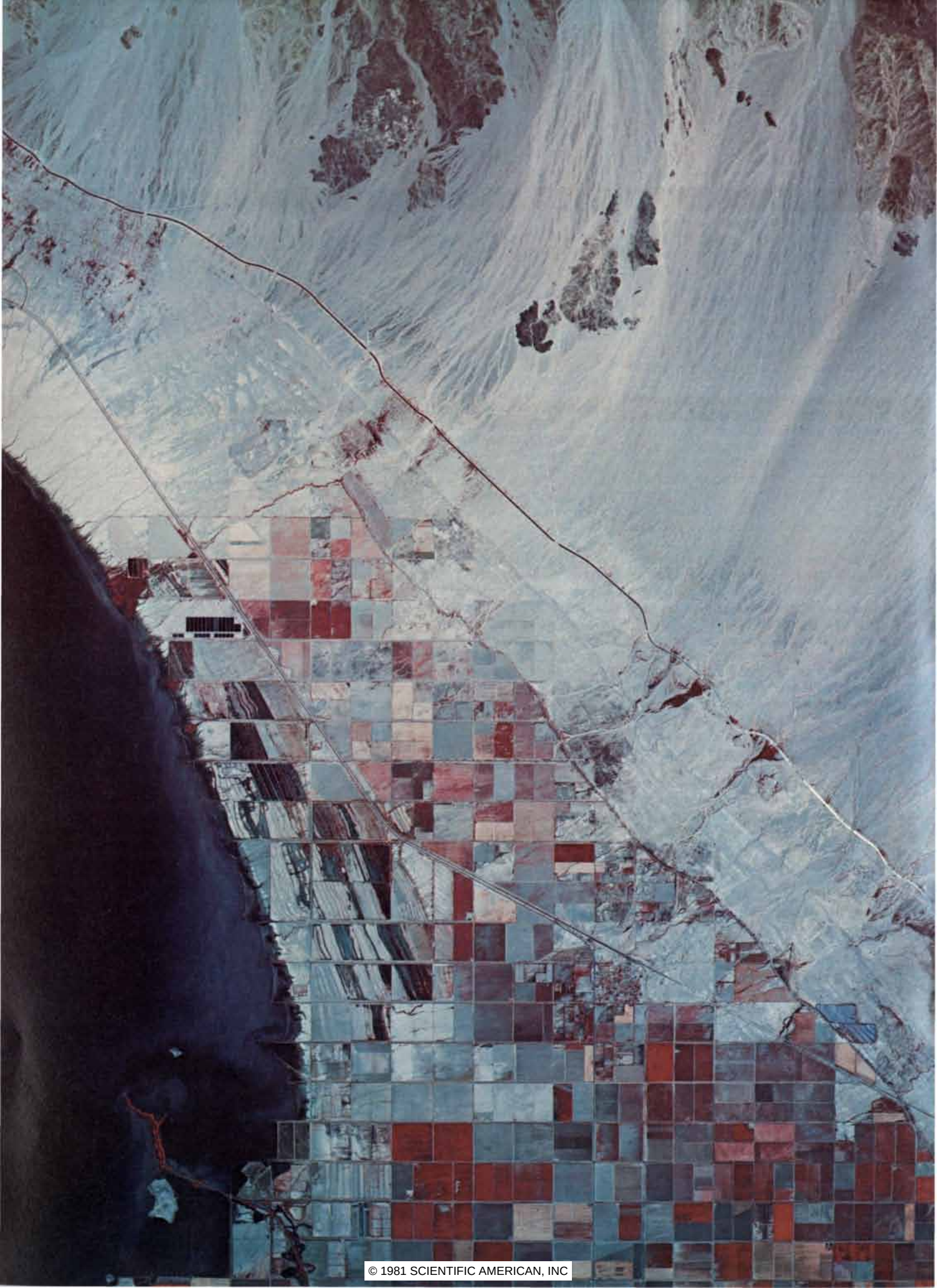
The process of the reintegration of society can also be observed in other social groups, such as among peasants and students. Many of the preexisting formal organizations are undergoing democratic rebirth. Many associations, for example, are electing their own boards and defining policies and tasks by democratic procedures without external control; in this way the members of such organizations are filling their own social vacuum with new social meanings.

By uniting and organizing, the people also lose what might be called the complex of Lilliputian v. giant. They can see that the newly organized unions and democratically run organizations give them the possibility of influencing things even at the national level. They lose the feeling of powerlessness and replace it with a feeling of dignity. There are important concepts that are not easily measured by batteries of indicators but that nonetheless are necessary for the proper understanding of some social situations. One cannot understand the events in Poland without reference to restored human dignity.

Observation of everyday life in Poland in the past few months indicates that much of the interpersonal irritation and aggressiveness has disappeared. People are being nice to one another even on the crowded tram at six o'clock in the morning or in the long lines in front of shops. If the elimination of the symptoms of frustration and aggression is confirmed by systematic surveys, it will be shown that essential needs of the people are being satisfied. Since the needs cannot be economic ones (in this respect the people are now more frustrated than ever), we must look for another explanation. In the end it may be found in the reintegration of our atomized society, in the elimination of the feeling of powerlessness and in the restoration of people's dignity.

It is possible now for people to articulate in public and work for the entire agenda of values they would like to see realized in a "good" and "just" social system. Consider the hierarchical order of values we found in our "generation gap" studies of 1972–74, an agenda similar to that found in many other studies. At the top of the list come "Equality of life opportunities," "Satisfaction of people's basic needs," "Freedom of speech" and "Influence of citizens on the way the society is being governed." Anyone who has followed events in Poland is aware that all these values also stood at the top of the list of the demands of the workers during the strikes of last summer and fall. They are now written into the agreements that the members of Solidarity signed with the government to end the strikes. A national opinion survey conducted in the fall of 1980 showed the same list of values in substantially the same order. From a methodological point of view it can be said that history has provided strong behavioral validation of our survey data.

Those who have been studying values and attitudes in Polish society have actually been studying social forces of great importance. These forces were at least partly shaped under the influence of the social system, but they have now imprinted themselves on the shape of the system itself and, at least to a certain degree, have caused its transformation.



The Salinity of Rivers

Rivers normally wash into the ocean the salts dissolved out of rock. When the flow is held up by irrigation, however, evapotranspiration concentrates the salt in the soil, creating problems for agriculture

by Arthur F. Pillsbury

Many ancient civilizations rose by diverting rivers and irrigating arid lands to grow crops. For such projects to succeed human beings had to learn to work cooperatively toward a common objective. The most fruitful of the ancient systems was created at the southeastern end of the Fertile Crescent, the broad valley formed by the Tigris and the Euphrates in what is now Iraq. From there civilization spread eastward through present-day Iran, Afghanistan, Pakistan, India and thence into China, wherever rivers disgorged through valleys of recently deposited alluvial soil. At its peak of productivity each irrigated region probably supported well over a million people. All these civilizations ultimately collapsed, and for the same reason: the land became so salty that crops could no longer be grown on it. The salts that were washed out of the soil at higher elevations became concentrated in the irrigated fields as the water evaporated from the surface and transpired through the leaves of the growing crops. Although floods, plagues and wars took their toll, in the end the civilizations based on irrigation faded away because of salination.

There was a notable exception to this common fate: the valley of the Nile, at the western end of the Fertile Crescent. The explanation is that until very recent times the Nile valley was not really irrigated. Instead the annual flood of the Nile inundated a broad stretch of land extending the length of the valley, depositing a new layer of fertile soil year after year. The flooding prevented the salts from accumulating because as the water drained into the soil it leached the salts out of it and carried them downward into the ground water, which eventually drained into the bed of the river itself. As a result salt balance was achieved more or less automatically; the salts were borne into the Mediterranean century after century as they had been for millenniums before the dawn of agriculture. With the construction of the Aswan Dam at the headwaters of the Nile and the introduction of conventional irrigation Egypt is now faced with the universal problem of keeping salts from accumulating in the irrigated fields.

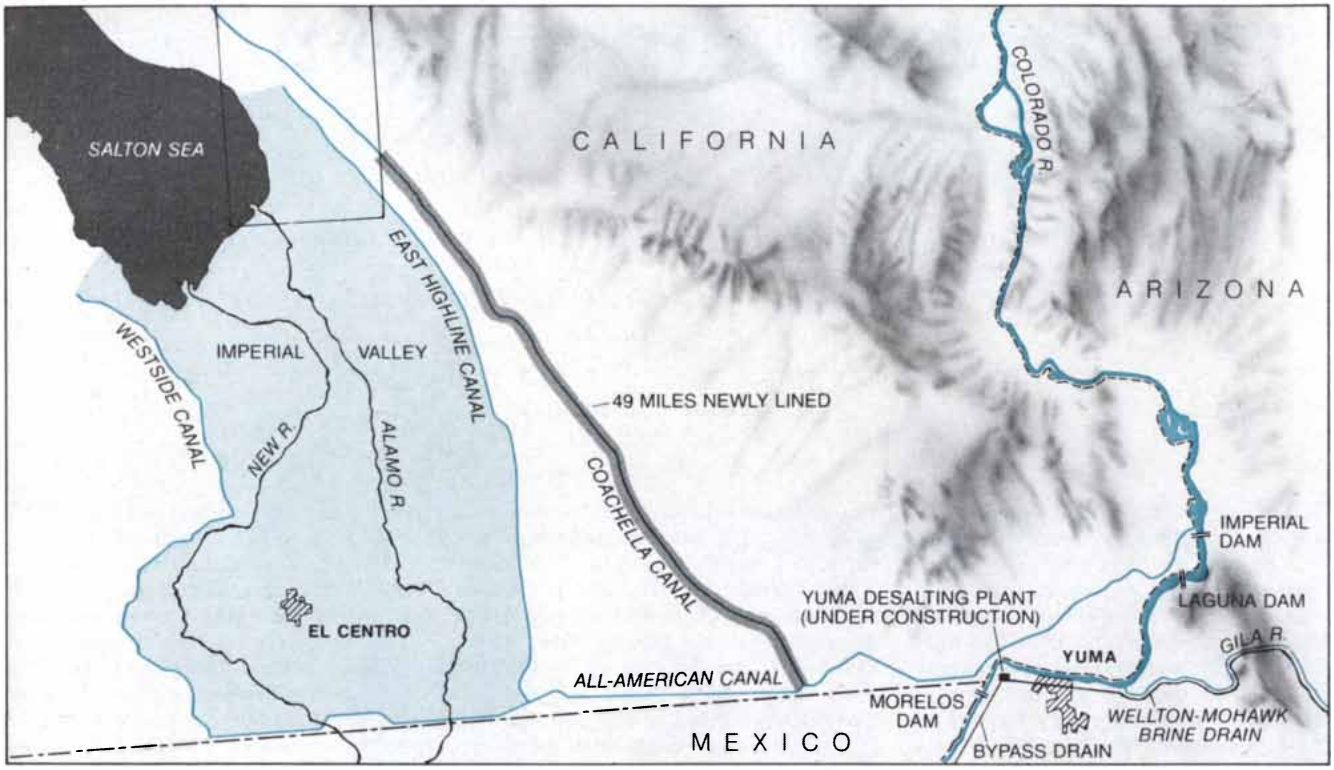
Irrigated agriculture has many advantages over humid-region agriculture in spite of the substantial costs of water and its application. Farmers can grow two or more crops per year of certain quicker-growing plants, such as grains and vegetables. Moreover, an arid climate can be expected to decrease the costs of cultivation, planting, harvesting and disease control and to increase yields and quality. Water can be supplied whenever it is needed to ensure optimum growth. Above all, the technology now exists to solve the salination problem and to make irrigated agriculture truly permanent. Here I shall describe that technology in terms of its application in the western U.S., where intensive irrigation has been practiced since the turn of the century.

All natural waters, including those described as fresh, contain salts. A virgin stream emerging from a mountain watershed may contain as little as 50 parts per million (p.p.m.) of "salt," or total dissolved solids. Ocean water averages about 35,000 p.p.m., or about 3.5 percent, of dissolved solids. I am speaking, of course, not simply of salt such as one finds on the dinner table: sodium chloride. Table salt does happen to be the most soluble of all the common salts, making up nearly a third of all the salt found in seawater, but in natural waters there are many other salts as well. The salts dissolved in such waters are usually dominated by the carbonates, chlorides and sulfates of calcium, magnesium and sodium.

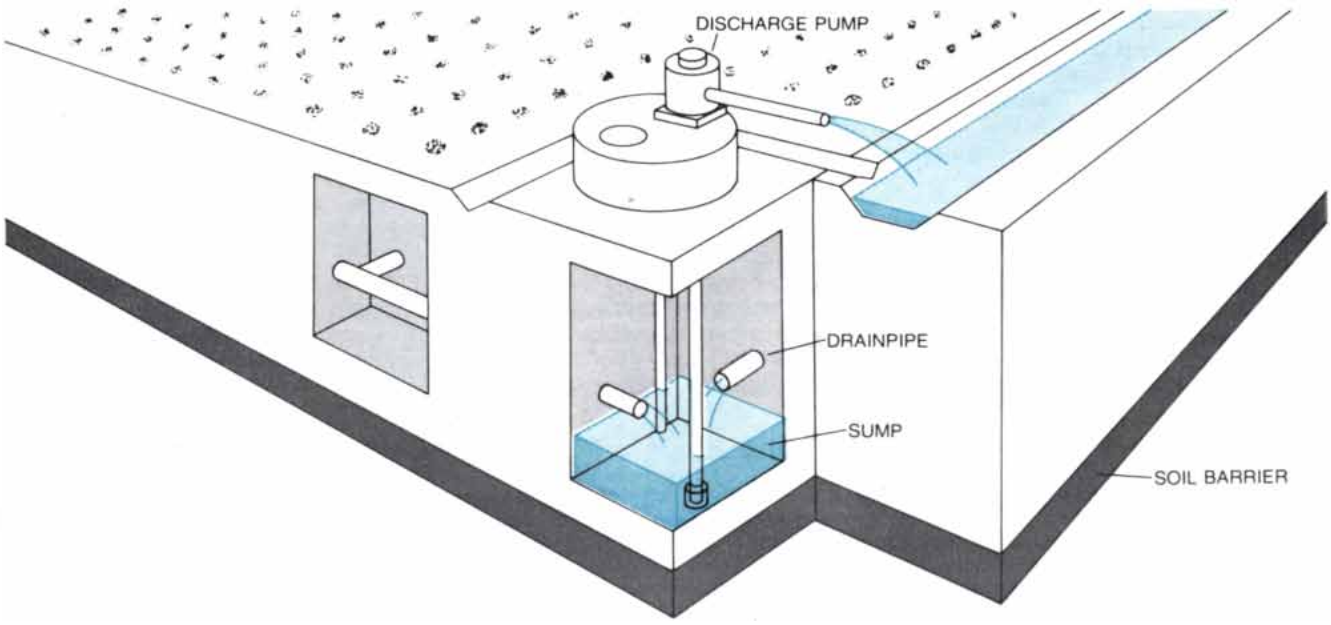
Accepting the commonly quoted estimate that the ocean basins contain 317 million cubic miles of salt water at an average concentration of 35,000 p.p.m. of dissolved solids, there are about 3.2×10^{18} tons of salt in the world's oceans. Additional large amounts of salt, largely in crystalline form, are sequestered in inland sinks and buried ancient sinks. Such salt deposits are one result of the geological processes summed up in the word weathering.

Weathering takes place under conditions where there is ample opportunity for the mineral crystals that constitute rock to oxidize. Although weathering embraces physical, chemical and biological processes, the physical processes are pervasive and central. Mechanical action fractures rock, exposing a far greater surface area to weathering agents. For example, the alternate freezing and thawing of water in the crevices of the rock exerts forces of compression and expansion that can break down the strongest material. Flowing water, wind and the grinding action of rocks in the bed of streams and the bottom of glaciers all contribute to physical weathering. Weathering manufactures both salts and the particles of rock that are borne from the uplands to the lowlands,

FORTUITOUS JUXTAPOSITION of the Imperial Valley and the Salton Sea in southern California near the border between the U.S. and Mexico makes it possible to export the brackish drainage waters from irrigated fields to a natural sink 370 square miles in area that is roughly as salty as the ocean. This photograph, taken from a U-2 aircraft operated by the National Aeronautics and Space Administration from a height of 65,000 feet, shows only the extreme northeastern tip of the nearly 500,000 acres that make the Imperial Valley the largest single expanse of irrigated agriculture in the Western Hemisphere. Growing vegetation, a strong reflector of infrared radiation, appears bright red in the Aerochrome infrared film used for the photograph. The water for irrigation is carried 80 miles from the Colorado River by the All-American Canal (*see map at top of next page*). The All-American Canal feeds the East Highline Canal, visible immediately adjacent to the fields along the Salton Sea, and the Coachella Canal, slightly to the east, which crosses the picture from upper left to lower right. The Coachella Canal carries Colorado River water another 123 miles to irrigated area north of Salton Sea.



ALL-AMERICAN CANAL originates at a reservoir behind Imperial Dam on the Colorado River. The 80-mile canal carries two-thirds of the 5.3 million acre-feet per year of water that California has recently been drawing from the Colorado. Ultimately the state will be limited to 4.4 million acre-feet per year. Some three million acre-feet per year is now applied to farmland in the Imperial Valley. The rectangle outlines the area shown in the U-2 photograph on page 54. The water as it is drawn from the Colorado carries about 800 parts per million (p.p.m.) of dissolved solids ("salts"). After about three-fourths of the water applied to crops has evaporated or transpired through the leaves of plants, the water drained from the fields contains about 3,200 p.p.m. of salts. Most of the drainage water is carried to the Salton Sea by the New River and the Alamo River. In past millenniums the Colorado emptied about half the time into the Gulf of California and half the time into the Salton Sea, forming the ancient Lake Cahuilla. When the Colorado was first seen by European explorers, it was flowing into the Gulf of California. It continued to do so until 1905, when it was again diverted into the Salton Sink (as it was then known) by a flood. It poured into the sink for two years before it could be rediverted to its former channel. Surface of the Salton Sea is now 230 feet below sea level, giving the water a maximum depth of 40 feet. The Salton Sea is the largest body of water in California.



DRAINAGE OF IRRIGATED FIELDS is often required to maintain salt balance in the root zone of plants. The irrigation water that percolates downward through the soil is enriched in salts because of evapotranspiration, the combination of direct evaporation and plant transpiration. To collect this brackish water the farmer installs pipes in parallel lines, usually about eight feet deep and 250 to 300 feet apart. The pipes, which are either loosely fitted or perforated, were once commonly made of tile or concrete but now are usually made of plastic. The drains form open channels through which ground water will flow when it reaches a level high enough to harm the roots. The brackish water usually flows to a drain but sometimes empties into sumps and is pumped into larger channels called collector drains.

where they are the principal constituents of soil.

The physical forces act most strongly at the highest elevations, where their extremes are greatest. As a result weathering and the consequent salt production are greatest at the highest elevations. An essential component of the weathering process is the water of precipitation, which serves to dissolve the salts. In areas of high precipitation, of course, the salts tend to be dilute.

When rain or snow falls, most of the water generally percolates into and through the ground; the excess beyond the capacity of the ground to retain water usually forms a water table at some depth below the surface. The ground water ultimately seeps into streams or reappears at the surface in the form of springs. It is this flow that makes streams and springs far more persistent than one would expect from the intermittent nature of the precipitation.

Much of the water retained in the soil and in the capillary fringe above the water table (the zone in which water can be drawn upward from the water table by capillary action) is available for the roots of plants to take up through the process of osmosis. Most of the water entering the roots is transpired through the leaves of the plant and passes as a vapor into the atmosphere. Water also evaporates directly from the surface of the soil. When the barometric pressure changes or when transient changes in pressure are induced by wind shear, the soil "breathes." Since the air in the soil is almost always essentially saturated with moisture, breathing represents additional evaporation. It is therefore not practical to determine how much of the moisture in the soil is lost by evaporation and how much is lost by transpiration, and so the total loss from land surfaces with vegetation is called evapotranspiration.

Water also evaporates directly from inland water surfaces such as those of creeks, rivers, lakes, marshes, canals and reservoirs. The result of all such evaporative processes is to distill pure water from the liquid phase, leaving all the salts behind in the water that remains. Since most of the watersheds in the western U.S. are quite thickly covered with vegetation, the natural concentration of salts in fresh waters is a widespread and significant process.

One of the principal natural resources of the western U.S. is its rivers, whose waters can be stored behind dams or in natural underground aquifers until they are needed. Another major resource is the alluvial soil of arid and semiarid valleys suitable for irrigated agriculture in a variety of excellent climates for growing crops. These resources and climatic conditions have been exploited to provide the nation and foreign markets with a diversified selection of foods and fibers that are largely complementary to the agricultural products of the humid regions.

Irrigated agriculture is expensive. Dams must be built to trap water and canals must be excavated to carry the water to where it is needed. If the water is in underground aquifers, energy must be expended to pump it to the surface. In California alone untold billions of public and private dollars have been invested in water development, chiefly for agriculture. The state is crisscrossed with thousands of miles of major irrigation canals and concrete pipelines that supply water to nearly nine million acres, a fifth of all the irrigated farmland in the nation. In 1980 the value of California's crops was $9.2 billion, representing more than 13 percent of the total value of all U.S. crops. California's farms provide some 40 percent of the nation's fresh food and vegetables.

The amount of water that must be applied annually to irrigated land is equivalent to between one foot and five feet of water covering the area cultivated, depending on the crop and the climate. Generally about three-fourths of the applied water is lost to evapotranspiration. The rest, which contains all the originally dissolved salts (except for the tiny amounts incorporated in the crop itself), percolates downward and laterally through the soil. It may enter an underground aquifer or it may reappear somewhere downstream as seepage into a river or into a natural sink, either directly or by way of a drainage ditch.

If, as seems reasonable, the average annual amount of water applied in irrigation in the western U.S. is equivalent to three feet covering the area cultivated, about 120 million acre-feet of water is applied annually to some 40 million acres of land. Roughly 90 million acre-feet of the total volume is lost by evapotranspiration. The remaining 30 million acre-feet holds essentially all the original salts: a fourfold concentration. (Water of good quality can be used at least twice, directly or indirectly, before it becomes too brackish for further use.) As a consequence of intensive irrigation the western states, particularly California, which applies more than 40 million acre-feet of water to its crops, face a major problem in isolating and disposing of brackish water.

In addition to the concentration of salts by evapotranspiration another kind of concentration takes place in the storage and transport of water before it ever arrives at the point of use. Evaporation occurs at the large storage reservoirs behind dams, along the extensive transmission channels and at the numerous regulating reservoirs. If the new water surfaces created are in mountainous areas where precipitation is high and where the land was formerly covered by forests and meadows, the increase in evaporation is probably of little significance; the reason is that the former evapotranspiration of natural vegetation is merely replaced by a roughly equal degree of simple evaporation. Most existing large reservoirs, however, are in desert areas or in environments formerly covered only sparsely by brush, grass-woodland or scrub timber, with the result that evaporation greatly exceeds the former evapotranspiration. Typical examples are the reservoirs formed by Hoover Dam on the Colorado River, Elephant Butte Dam on the Rio Grande and Grand Coulee Dam on the Columbia River.

The great rivers of the western U.S. originate in mountains where there is generally high annual precipitation. Characteristically in or near the mountains the concentration of total dissolved solids is low, commonly in the range of 50 p.p.m. for most rivers. Since a number of the western rivers flow through arid or semiarid regions for great distances, particularly in their lower reaches, the concentration of salts through evaporation rises steadily with distance downstream. If part of the flow is diverted for irrigation, the salt concentration is amplified by evapotranspiration. Increasingly water serves for cooling purposes, as in the cooling systems of fossil-fuel and nuclear power plants, which therefore also "consume" water through evaporation and thereby concentrate salts.

Some years ago I thought it would be interesting to analyze several major river systems to see how the quantity of salts carried by a river at a given point was related to the quantity of water that had entered the river above that point. To this end I calculated the annual salt production in tons per square mile of watershed above a given point on the river and the annual water production in acre-feet per square mile of the same watershed. The salt content and the flow of major rivers at various points are recorded by a network of stations run by the U.S. Geological Survey. I selected 10 years as the period over which to average the records because that was generally the maximum length of record available. For the survey I selected the following rivers: the Missouri, the Arkansas, the Rio Grande, the Pecos, the Colorado, the Gila, the Salt, the Agua Fria, the Sevier, the Humboldt, the San Joaquin, the Mokelumne, the American, the Columbia and the Willamette. These 15 rivers carry the bulk of the precipitation that falls on the 11 westernmost states of the U.S.

When salt production is plotted against water production on a logarithmic scale, the relation not only proves to be remarkably linear but also refutes the common assumption that rivers flowing through the more arid regions carry the

most salt [*see bottom illustration on opposite page*]. Actually the higher the water production per square mile, the higher the production of salts. The relation is all the more remarkable when one considers the great variation in the type and age of the rocks in the several river basins, the relative shortness of the 10 years of records available and the effects of diversions across watersheds, which transport substantial volumes of water of low salt content from one river basin to another.

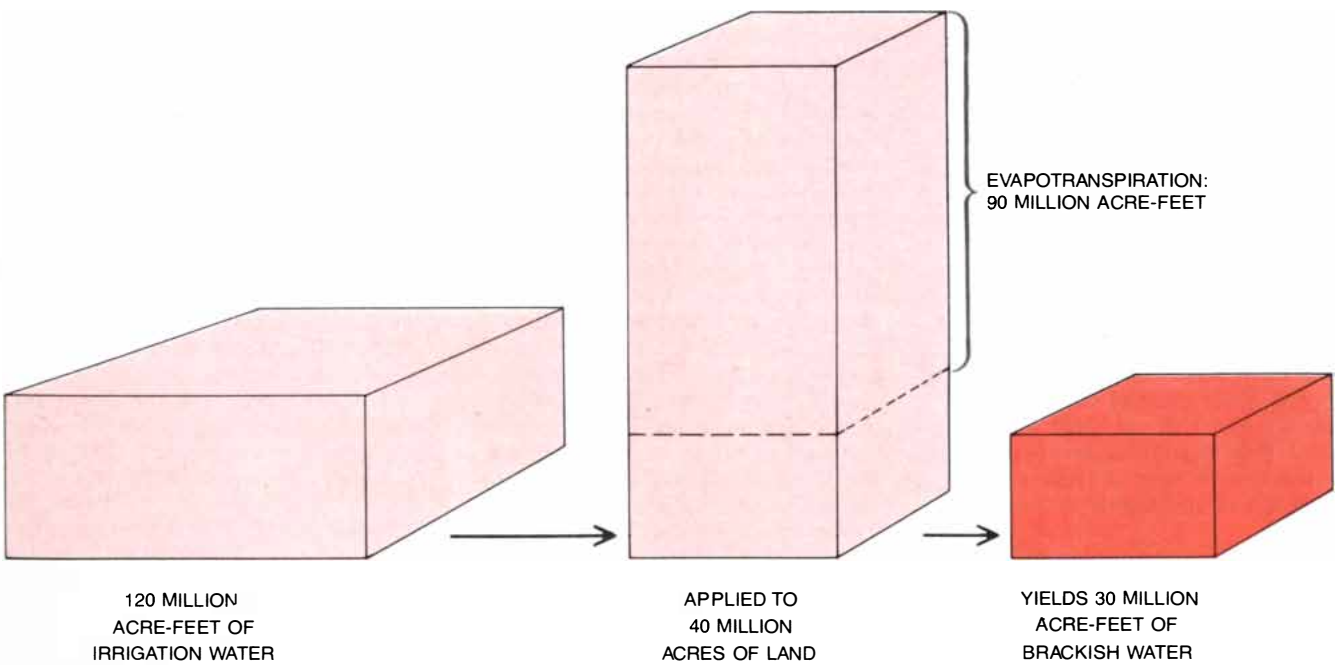
The most productive river in both salts and water is the Willamette in Oregon, which flows through a region of very high precipitation. The least productive in both salts and water is the Gila below Gillespie Dam in southwestern Arizona, a desert region. Even today it is asserted that rivers such as the Arkansas, the Pecos and the Colorado are unusually salty because they are slowly leaching away ancient buried salt beds. The evidence does not support such an assertion. The saltiness of rivers is simply a matter of the relative amount of water that is turned into water vapor through consumption, whether it is natural or the result of human activity.

The key to maintaining a salt balance in irrigated fields is adequate drainage. Whether it is natural or artificial, drainage refers to the removal of water from a place where it is not wanted to some other place, through a pipe or channel that can be on, above or below the land surface. The term agricultural drainage refers specifically to measures intended to lower the depth of a water table that is too close to the surface to allow the successful growing of crops. In humid regions the water table may have to be lowered in order to provide aerated soil around plant roots and to increase the firmness of the soil for tillage and other farm operations. For this purpose in such regions a network of ditches or tile drains is normally laid out at a depth of three to five feet below the surface.

In arid regions drainage must serve the additional function of maintaining a satisfactory salt balance in the vicinity of plant roots. When the weather is rainy or when irrigation water is being applied, water and salts both percolate downward. In dry weather and between irrigations the water and salts percolate upward through capillary action. In a humid area the amount of salt entering the soil is low to begin with, and so there is far less to travel upward through capillary action in dry periods. Moreover, the dry periods are usually short. In arid climates the drainage ditches or tile drains must be deeper than they are in humid regions in order to prevent a net upward movement of salts. To be effective the drainage channels must usually be at least six feet below the surface. The effluent from the drainage network must be discharged in such a way that the salt imported with the irrigation water is exported without harming the interests of water users downstream.

A remarkable and fortuitous facility for the disposal of brackish drainage water exists in southern California adjacent to the Imperial Valley, the largest single expanse of irrigated agricultural land in the Western Hemisphere. The water needed to irrigate the valley's more than 500,000 acres is carried a distance of 80 miles from the Colorado River by the All-American Canal. The All-American Canal also supplies the Coachella Canal, which carries Colorado River water an additional 123 miles to another rich agricultural area of some 65,000 acres. (The Colorado supplies a little more than half of all the water used in southern California, including the municipal water of the state's two largest cities, Los Angeles and San Diego.) Brackish irrigation water that drains from the Imperial Valley and Coachella agricultural districts is channeled into the Salton Sea, which at present is a little saltier than the ocean. Some 90 percent of the surface inflow to the Salton Sea is waste water from the Imperial Valley, Coachella and Mexicali districts.

Since 1955 the Imperial Valley irrigation district has been a net exporter of salts, draining out about 15 percent more salt than the All-American Canal carries into the district from the Colorado. The water drained into the Salton Sea contains about 3,500 p.p.m. of salts



IRRIGATION OF AGRICULTURAL LANDS represents the biggest consumptive use of water in the U.S. Water applied to crops is termed consumptive because three-fourths of it is dissipated into the atmosphere through evapotranspiration. It is estimated that about 120 million acre-feet per year of water is applied to some 40 million acres of land in the western states, or about three feet of water to every irrigated acre over the growing season. Assuming that three-fourths of the water, or 90 million acre-feet, is lost to evapotranspiration, the salts in the original volume, represented here by shades of color, are concentrated in the remaining 30 million acre-feet of water. This water, often containing more than 2,000 p.p.m. of salts, must be drained from the fields and then disposed of to prevent a buildup of salts.

and serves to retard the rate of increase in the Salton Sea's overall salinity level. The Salton Sea itself, which lies 230 feet below sea level, was a dry, salt-encrusted depression until 1905, when a flood on the Colorado broke through natural levees. The water of the Colorado poured into the sink for two years before it was rediverted into its former channel. With an area of 370 square miles, the Salton Sea is California's largest lake and a major recreational area.

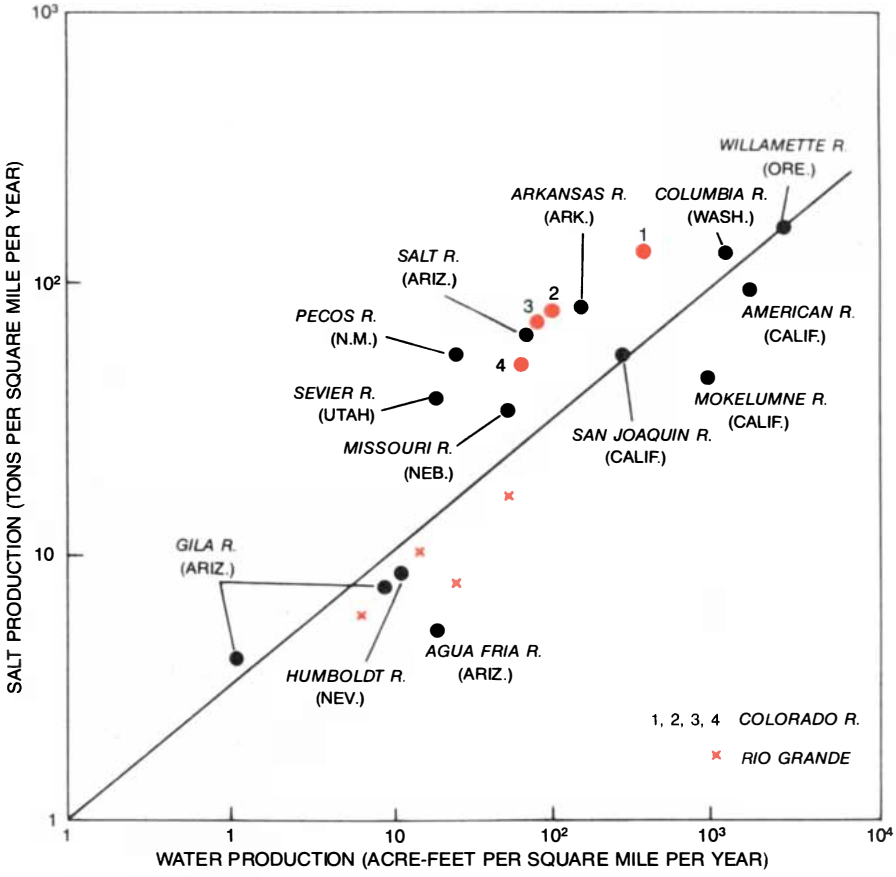
Where no Salton Sea or its equivalent exists to accept the drainage from irrigated fields the problem of achieving salt balance is more complex. People cherish the notion that the water of a river not only is fresh but also should be kept fresh right down to the river's mouth or to its entry into an estuary. In the humid regions of the world the departure from this ideal is seldom great, but the ideal is unrealistic in the more arid regions, where many of the rivers have been developed for irrigated agriculture. Before man began harnessing the rivers the seasonal floods were highly effective in carrying salts to the ocean and keeping the river basin in reasonably good salt balance. Today, with river flows being regulated by storage systems, and with high consumptive use of the released water, there is not enough waste flow left to achieve anything approaching balance. The salt is being stored, in one way or another, within the river basins.

Not only are salts getting bogged down somewhere in the system but also various measures are being taken that deliberately impede the flow of salts to the sea. In the U.S. it is the law of the land, reflecting the demands of both environmentalists and water users, that rivers remain, if not forever "wild and scenic," at least fresh for their entire length. The measures being planned and effectuated to accomplish this ideal are dangerous for the future. The general concept is to divert saline flows, where they can be found, into evaporation basins. There water will evaporate from the surface, leaving behind layer on layer of crystalline salts. It is proposed that the evaporation basins be situated either where the underlying ground is already saline or where the soil is relatively nonporous. Where neither is the case the ponds are to be lined with a presumably impervious material. Such schemes, designed to store the salts in the river basins themselves, may work for a few years or decades but are bound to be disastrous in the long run.

Why? The schemes will fail for any of several reasons. Although the ground waters under the evaporation basin may well be brackish or saline, every groundwater basin with a flow gradient must have an outlet somewhere near its lower

RIVER AND LOCATION	TOTAL DISSOLVED SOLIDS (PARTS PER MILLION)	ELECTRICAL CONDUCTIVITY (MILLIMHOS PER CENTIMETER)	SODIUM-ABSORPTION RATIO
COLUMBIA AT WENATCHEE, WASH.	78	.15	.2
SACRAMENTO AT TISDALE, CALIF.	180	.16	.6
MISSOURI AT WILLISTON, N.D.	574	.84	2.0
COLORADO AT YUMA, ARIZ.	740	1.06	2.2
RIO GRANDE AT EL PASO, TEX.	754	1.16	3.6
ARKANSAS AT LA JUNTA, COLO.	981	1.21	1.5

QUALITY OF IRRIGATION WATERS in the western U.S. varies greatly. The wide diversity is reflected in these samples taken from six rivers. The Columbia in the extreme northwest, 1,240 miles long and with a total discharge of 455 million acre-feet per year, is second in total flow only to the Mississippi. The Arkansas, which is even longer than the Columbia (1,460 miles), discharges only a sixth as much water and ranks 13th among U.S. rivers. The discharges of the other four rivers are considerably smaller. The term "total dissolved solids" is still widely used as a measure of salinity, but for the farmer the other two characteristics in the table are even more significant. Electrical conductivity has a greater influence on plant growth than salinity alone. Low values are preferable. Sodium-absorption ratio is calculated from the abundances of sodium, calcium and magnesium ions, expressed in milliequivalents of each ion per liter. Broadly speaking, the ratio expresses the excess of sodium, or the deficiency of calcium, that adversely affects the permeability of water in the soil. Values below 10 are satisfactory.

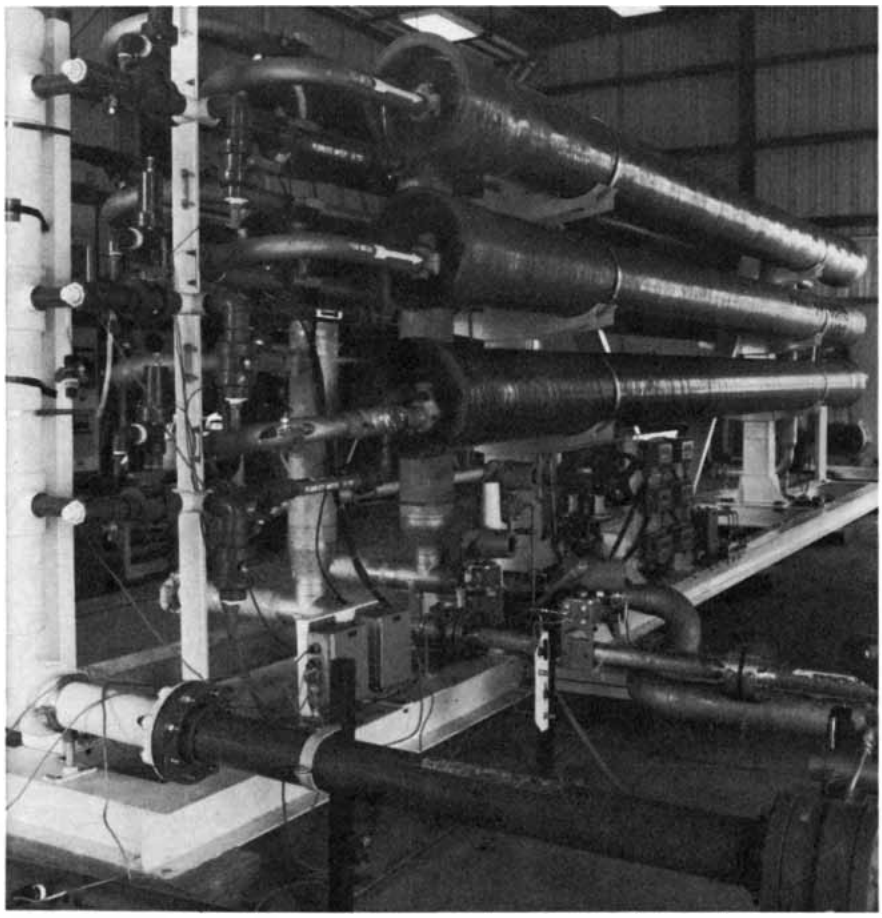


LEACHING OF SALTS FROM THE GROUND increases directly with the amount of precipitation. The higher the runoff of water per unit of watershed area is, the more salt is carried into the river that drains the watershed. To quantify the relation the author analyzed U.S. Geological Survey records for a 10-year period showing the salt content and river-flow volume at selected sampling points for 15 rivers in the western U.S. The salt content and flow rates were then related to the areas of the watersheds above the sampling points in order to derive values for salt production in tons and water production in acre-feet per square mile per year. A logarithmic plot of the two values falls remarkably close to a straight line. The Willamette River, which drains a watershed with the highest precipitation per square mile, also yields the most salt per square mile of watershed. The least productive river in both water and salts is the Gila River in extremely arid Arizona. The four sampling points (*colored dots*) on the Colorado River and the watersheds to which they correspond are shown in the map on page 62. Colored crosses are values for the Rio Grande, the only other river for which four sampling points are included.

end. The saline water in the evaporation basin will serve to increase the "head," or hydraulic pressure, on the saline waters below and will thereby increase the rate of discharge at the natural outlet, wreaking havoc in downstream ground waters and downstream lands. If the evaporation basin is situated above soil shown to be impermeable to fresh water, it will be found that the soil will gradually become more permeable when the waters are saline. This fact is well established. Many types of materials have been proposed for making evaporation basins impervious: rubber and plastic sheeting, asphaltic mixtures and special types of concrete. Conceivably some linings will be effective for as long as 50 years, but ultimately one must expect them to fail. In all probability their lifetime when they are exposed to saline water will be much shorter than their lifetime is when they are exposed to fresh water, for which they are normally intended.

Another prime effort today, designed to keep waters fresh in the lower reaches of river-basin systems, is the construction of "brine lines." These are lined canals or pipelines through which brackish waters are conveyed to the ocean or some other sink. The lines must be elevated above nearby rivers or the adjacent systems for distributing irrigation water to ensure that the saline drain water does not recontaminate the fresh water. This means that the effluent from field-drainage systems, along with any brackish or saline water from wells, must be pumped into the brine lines at a considerable cost in energy. Even if the energy cost is accepted, brine lines alone cannot establish a salt balance because there will still be ground-water flow below the drainage lines or above the saline aquifers that might be pumped.

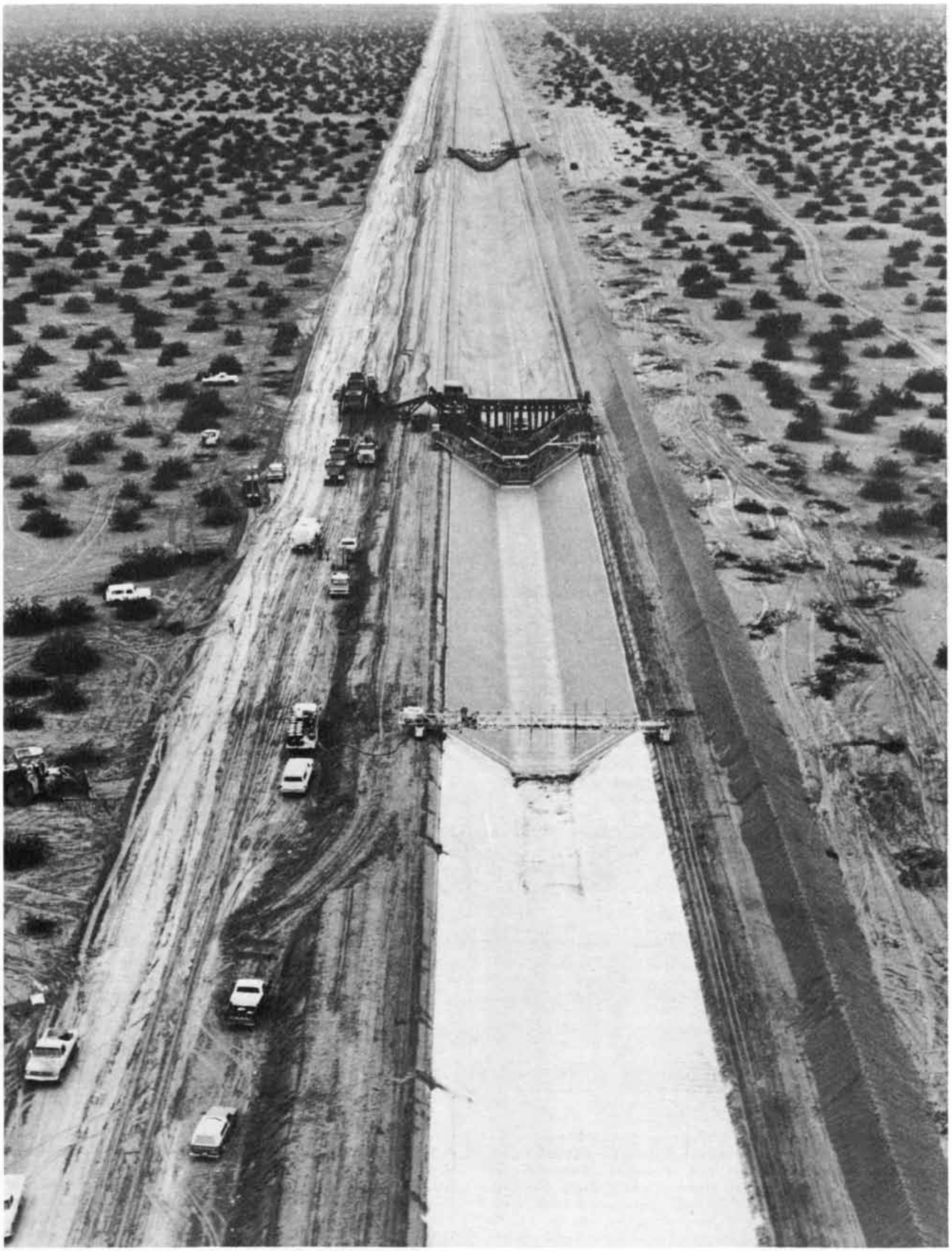
The U.S. Water and Power Resources Service (formerly the U.S. Bureau of Reclamation) has recently completed 82 miles of an open-channel brine line, roughly parallel to the San Joaquin River in California, that is designed to drain up to 30,000 acres of prime agricultural land in the state's San Joaquin Valley. Probably the longest brine line yet built, it now discharges into the Kesterson Reservoir south of Modesto. From there the saline waters gradually seep into the San Joaquin River as it approaches "the Delta," the estuarial area at the head of San Francisco Bay formed by the confluence of the San Joaquin and Sacramento rivers. The section of the drain that will carry the brine directly into the Delta has yet to be built. The existing canal is the first segment of a proposed "Master Drain," 290 miles long, to be financed with Federal and state funds. The estimated cost is more than $1.2 billion. By the year 2005 the proposed drain will serve 500,000 acres and will have the capacity to remove more than three million tons of salt per year.



DESALTING TEST FACILITY is one of two installations at Yuma, Ariz., recently built for the U.S. Water and Power Resources Service as the first stage in a program to treat brackish water from the Wellton-Mohawk Irrigation and Drainage District that is now too salty to be returned to the Colorado River. At present the brackish water is carried by a 51-mile bypass drain to the Gulf of California. This test facility, built by the Fluid Systems Division of UOP, is designed to extract some 325,000 gallons per day of low-salinity water (about 250 p.p.m. of solids) from a brackish input stream of 470,000 gallons per day with a salinity of about 2,800 p.p.m. Desalting is accomplished by a process of reverse osmosis. Saline water is forced against a plastic membrane at a pressure of 300 to 400 pounds per square inch. Water passes through and salt is left behind. The full-scale plant, using two different but similar processes, will have a design capacity of some 72 million gallons of desalted water per day. The product stream will be blended with untreated drain water to yield about 92,000 acre-feet per year of water (with a quality higher than that required by treaty) that can be added to the Colorado for export to Mexico. An equivalent volume of water will be made available for irrigation upstream in the U.S.

The difficulty of achieving salt balance in river basins where there is high consumptive use of water can be better appreciated if one reflects on the basins' complex hydrological history. Valley and basin lands consist primarily of soils that have been deposited by floods. The river channel that winds through the valley and basin today has a look of permanence that is deceptive. In a flood heavy debris is tumbled along the deepest part of the channel and the river overflows its banks. As the water moves laterally there is a marked decrease in the velocity and depth of its flow. As a result sand and other small particles settle out, creating natural levees. The finest soil particles settle out at a considerable distance from the central channel and at a much lower rate. The soil that settles in the valley lands is thus of medium texture, chiefly loams.

Where the velocity of flow falls almost to zero in the broad basin lands farther downstream the fine-textured clays settle out. The channel itself, including the natural levees, tends gradually to rise above the surrounding land. Later, in some exceptional flood, the river will overflow its banks again and create a new channel where the slope below it is steeper. The new channel will capture the old channel as it erodes upstream. In time, as deposition from flood after flood raises the land, an interlacing network of buried channels is covered over. Both shallow and deep ground water tend to follow such inter-



TO CONSERVE WATER formerly lost through seepage 49 miles of the 123-mile Coachella Canal have recently been rebuilt and lined with concrete. The new section, a part of which is shown near completion here, is expected to save 132,000 acre-feet of water per year, reducing the amount drawn from the Colorado River via the All-American Canal from 498,000 acre-feet per year to 366,000 acre-feet.

COLORADO RIVER BASIN drains an area of 242,000 square miles in seven states. Nine major reservoirs on the river have a total storage capacity of about 65 million acre-feet, or roughly 4.5 times the annual flow of 14 million acre-feet measured at Lee Ferry. The largest reservoir, with a capacity of 26 million acre-feet, is Lake Mead, made by Hoover Dam. Under the terms of the Colorado River Compact of 1922 the four states in the upper basin above Lee Ferry are allotted 6.5 million acre-feet annually and the three states in the lower basin (Arizona, Nevada and California) are allotted 7.5 million acre-feet. When the Central Arizona Project, designed to carry 1.2 million acre-feet per year, is completed in 1985, California will be limited to 4.4 million acre-feet, leaving 1.5 million acre-feet for Mexico. In the drought year of 1977 California drew on the Colorado River for 5.6 million acre-feet of water. The four sampling points selected by the author in his study of salt and water production in the Colorado basin, shown in the bottom illustration on page 59, are Glenwood Springs (*1*), Lee Ferry (*2*), Grand Canyon (*3*) and a point below Hoover Dam (*4*). The watersheds corresponding to these areas are identified by the light-colored broken outline. Whereas the Glenwood Springs watershed produces 350 acre-feet of water and 135 tons of salt per square mile per year, the Hoover Dam watershed yields only 60 acre-feet of water and 50 tons of salt per square mile per year.

lacing channels, with the deeper water appearing in what are called finger aquifers. Particularly in the basin lands the shallow ground water commonly seeps into the present river channel. Such diffused flow is deeper than the ditches or tile drains constructed for agricultural drainage.

Tile drains are most effective, as a rule, in the basin and basin-rim lands and in certain stratified soils where the "semiperched" water table rises closer to the surface than six or seven feet. The water table is said to be semiperched when the variable layering of the river-deposited soil tends to isolate the water near the surface from the main body of ground water lying at deeper levels. Under those conditions there is little opportunity for the irrigation water, enriched in salts, to percolate downward and degrade the deep ground water, which remains available for irrigation or other uses. Farther upriver in the valley lands the subterranean structure is such that near-surface water cannot be isolated from deeper water, with the result that ditch or tile-drain systems are powerless to preserve the quality of the ground water.

Before the advent of intensive irrigation the ground waters of the western valleys and basins were almost uniformly of high quality. The underground aquifers were largely recharged at the upper end of the valleys where the rivers disgorged onto the valley lands. The ground waters subsequently discharged into the basin lands and for the most part into the rivers themselves in the form of diffused flow. When farmers began pumping irrigation waters from the aquifers, the net effect was to lower the water table downstream from the pumping, thereby reducing the discharge flow back into the rivers. Eventually the water table would fall so low that there was no discharge at all and the aquifers became closed basins. Salts could then no longer be passed along downstream and simply accumulated.

In most regions the water pumped from aquifers supplies both irrigation and urban needs. The urban wastes collected as sewage generally show an increase in total dissolved solids of 300 to 350 p.p.m. Where such wastes are not discharged directly into the ocean all the salts coming into the basin remain trapped and build up within it.

When the aquifers being pumped for agricultural and urban needs are near the coast, the water table has often been pumped below sea level, with a consequent intrusion of seawater into the ground-water basin. The usual way to stop the intrusion has been to drill a series of injection wells parallel to the coast. The water pumped into the injection wells can be somewhat brackish and in some cases is treated urban sewage. The technique has been successful in creating "mounds" of water that repel the seawater. The objection to such schemes is that they totally block the export of salts that would otherwise be carried to the sea by ground waters. Obviously corrective measures must eventually be taken.

The only effective way to keep ground-water basins fresh is to pump from wells near the lower end of each basin, where the salinity is highest, and to hurry the effluent on its way to the ocean or some other sink. At the same time it will probably be essential to augment the recharge near the upper end of each basin. Unless these steps are taken one can foresee the day when the aquifers will be destroyed by salinity.

The custom traditionally followed in the U.S. in developing water resources has been to expect the river itself to carry supplies of fresh water to points of diversion almost down to the river's mouth. Such a design is generally the cheapest and has the advantage of capturing floodwaters that upstream storage would miss. This ignores the basic principle, essential for the long term, of going upstream for supply and allowing the lower rivers to become brackish. Let me illustrate with three examples: the Rio Grande system, the Colorado River system and the Delta east of San Francisco Bay.

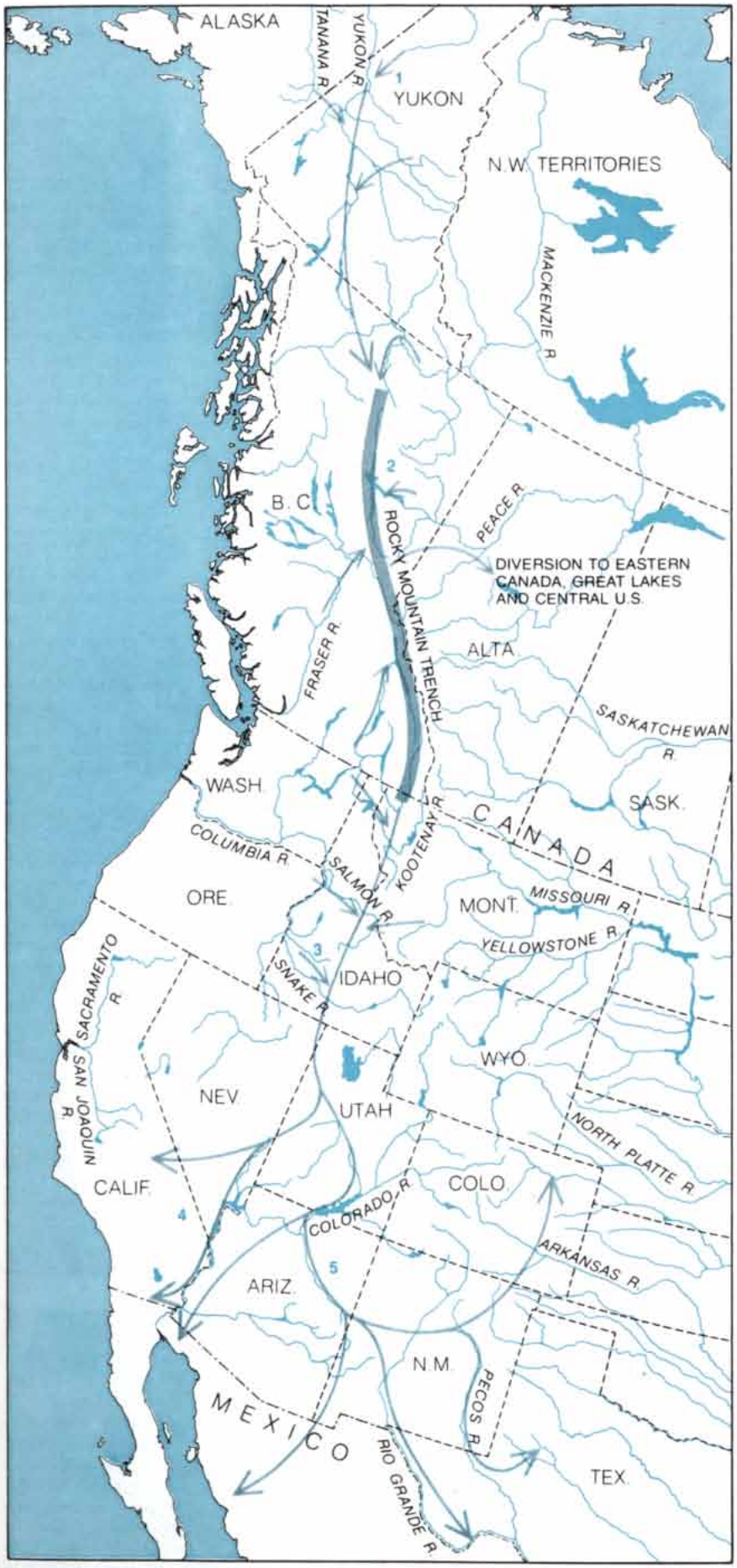
The Rio Grande, which rises in southwestern Colorado and empties into the Gulf of Mexico, is nearly 1,900 miles long. It is the third-longest river in the U.S., yet like the Colorado it does not qualify for inclusion in the Geological Survey list of 33 rivers with the highest discharge. Over the final 800 miles of its length it is the principal boundary between the U.S. and Mexico, and the allocation of its water has long been a matter of contention between the two countries. Essentially the entire flow of the upper Rio Grande, except during floods, is stored and utilized upstream from El Paso, at the extreme western end of the U.S.-Mexico border. There is almost no waste that would make it possible even to approach salt balance. Severe salt problems are gradually developing in southern New Mexico and western Texas. Along the lower river, between El Paso and the mouth in the Gulf of Mexico, there are three major international dams: Amestad, for storage; Falcon, for more storage and hydropower generation, and Anzalduas, for diversion, chiefly irrigation. (Mexico calls the lower river from El Paso to the Gulf the Rio Bravo del Norte; most of the water entering it is runoff from mountains in Mexico.) The present plan of development eliminates any chance of achieving salt balance either above El Paso or below it.

The Colorado River, 1,450 miles long, supplies more water for consumptive use than any other river in the nation. Its well-known Hoover Dam created Lake Mead, a storage reservoir and recreational area of some 250 square miles. Below Hoover Dam and its 1,345-megawatt hydroelectric power plant are seven more dams, two of which also serve to generate power. The dam farthest downstream is Morelos Dam, which stores water for irrigating the Mexicali Valley in Mexico. Irrigated regions near the river send back their drainage water enriched fourfold in salt. About 70 percent of the total flow below Hoover Dam, containing about 700 p.p.m. of salt, is exported to California, chiefly through the Colorado River Aqueduct, which supplies Los Angeles and San Diego, and the All-American Canal, which supplies the Imperial Valley and Coachella agricultural regions.

Primarily because so much water (and salt) is exported, the lower reaches of the Colorado are in reasonable salt balance. The problem is that the salt content of the lower river is high: more than 800 p.p.m. In 1974, at the behest of the Environmental Protection Agency, the seven states of the Colorado River basin agreed on a program to maintain the salinity in the lower basin at or below the level measured in 1972: 723 p.p.m. below Hoover Dam, 747 p.p.m. below Parker Dam and 879 p.p.m. at Imperial Dam.

The same year, in partial satisfaction of the 1944 Mexican Water Treaty, the U.S. agreed that the salinity of the Colorado River water delivered to Mexico at Morelos Dam should not be allowed to exceed the average salinity of water arriving at Imperial Dam by more than 115 ($\pm$30) p.p.m. The treaty with Mexico provides that the U.S. shall deliver to Mexico 1.36 million acre-feet of Colorado River water and another 140,000 acre-feet from well fields adjacent to the Colorado at the U.S.-Mexico border, for a total of 1.5 million acre-feet. In order to ensure the required river flow at the agreed salinity, the Water and Power Resources Service has undertaken to build a desalting plant at Yuma, Ariz., that will process a large volume of brackish drain water from the Wellton-Mohawk Irrigation and Drainage District in Arizona. The brackish water was formerly discharged into the Colorado below Morelos Dam; it is now exported to the Gulf of California through a bypass drain 51 miles long, mostly in Mexico, that was completed in 1977 at a cost of $27 million, paid by the U.S.

The desalting plant, when it is completed in the mid-1980's at an estimated cost of $216 million, will be the largest of its kind. It will take in about 107,000 acre-feet per year of water with an average salinity of 2,800 p.p.m. and will



yield a cleansed stream of 73,000 acre-feet (65 million gallons a day) with a salinity of only 255 p.p.m. and a brine stream of 34,000 acre-feet with a salinity of 8,200 p.p.m. The brine stream will continue to be sent on to the Gulf of California. If low-salinity makeup water is needed to meet treaty obligations, the cleansed stream can be returned undiluted to the Colorado at Yuma. Normally, however, the cleansed stream will be blended with untreated drainage water to yield up to 92,000 acre-feet of water with a salinity of less than 800 p.p.m. The reclaimed water will cost about $250 per acre-foot, more than 30 times the cost of irrigation water in the Imperial Valley. The desalination will be accomplished by plastic membranes that remove salts by means of reverse osmosis.

To further ensure the flow to Mexico and to reduce the volume of water taken by California, 49 miles of the Coachella Canal have recently been rebuilt and lined with concrete at a total cost of $45 million in an effort to save 132,000 acre-feet per year of water previously lost by seepage into the surrounding desert. It is expected that the amount drawn from the Colorado River via the All-American Canal can be reduced from 498,000 acre-feet per year to 366,000 acre-feet. The lining of the canal will also help California to live within its ultimate al-

ELABORATE SCHEME for diverting water from abundant sources in Alaska and Canada to regions that are already growing short of water was originally proposed by the Ralph M. Parsons Company of Pasadena, Calif., in 1964. Only the western section of the total plan, known as the North American Water and Power Alliance (NAWAPA), is depicted schematically here. An eastward extension of the plan would divert water from British Columbia to the upper Mississippi River, the Great Lakes and eastern Canada. It is estimated that the entire project would cost at least $200 billion and take 30 years to complete. NAWAPA would yield about 160 million acre-feet of water per year for industrial, municipal and agricultural uses. Roughly half would be allocated to the U.S. and the balance to Canada and Mexico. A series of hydroelectric plants would generate all the power needed for pumping and provide a surplus capacity of at least 70,000 megawatts. Water collection would begin at the headwaters of the Yukon and Tanana rivers in Alaska (*1*) and would be supplemented by utilizing a series of streams extending down to the Peace River (*2*). The collected water would flow into a 500-mile reservoir, the Rocky Mountain Trench, made by damming the upper reaches of the Columbia, Fraser and Kootenay rivers. More water would enter the system from the Clark, Snake and other rivers in the northwestern U.S. (*3*). Water for irrigation and other purposes would be sent to southwestern states and northwestern Mexico (*4*). A branch would supply New Mexico, Texas, Colorado, Kansas, Nebraska, Oklahoma and other parts of Mexico (*5*).

location of 4.4 million acre-feet of Colorado water. (It has recently been drawing 5.3 million acre-feet per year, and it drew 5.6 million acre-feet in the drought year of 1977.) Although these various costly measures should enable the U.S. to meet its treaty obligations to Mexico, it is clear that the Colorado can yield no additional water for the expansion of agriculture.

The Delta is an estuarial area of some 1,200 square miles formed by the confluence of the Sacramento, the San Joaquin and several smaller rivers near the middle of the rich Central Valley agricultural region. This large area was originally bulrush-covered marshland at or near mean sea level. Beginning in a small way sometime after the passage of the California Swamp and Overflow Act of 1850 but not before the rare great flood of 1861–62, the land was gradually reclaimed by the building of levees around numerous small islands outlined by the rivers and sloughs. When better pumps became available around the turn of the century, the process was greatly accelerated. The reclamation continued into the 1920's. The material for the levees came from dredging the rivers and sloughs. As with the polders of the Netherlands, the reclamation requires that drainage water and rainwater be continually pumped out into an adjacent waterway. Irrigation water is obtained simply by siphoning water over the levees. There are some 550,000 acres of cropped land in the Delta complex, and the labyrinth of waterways is heavily exploited for recreation.

The land was originally all peat, often to depths of 35 feet or more. As happens whenever peat soil is drained, there has been widespread subsidence of the ground level, amounting to about three inches per year. Some land is now as much as 21 feet below mean sea level. Not surprisingly there have been many levee failures. Over the past two years four large islands were inundated when their levees broke. State and Federal agencies are now working hard at public expense to repair and strengthen the levees involved. It is reasonable to question whether the expense is justified, given the fact that the subsidence problem is clearly a long-term one.

An alternative approach to the Delta problem, proposed some years ago, would be to provide the stablest farmlands in the Delta with an overland supply of fresh water from the projected Peripheral Canal. The canal would also serve to carry water from the Feather River, north of the Delta, around the eastern end of the Delta to the San Joaquin Valley on the south. Although the California legislature authorized construction of the Peripheral Canal last year, the project has run into heavy opposition from environmentalists, among others, who are demanding that the issue be submitted to a statewide referendum. In addition to the construction of the Peripheral Canal, at a currently estimated cost ranging from $700 million to $1.3 billion, the legislature authorized the expenditure of $4.1 billion for construction over the next quarter century of dams and other water-development facilities in the northern part of the state.

At present the water exported south and west must move through the Delta sloughs before it reaches the pumps that send it to the ultimate users. Periodically fresh waters that normally flow downstream through the Delta reverse their direction and allow salt waters from San Francisco Bay to travel upstream toward the export pumps. As a result the Delta waters and the water destined for irrigation are both seriously degraded in quality.

The Peripheral Canal has been opposed by many in northern California who evidently do not recognize that the water contractors of the State Water Project have long been paying in full on the Oroville Dam complex on the Feather River that is designed to create the "new" regulated water destined for the canal. The export pumps on the Delta cannot possibly provide the amounts of water the state has contracted to deliver unless the Peripheral Canal is built. Opposition now comes from those who advance the concept that the water belongs in the "area of origin" and that those who are paying for it must now surrender their right to it.

It should be obvious that in order to maintain and ensure the long-term viability of irrigated agriculture and to provide enough water to carry the salts to the ocean or some other natural sink, the development of water resources should be intensified. First there must be much more development on the local scale and much more attention paid to local practices. Eventually, however, some grand-scale water-diversion concept will be needed, simply because much of the West, including the High Plains of Texas and New Mexico and from there northward through Nebraska, is short of water. All parts of northern Mexico, the U.S. and Canada with an irrigation potential should have an adequate supply of water, and that means a supply sufficient to flush all the salts to the ocean or some other natural sink. It should also be recognized that there are important economies of scale: doubling the amount of water to be supplied by a project does not double the cost. The cost increases by a factor not of 2 but of something less than 1.5.

Over the years a number of visionary schemes have been proposed for diverting water from rivers in the humid northwestern U.S. to more arid regions to the south. One plan, conceived by a private engineer, William G. Dunn, proposes to transfer 2.4 million acre-feet per year from the Snake River in Idaho to a point below Hoover Dam on the Colorado River, a distance of some 600 miles. The Snake, which discharges about 30 million acre-feet per year into the Columbia River, is the nation's 12th-largest river. The present flow of the Colorado below Hoover Dam is about 14 million acre-feet per year, of which California will be entitled to take 4.4 million after completion of the Central Arizona Project, beginning in about 1985. A more ambitious plan, put forward by Frank Z. Pirkey, another experienced consulting engineer, proposes the transfer of 15 million acre-feet per year from the Columbia River to Lake Mead, behind Hoover Dam. In Pirkey's plan the water would first be pumped 4,900 feet over the mountains to Goose Lake on the Oregon-California border and then to Shasta Lake behind Shasta Dam in northern California before its final transfer to Lake Mead. Along the way substantial amounts would be available for irrigation.

By far the most audacious scheme yet advanced is the one proposed in 1964 by the Ralph M. Parsons Company of Pasadena, Calif., one of the engineering firms that built Hoover Dam in the early 1930's. Known as the North American Water and Power Alliance (NAWAPA), the scheme would divert waters from Alaska and northern Canada to many parts of Canada, the U.S. and Mexico. Hydroelectric plants along the way would generate substantially more power than what would be needed for pumping. The total drainage area envisioned by the plan covers about 1.3 million square miles where precipitation is plentiful. Of a total runoff of more than 800 million acre-feet per year NAWAPA would divert some 160 million acre-feet southward for consumption and waterway control. The completed system, estimated to take 30 years to build and to cost more than $200 billion, would provide a surplus hydroelectric capacity of about 70,000 megawatts, equal to nearly 25 percent of present U.S. average production.

Although the magnitude of the NAWAPA plan is staggering and the plan would have to surmount formidable political hurdles before it could be implemented, it is in my opinion the only concept advanced so far that will enable the lower reaches of western rivers to achieve the salt balance necessary for the long-term health of western agriculture, on which the entire U.S., and indeed the world, has much dependence. Unless the lower rivers are allowed to reassert their natural function as exporters of salt to the ocean, today's productive lands will eventually become salt-encrusted and barren.

The Atmosphere of Venus

A decade of exploration by spacecraft now shows that it consists almost entirely of carbon dioxide. Its clouds of sulfuric acid are driven by winds that attain a speed of 360 kilometers per hour

by Gerald Schubert and Curt Covey

Until a few decades ago the atmosphere of Venus was unknown. Three centuries of observations with telescopes had revealed only that the planet is covered by an unbroken deck of seemingly uniform, featureless clouds. Today observations made from the earth have been extended to ultraviolet wavelengths. In this part of the electromagnetic spectrum markings are seen in the clouds. Moreover, since 1967 a series of interplanetary spacecraft have made observations at closer range, and a total of 13 probes have penetrated deep into the atmosphere of Venus before the heat of the atmosphere disabled them. Remarkably, two of the spacecraft in the Venera series launched by the U.S.S.R. survived long enough to return data from the surface, where the year-round temperature of 460 degrees Celsius is high enough to melt metals such as zinc. A U.S. spacecraft, the Pioneer Venus orbiter, which arrived at Venus in December, 1978, continues to circle the planet and transmit information back to the earth.

Today the atmosphere of Venus is known to be predominantly (96 percent) carbon dioxide. The rest is nitrogen with trace amounts of other substances, including water vapor. The atmosphere of Venus is more than 90 times as massive as the atmosphere of the earth, so that the pressure at the surface of Venus is more than 90 times the pressure at the surface of the earth. It is a pressure encountered at a depth of one kilometer in the oceans of the earth. The clouds in the atmosphere of Venus are now known to form continuous layers at an altitude roughly between 45 and 60 kilometers. (On the average the earth is half covered by clouds that lie at an altitude of less than 10 kilometers.) Investigators distinguish three layers of clouds on the basis of the concentration and the size of suspended particles. It is hypothesized that the particles may differ from one layer to another, but the only particles identified so far are liquid droplets composed of sulfuric acid with an admixture of water.

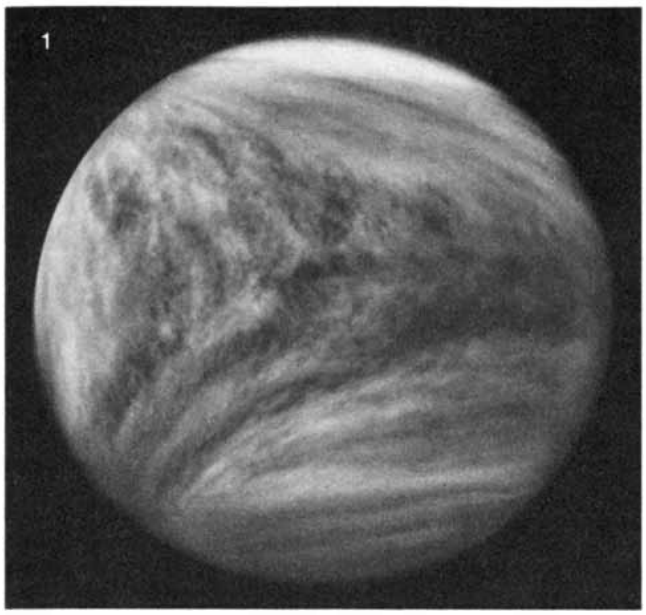
The winds of Venus are now known to be dominated by a planetwide east-to-west circulation that attains a velocity of 100 meters per second (360 kilometers per hour) at the altitude of the cloud tops. At that speed a parcel of the atmosphere would circle Venus in four earth-days. The planet itself also turns east to west, but it takes 243 earth-days to complete a single rotation. In a word, the atmosphere of Venus superrotates: at the cloud tops it moves more than 60 times as fast as the planet does. In contrast the earth turns west to east, and the atmosphere of the earth (considered on a global scale) rotates synchronously with the solid planet below it.

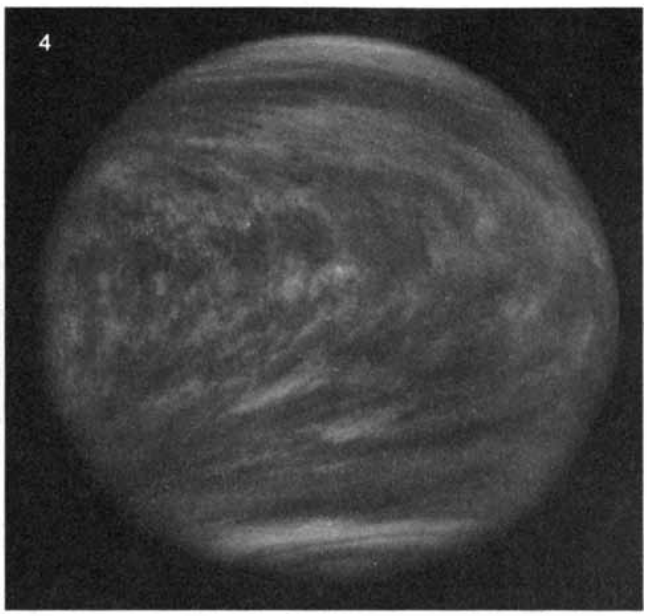
The focus of research on the atmosphere of Venus is now changing. The challenge is no longer to discover what the atmosphere is like. The challenge is to explain why the atmosphere of Venus should be so different from that of the earth in spite of the fact that Venus and the earth are nearly the same in size and mass.

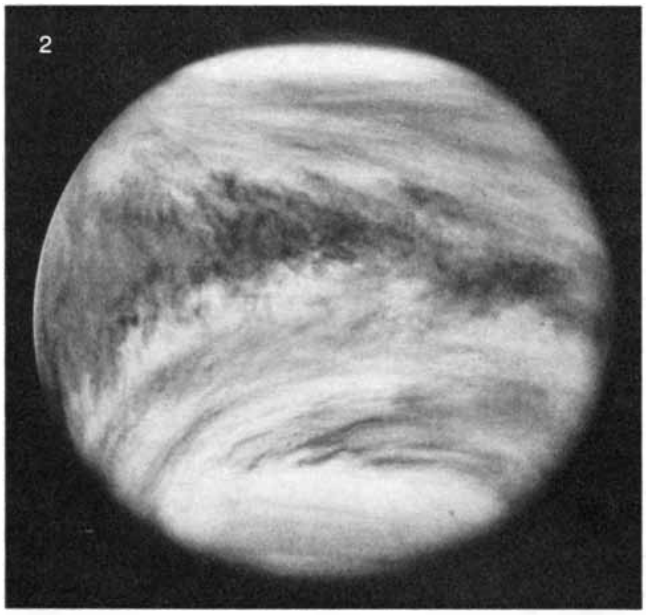
The atmosphere of a planet is well characterized by the way its temperature varies with altitude. On Venus the pattern of variation divides the atmosphere into two regions. In the lower one, which extends from the surface to an altitude of about 100 kilometers, the temperature decreases with height. This region is called the troposphere by analogy with the lower part of the atmosphere of the earth (the part below 10 kilometers), in which the temperature also decreases. Throughout most of the height of the troposphere of Venus the temperature falls at a rate of about 10 degrees C. for each kilometer of altitude. At the surface the temperature of 460 degrees changes little from day to night. Indeed, one must make measurements at the height of the clouds to find diurnal changes. The reason for this lack of variation is clear. The lower atmosphere is so dense that it has great thermal inertia: it can store a large amount of heat. For the same reason one would not expect to find a significant difference between the daytime and the nighttime temperature in the oceans of the earth.

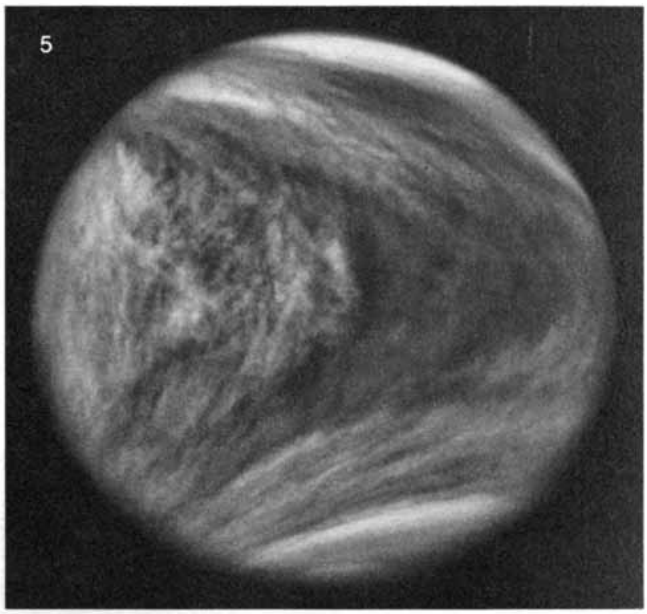
The high temperatures below the clouds, culminating in the surface temperature of 460 degrees, cannot be attributed simply to the fact that Venus is closer to the sun than the earth is; indeed, the ubiquitous cloud cover on Venus is so reflective that the planet absorbs less solar radiation than the earth. Furthermore, as has been shown by Pioneer Venus probes and by the Venera series of probes, only a small fraction of the radiation absorbed by Venus penetrates the clouds and the massive lower atmosphere and reaches the surface. The planet remains hot in spite of all this because the surface reradiates the solar energy in the infrared part of the electromagnetic spectrum; constituents of the atmosphere such as carbon dioxide, sulfur dioxide, water vapor and cloud particles are efficient absorbers of infrared radiation, and so the heat is trapped in the troposphere. In short, the tropo-

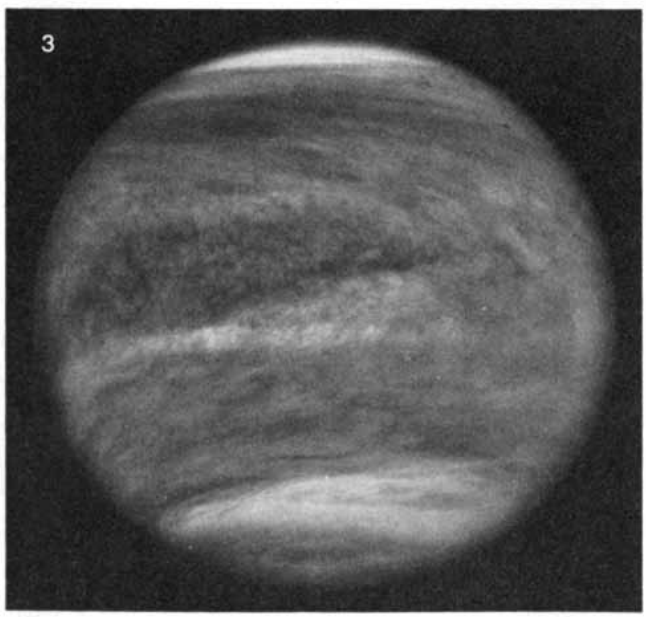
DOMINANT WIND ON VENUS is revealed in the series of photographs on the opposite page, which record the image of the planet in ultraviolet radiation. The photographs were made on consecutive days by the Pioneer Venus orbiter, a U.S. spacecraft that took up its trajectory around the planet in December, 1978. In the first photograph, made on February 15, 1979, a *Y*-shaped marking in the clouds of Venus opens toward the west of the planet. The stem of the *Y* is a broad, dark band along the equator. In the second photograph the vertex of the *Y* is near the west limb of the planet; in the third photograph the vertex is out of view; in the fourth the arms are returning; in the fifth, made on February 19, the complete *Y* is again exposed. The dominant wind is thus a rotation of the atmosphere from east to west that drives cloud markings around the planet in only four earth-days. Venus itself also rotates westward, but the rotation of the solid planet is more than 60 times slower than that of the atmospheric markings. The final photograph shows Venus on February 20, 1979. The photographs were provided by Larry D. Travis and Anthony Del Genio of the Goddard Institute for Space Studies.

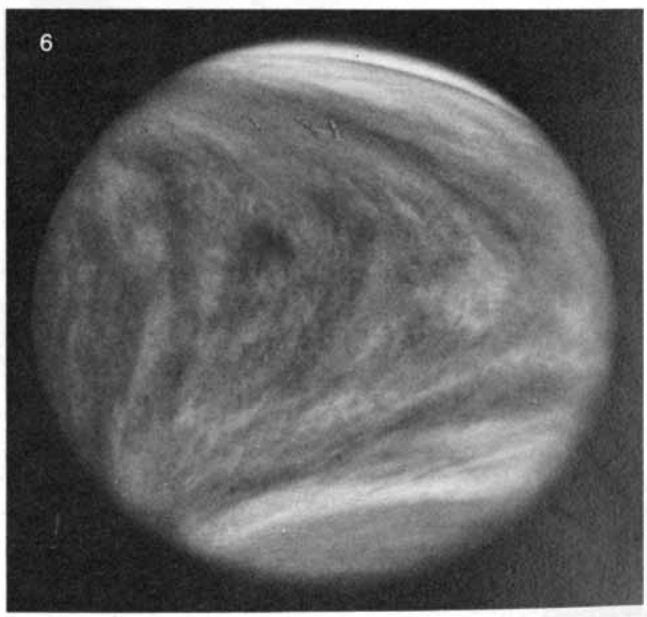






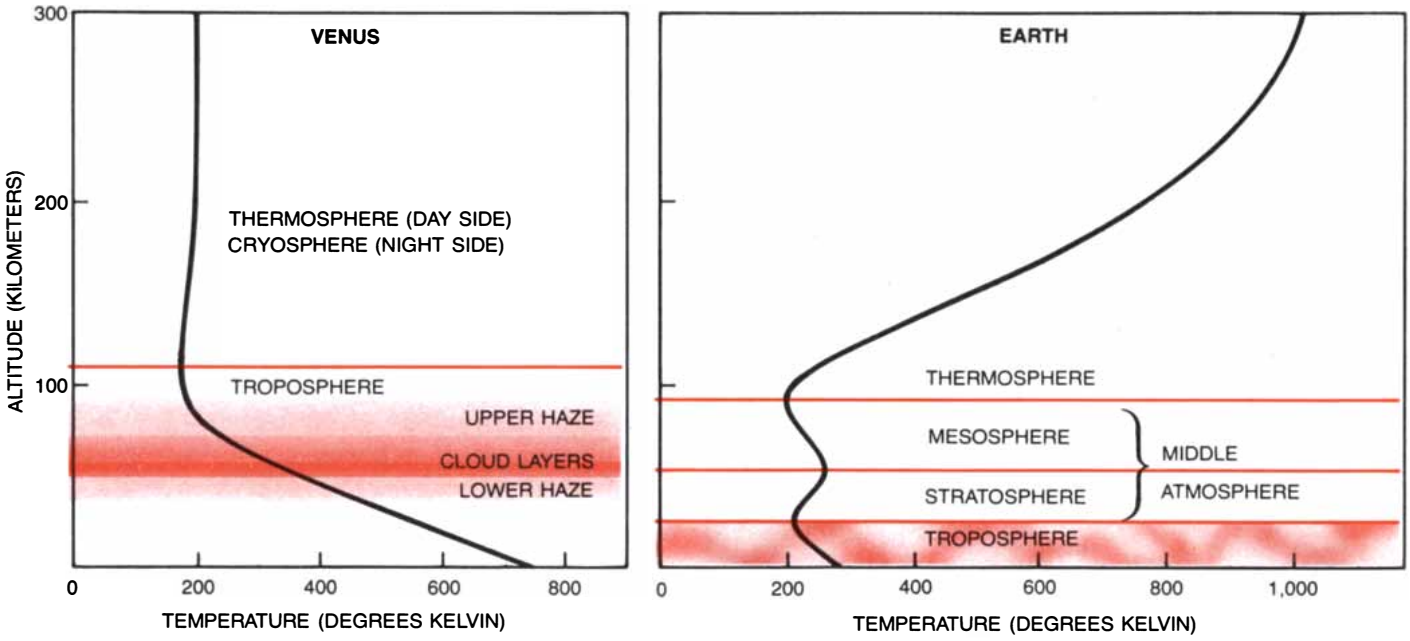




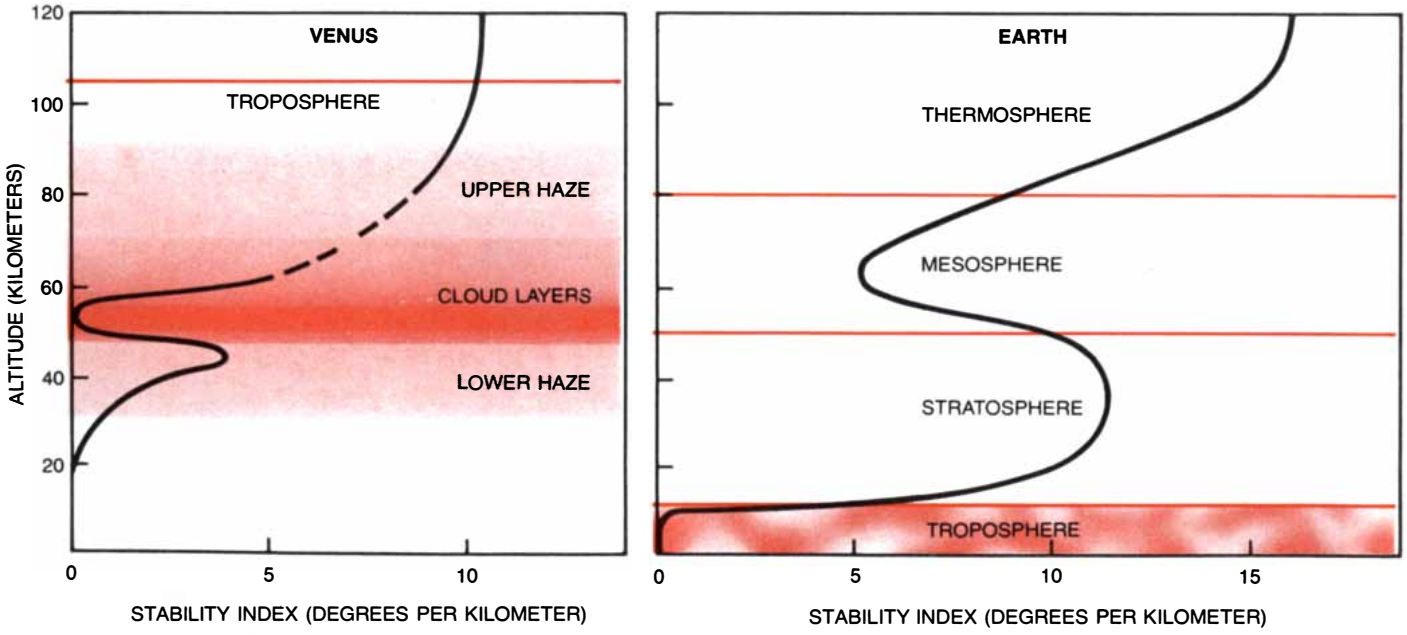


sphere of Venus is heated by the greenhouse effect, as has been proposed by Carl Sagan of Cornell University and James B. Pollack of the Ames Research Center of the National Aeronautics and Space Administration.

Above the troposphere lies a thinner upper atmosphere. In daylight the upper atmosphere is heated directly by ultraviolet radiation from the sun and the temperature therefore increases with height. Such an increase is also observed in the upper atmosphere of the earth, which is accordingly called the thermosphere. On the earth the thermosphere is present day and night; the large-scale rotation of the atmosphere with the planet carries the heated day-side upper atmosphere to the night side of the planet. On the night side of Venus, however, the thermosphere disappears; the upper at-

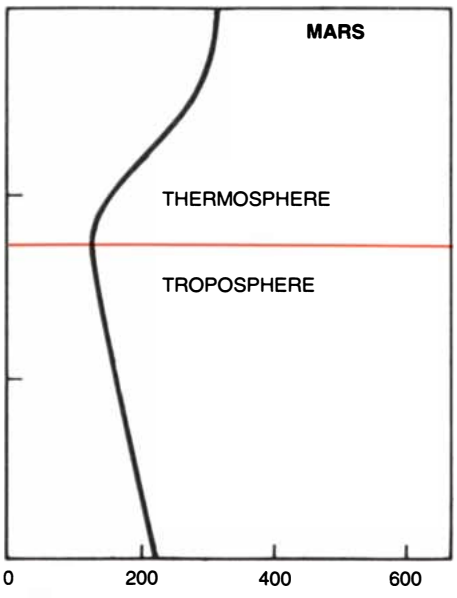


TEMPERATURE PROFILES of the atmosphere of Venus, earth and Mars are compared. The atmosphere of Venus (*left graph*) has two subdivisions. In the lower one, called the troposphere, the atmosphere is heated mostly by the greenhouse effect (the trapping of the heat the surface radiates). Hence the temperature decreases with altitude. In the upper subdivision, called the thermosphere, the atmosphere is heated directly by the absorption of solar radiation; the temperature increases with altitude. The profile for Venus is an average of daytime and nighttime measurements; actually the Venusian thermosphere could be said to disappear at night because the temperature of the atmosphere then decreases monotonically with altitude. The atmosphere of Mars (*right graph*) has the same two subdivisions

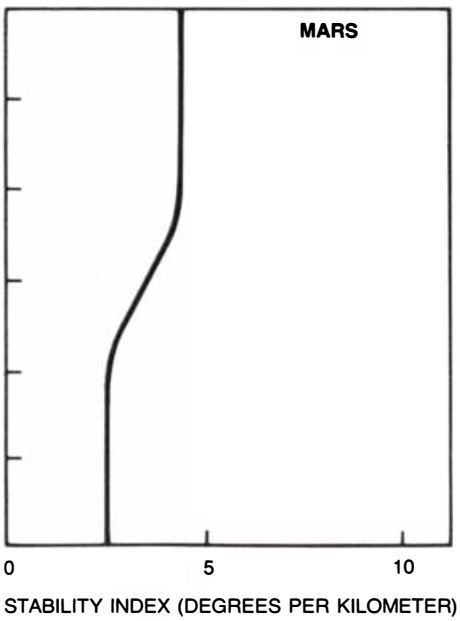


STABILITY PROFILES of the atmosphere of Venus, the earth and Mars indicate the degree to which the atmosphere at a given altitude resists convection. Basically a parcel of the atmosphere at any given altitude must support the weight of the atmosphere above it. Thus atmospheric pressure decreases with altitude. A parcel displaced upward will therefore expand, and as a result it will cool. If the gradient of adiabatic cooling (the cooling caused solely by the reduction in pressure) exceeds the atmosphere's temperature gradient, the displaced parcel will be cooler than its new surroundings and will sink back toward its original level. In that circumstance the atmosphere is stable against convection. If the adiabatic gradient is less than the temperature gradient, the displaced parcel will still be warmer than its surroundings. It will be buoyant and will continue to rise. The atmosphere is therefore unstable. The illustration plots the difference between the adiabatic gradient and the temperature gradient. Positive numbers indicate stability; zero suggests instability. (A negative value would not persist because convection would set in and tend to restore the balance.) On the earth the stability of the stratosphere confines

mosphere cools quickly after sunset to temperatures far below those of the troposphere. The nightly disappearance of the thermosphere of Venus was discovered on the Pioneer Venus mission. It has not yet been explained. The upper atmosphere of Venus rotates fast enough



as the atmosphere of Venus. The atmosphere of the earth (*middle graph*) is more complex than the others; it has a middle subdivision, in which the temperature profile attains a local maximum as a result of the absorption of ultraviolet radiation by the layer of ozone there.



convective mixing (and therefore clouds) to the troposphere. On Venus the pattern is similar, although a moderately stable layer lies subjacent to the clouds. On Mars the atmosphere is too thin for convective instability to develop. The data for the atmosphere of Venus and Mars were recorded by spacecraft. Data for the earth were provided by Richard Walterscheid of the Aerospace Corporation.

to carry a large amount of heat to the night side of the planet, but the heat is somehow lost. In any event we have suggested that the night-side upper atmosphere, which is the coldest part of Venus, be called the cryosphere.

The basic difference between the atmosphere of Venus and that of the earth is that the atmosphere of Venus is hot at the bottom and cold at the top, whereas on the earth the reverse is true. The lower atmosphere of the earth is not massive enough to sustain a large greenhouse effect. A second difference is that the earth's atmosphere has a middle region in which the temperature rises to a local maximum. The heating there results from the absorption of ultraviolet radiation by a layer rich in ozone.

In the troposphere of Venus horizontal variations in temperature are much smaller than the vertical variations. The greatest latitudinal differences the Pioneer probes detected were at the level of the upper clouds. Here the north probe, which entered the atmosphere at a latitude of 60 degrees, measured temperatures some 10 to 20 degrees lower than those measured by three other Pioneer Venus probes. The highest temperatures were measured by a probe that entered the atmosphere near the equator. The explanation of this pattern on Venus is the same as it is on the earth. In the equatorial region of each planet the sun's radiation enters the atmosphere at angles most nearly perpendicular to the surface. Hence the deposition of solar energy per unit area is greatest near the equator. On Venus the incident solar energy is absorbed mostly in the upper regions of the clouds.

Between an altitude of 70 kilometers and the beginning of the upper atmosphere at 100 kilometers the temperature gradient from the equator to the pole reverses: at these heights the polar regions are generally warmer than the equator. The reversal, discovered during the Pioneer Venus mission, has not been explained. A similar reversal of the temperature gradient is known to exist on the earth, where the winter pole is the warmest part of the middle atmosphere. There too the cause is not understood.

The variations of temperature with respect to longitude in the troposphere of Venus are much smaller than the variations with respect to latitude. This was shown by the measurements made by two other probes of the Pioneer Venus mission, the day and night probes, which entered the atmosphere at the same latitude (30 degrees south) but were separated by more than 110 degrees of longitude. The temperatures measured by these probes at any given altitude never differed by more than five degrees C. In the lowest 10 to 20 kilometers of the atmosphere the smallness of this difference results from thermal inertia; the lower atmosphere is so massive that it retains most of its heat even during the long period it spends in darkness. At higher altitudes in the troposphere the rapid east-to-west circulation carries heat around the planet and reduces the temperature gradient.

Although the horizontal temperature differences are small, they must nonetheless be the source of the forces that drive atmospheric motions. In this respect the latitudinal gradient is particularly important. The solar energy entering the atmosphere of Venus warms the equator more than it warms the poles. On the other hand, the energy the planet emits in the infrared has virtually the same intensity at all latitudes because of the relatively small variation of temperature with respect to latitude. The imbalance between the incoming solar energy and the outgoing infrared energy would rapidly cool the poles and heat the equator were it not for large-scale motions of the atmosphere that carry heat from the equator toward the poles. Such motions are found on the earth, although the imbalance between the incoming and the outgoing radiation at the poles is less than it is on Venus.

How does heat move on Venus from the equator toward the poles? One would expect that a given parcel of the lower atmosphere in the equatorial latitudes of Venus would rise. Its place would be taken by adjacent parcels flowing in from the north and the south. In this way a steady current would ultimately become established, in which the atmosphere flows poleward at high altitude, bearing excess heat. After warming the polar regions the current would sink and then flow back toward the equator at a lower level.

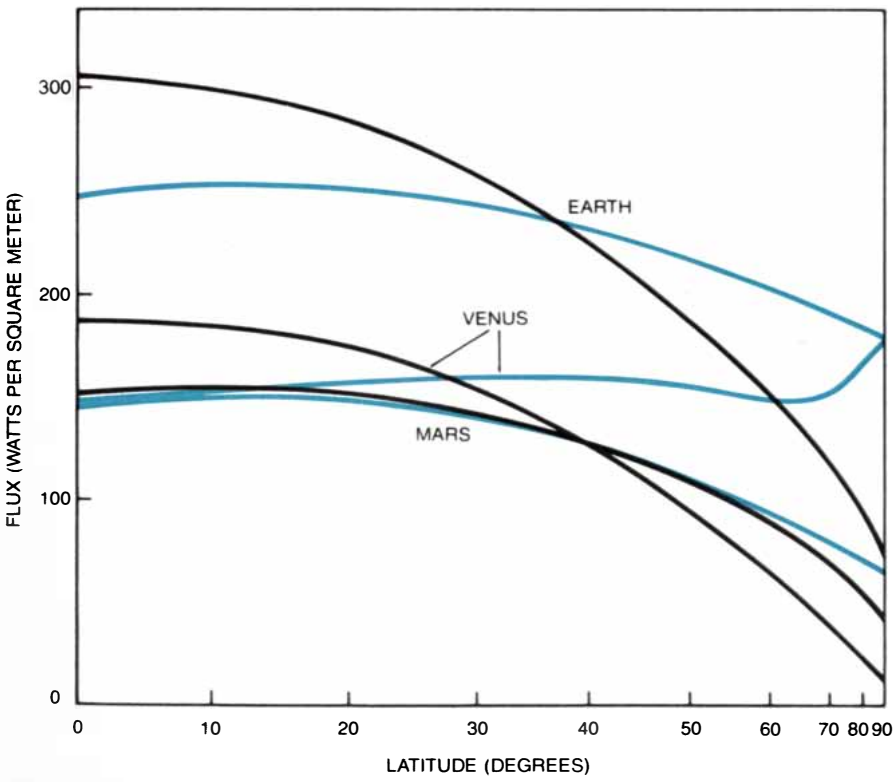
North-south circulation of this kind is known as a Hadley cell. The evidence suggests that Hadley cells have formed on Venus at the level of the cloud deck. In particular, ultraviolet photographs made by the U.S. *Mariner 10* spacecraft when it flew by Venus in 1974 showed poleward motion at the tops of the clouds, whereas measurements made by all the Pioneer Venus probes showed winds toward the equator near the base of the clouds. A Hadley cell would lie at cloud heights on Venus because most of the incident solar energy is absorbed there. The cell would extend all the way from the equator to the pole because the slow rotation of the planet does little to deflect the circulation from its north-south trajectory. On the earth a Hadley cell lies just above the surface, the place where most of the solar energy is absorbed. Because the earth rotates rapidly the terrestrial Hadley cell extends no farther than the mid-latitudes, where the poleward transport of heat is taken over by complicated wavelike motions called baroclinic eddies.

A small fraction of the solar energy reaching Venus is absorbed at the sur-

face, and since this energy is deposited predominantly at the equator, it too must be transported toward the poles. It is unlikely, however, that the cloud-level Hadley cell participates in this transport. The most important factor in analyzing the Hadley-cell circulation is the adiabatic gradient, which represents the cooling or heating a parcel of the atmosphere undergoes as a result of expansion or contraction when it is displaced upward or downward. If a parcel is displaced upward, for example, it enters a region where the pressure is lower, and it therefore expands and cools. In essence the heat in the parcel is distributed throughout a larger volume. The parcel may now be colder than the atmosphere around it. In that case it will be heavier than its surroundings and will tend to sink back to its original level. Under this condition the atmosphere is stable. On the other hand, the parcel may still be hotter (in spite of the adiabatic cooling) than the atmosphere around it. In that case it will continue to rise. The atmosphere is then unstable and is susceptible to overturning and the transport of heat by convection.

In evaluating the stability of an atmosphere one therefore compares two quantities: the rate at which temperature changes with altitude and the rate of adiabatic cooling. Wherever the latter is the smaller the atmosphere is unstable. When the atmosphere of Venus is analyzed in this way, a stable layer some 20 kilometers thick is found to lie immediately below the clouds. Since vertical motions would be suppressed in this layer, it is doubtful that a single Hadley cell could extend from the surface to the clouds.

A separate Hadley cell might operate near the surface, but the presence of such a cell is still a matter for speculation. The north-south winds measured deep in the atmosphere by the Pioneer Venus probes are small; they have magnitudes of only a few meters per second and their directions reveal no large-scale pattern. If a Hadley cell does lie deep in the atmosphere, the mean north-south circulation in the atmosphere of Venus would have to consist of at least three Hadley cells stacked one on top of the other because the cloud-level cell and the deep cell would drive at least one counterrotating cell between them. If there is no deep Hadley cell, the solar energy that penetrates to the surface must be carried poleward by the net effect of eddy motions.



RADIATION BALANCE between incoming and outgoing solar energy is attained on Venus, the earth and Mars because the atmosphere of each planet transports heat from its equator to its poles. At the equator the amount of radiation each planet receives from the sun (*black lines*) exceeds the amount it emits in the form of infrared radiation (*colored lines*). The excess is carried toward the poles. At the poles the amount of radiation the planet emits would exceed the amount it receives if it were not for the poleward flow. Because the radiation balance has not been measured extensively on Mars the curves for that planet are hypothetical. The difference between the curve representing incoming radiation and the curve representing outgoing radiation on Mars is small because the atmosphere of Mars is too tenuous to hold much heat.

Since Venus rotates slowly, one might expect north-south circulation to constitute the totality of the atmospheric circulation. For one thing, seasonal differences on the planet should be insignificant. Venus' elliptical orbit deviates only slightly from a circle, so that the planet maintains an almost constant distance from the sun. Moreover, the planet's axis of spin is almost perpendicular to the plane of the orbit. At any latitude, therefore, the deposition of sunlight changes only slightly through the year.

Surprisingly, the expectation of a simple pattern of circulation in the atmosphere has been contradicted by the evidence. The mean north-south circulation is overshadowed by the far stronger east-to-west wind, which begins some 10 kilometers above the surface and extends to heights of 90 kilometers or more. Up to the height of the clouds the wind speed increases with altitude. It reaches a velocity of 30 meters per second at an altitude of 30 kilometers and 100 meters per second at cloud heights of 60 kilometers. At the latter speed the wind gives rise to a four-day superrotation of the ultraviolet markings in the clouds. At cloud heights, moreover, the superposition of the westward circulation and the weaker Hadley circulation gives rise to a vortex that converges at each pole. The vortex was first recognized in a composite photograph of the south pole that Verner E. Suomi and Sanjay S. Limaye of the University of Wisconsin at Madison constructed in 1978 from ultraviolet photographs of Venus made by *Mariner 10* at latitudes nearer the equator.

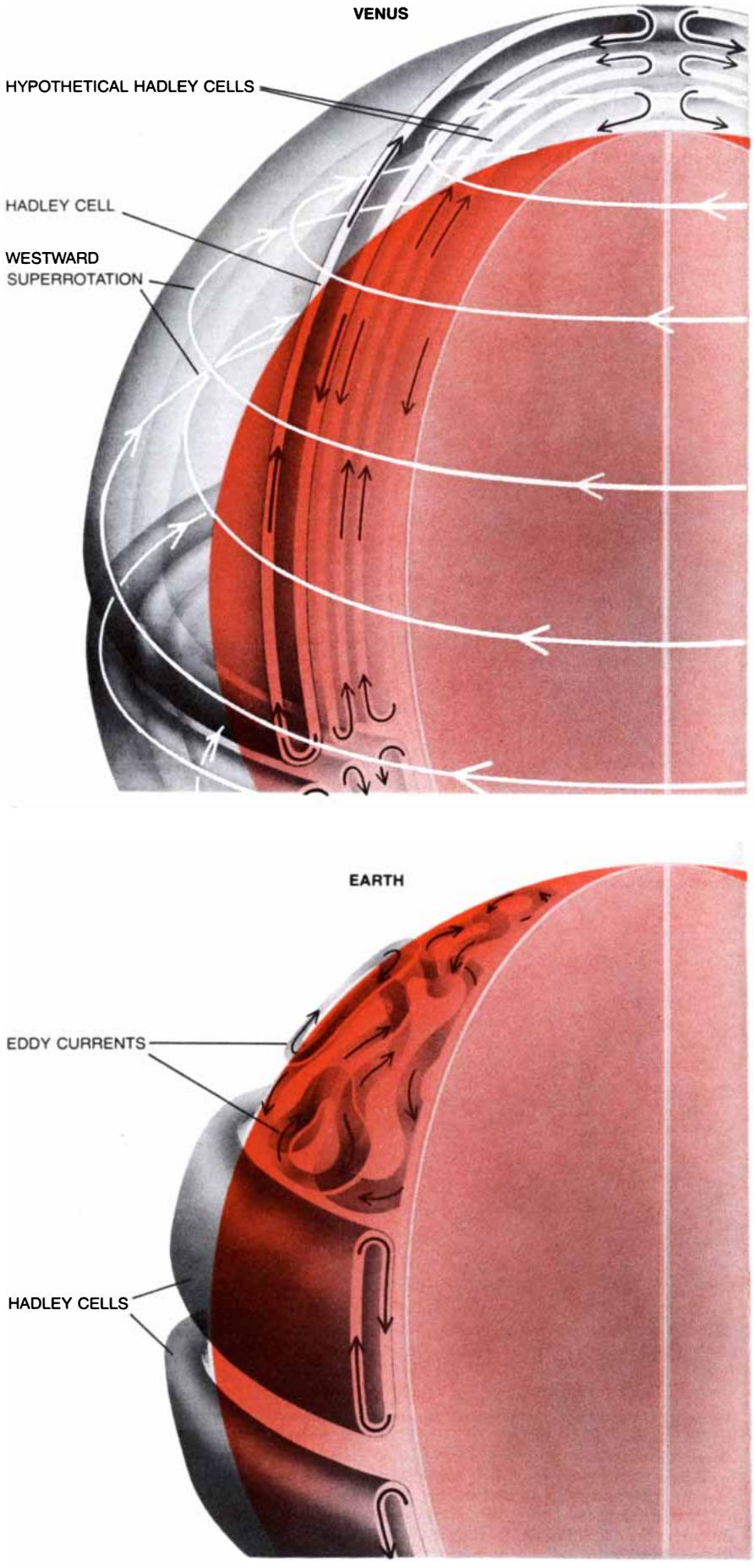
The superrotation had first become apparent when observations from the earth showed that large-scale features in the atmosphere of Venus circle the planet about once every four days. The features could be seen in ultraviolet photographs of the planet. Until the Venera and the Pioneer Venus probes measured the wind speed, however, one could not be certain whether the movement of the ultraviolet features represented bulk motion of the atmosphere or was instead the propagation of a wave—a sloshing, so to speak—in an otherwise placid atmosphere. It is now generally accepted that the large-scale features are caused by planetary waves but that the sloshings propagate slowly with respect to the rotation of the atmosphere. What propels the features around the planet is indeed the westward wind.

The precise nature of the planetary waves remains hard to determine. The problem is that the substances responsible for absorbing ultraviolet radiation and thereby making the large-scale fea-

tures visible have not all been identified. Specifically, the absorption of radiation at short ultraviolet wavelengths can be attributed to sulfur dioxide gas, but the absorption at longer wavelengths remains to be explained. The absorber at longer wavelengths cannot be sulfuric acid, the main constituent of the clouds, because sulfuric acid is transparent to ultraviolet radiation. The absorbers of the radiation may turn out to be concentrated below the cloud tops in areas of the planet that are bright in the ultraviolet. Vertical motions induced by the passage of a planetary wave could then give rise to contrast in the ultraviolet image of the planet by lofting the absorbers to a higher altitude, where their absorption of ultraviolet radiation would darken parts of the image.

The extent to which the four-day circulation persists with increasing height above the cloud tops is also not well known. The Pioneer Venus data suggest that the wind speed reaches a maximum of about 150 meters per second at an altitude of 70 kilometers and then decreases with height between 70 and 90 kilometers.

Above 150 kilometers the circulation of the atmosphere has been measured only by indirect methods. For example, a mass spectrometer on board the Pioneer Venus orbiter measures the density of substances such as nitrogen, carbon dioxide and helium high in the atmosphere. In addition the orbiter is tracked as it revolves about the planet. The slow decay of its orbit is caused by drag, which in turn depends on the density of

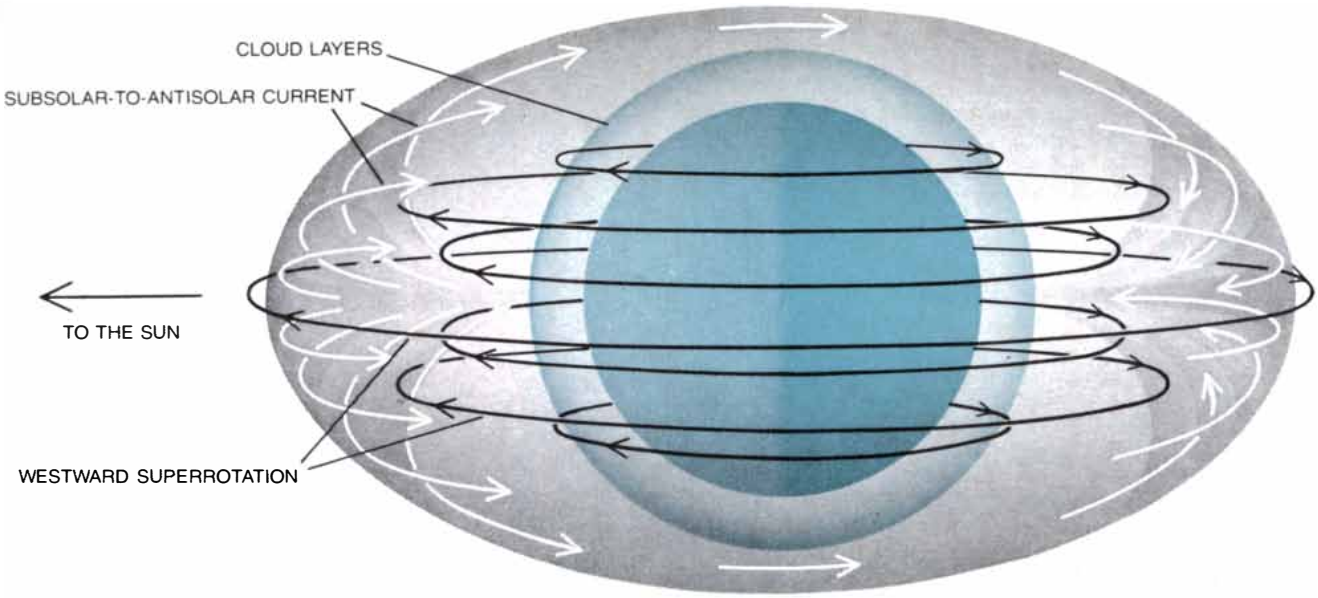


PATTERN OF WINDS in the lower atmosphere of Venus differs markedly from the pattern on the earth. On Venus (*upper drawing*) the data recorded by probes suggest that at cloud heights of roughly 60 kilometers the atmosphere circulates in a current that rises at the equator, travels poleward, sinks near the pole and returns to the equator. The pattern, which is called a Hadley cell, represents the poleward flow of the heat the sun deposits in the clouds. Two more Hadley cells are shown, but their presence is hypothetical. The cell nearest the surface may be needed to carry heat in the lower atmosphere; the stable layer just below the clouds seems to prevent the heat deposited near the surface from rising to cloud heights. If the bottom cell exists, there must be a middle cell in counterrotation between the other two. The north-south flow in the Hadley cells has a speed of only a few meters per second. Superimposed on this motion is the westward rotation of the atmosphere (*white lines*), which attains a speed of approximately 100 meters per second at the altitude of the clouds. On the earth (*lower drawing*) a Hadley cell lies near the surface. It extends only to mid-latitudes, where the rotation of the earth disrupts its north-south path. From there the poleward transport of heat is the net effect of ever changing patterns of eddies.

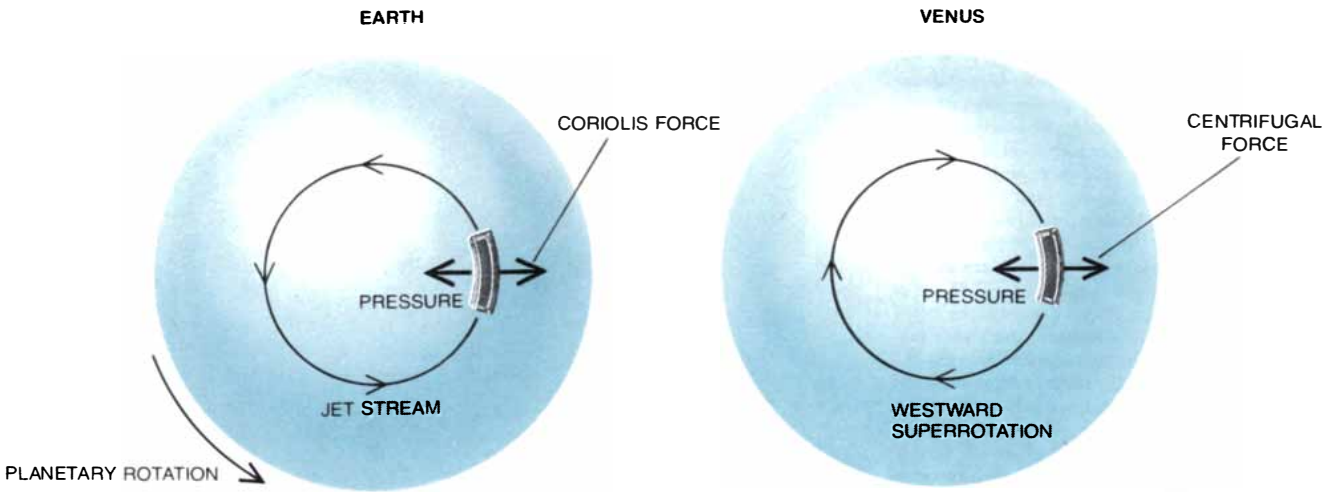
the atmosphere. By both of these methods it has been shown that the place of lowest density in the upper atmosphere lies well to the west of the point opposite the sun on the night side of the planet.

This finding suggests that an earlier hypothesis should be amended. Before the Pioneer Venus mission many investigators had expected that the main influence on the circulation of the upper atmosphere of Venus would turn out to be the heating of the part of the upper atmosphere facing the sun. On that hypothesis a parcel of the upper atmosphere would ascend in the region facing the sun, flow to the night side of the planet and descend. The overall pattern of circulation would thus be symmetrical about a line connecting the center of the sun with the center of Venus. The Pioneer Venus measurements imply a more complex pattern, in which a westward superrotation may deflect the descending limb of the circulation.

Why is the dominant circulation at all altitudes a westward superrotation? That remains the greatest mystery about



WINDS IN THE UPPER ATMOSPHERE of Venus are hypothesized to result from the superposition of two basic flows: the westward rotation of the atmosphere (*black lines*) and a symmetrical circulation (*white arrows*) in which the winds ascend on the day side of the planet and descend on the night side. Although the winds in the upper atmosphere have not been measured directly, the superposition shown is supported by measurements indicating that the density of the upper atmosphere is lowest at a point on the night side of the planet well to the west of the point opposite the sun. It is as if the sinking current were being deflected by the westward current.



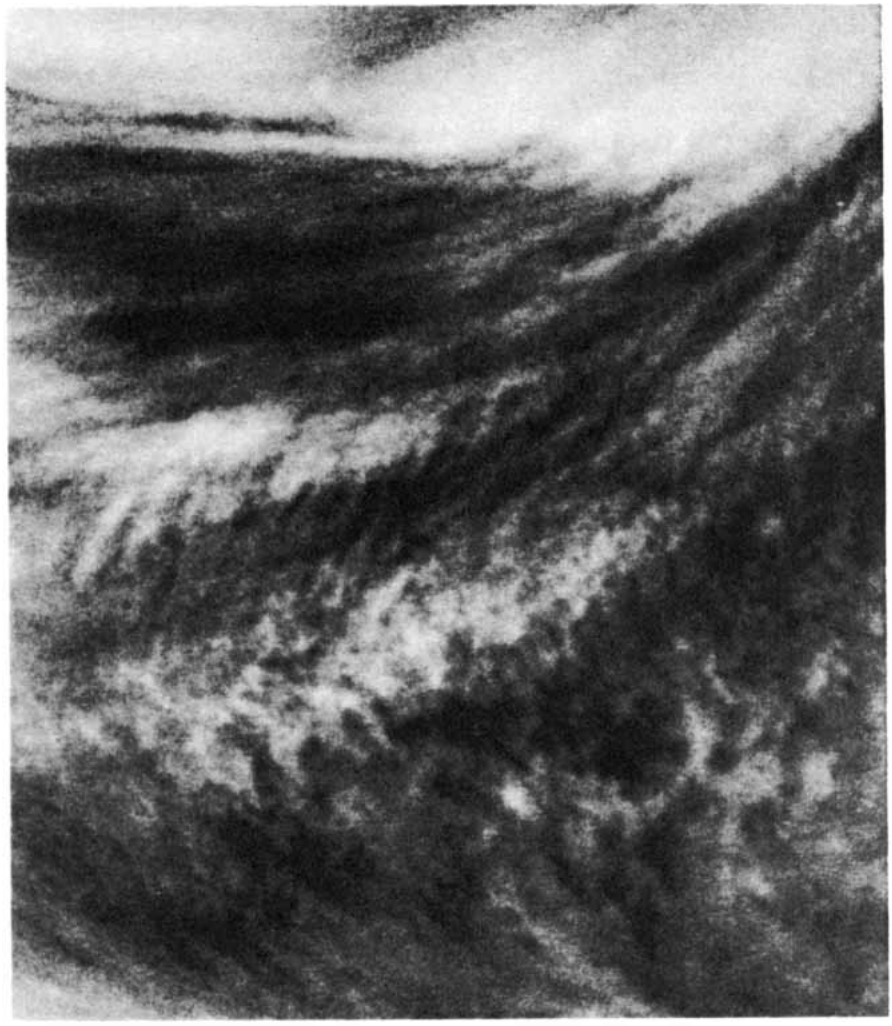
BALANCE OF FORCES that maintains atmospheric motions on a rapidly rotating planet (such as the earth, Mars, Jupiter or Saturn) differs from the balance on a slowly rotating planet such as Venus. In each case the greater deposition of heat at the equator of the planet establishes a pressure gradient that drives the atmosphere toward the pole. On the earth (*left*) the pressure is opposed primarily by the Coriolis force: the deflection of a parcel of the atmosphere at a right angle to its trajectory because it is on a rotating sphere. The earth's jet stream, for example, is supported by such a balance; the Coriolis force drives the jet stream toward the Equator and counteracts the pressure that drives it toward the pole. The condition is called geostrophic balance. On Venus (*right*) the Coriolis force is negligible. Instead the motion of a parcel circling the planet as part of the westward rotation of the atmosphere subjects it to a centrifugal force, which balances the poleward pressure by pushing the parcel away from the planet's spin axis. The condition is cyclostrophic balance.

the atmosphere of Venus. The essential difficulty is that one expects an atmosphere to rotate more or less in step with the solid planet below it; if it rotates faster, one must imagine a mechanism by which angular momentum is transported upward from the surface of the planet. In effect the surface of the planet must push the atmosphere.

A mechanism that might account for the upward transfer of angular momentum is the moving-flame effect. Imagine a flame under a vessel holding a layer of fluid, and imagine initially that the flame stands still. Over a period of time a convection cell develops, in which the fluid directly over the flame rises and the fluid around this upwelling sinks. Since the flame is stationary, both the ascending and the descending limbs of the convection cell are vertical. Now imagine that the flame is moved continuously from left to right under the layer. The bottom of the cell will come to lie to the right of the top. Hence the limbs of the cell will be tilted: the upwelling fluid will move to the left as it rises and the descending fluid will move to the right as it sinks. In laboratory experiments employing a ring of fluid it has emerged that these effects reinforce each other, and the entire fluid flows to the left. The fluid flows, then, in a direction opposite to the direction in which the flame is moved.

On Venus, which rotates toward the west, the sun moves toward the east in the sky. A moving-flame effect caused by this eastward motion of the sun might therefore build up the westward superrotation. To be sure, the sun is above the atmosphere of Venus whereas the flame is under the fluid layer. On Venus, however, the situation is complicated by the stable layers of the atmosphere, which tend to resist the formation of convection cells. An example is the layer below the clouds. In a calculation one of us (Schubert) made with Richard E. Young of the Ames Research Center, the hypothetical convection cells on Venus turn out to be not merely tilted: their vertical profile has the shape of a boomerang. According to the calculation, it is conceivable that such cells could impart to the atmosphere a net westward rotation.

Another hypothesis, suggested by Peter J. Gierasch of Cornell University, is that the Hadley circulation transports angular momentum high into the atmosphere. In this view a parcel of the atmosphere that rises at the equator takes with it an angular momentum corresponding to that of the surface at the equator, and as the parcel moves poleward in a Hadley cell it gives up some of its momentum to the surrounding atmosphere. When the parcel returns to the surface, it regains momentum, and so the cycle continues. Perhaps the cycle works in spite of a stacking of three or more Hadley cells. An analysis of ultraviolet photographs made during the *Mariner 10* flyby of Venus shows that at latitudes of from 40 to 45 degrees the cloud markings were moving westward in 1974 somewhat faster than they would move if the atmosphere were rotating rigidly, that is, like a solid body. It has been suggested by William B. Rossow and his colleagues at the Goddard Institute for Space Studies that angular momentum carried upward by the Hadley circulation might have built up this jet and that the turbulence around the jet might disperse the angular momentum throughout the atmosphere. In the Pioneer Venus data, however, no such jet has been found.



SMALL ATMOSPHERIC WAVES are visible in ultraviolet photographs made by the Pioneer Venus orbiter. Here the waves are a file of dark diagonal lines at the center of the image. Each line is roughly 1,000 kilometers long and 200 kilometers from its neighbor on each side.

One thing is certain: the atmosphere of Venus responds to a balance of forces fundamentally different from that of the atmosphere of the earth and Mars, or for that matter of Jupiter and Saturn. The reason is the slow rotation of Venus. On any planet the poleward transport of heat by winds corresponds to a gradient in atmospheric pressure: the pressure is greater at the equator than it is at the poles. On a rapidly rotating planet such as the earth the resulting poleward drive on a given parcel of the atmosphere is balanced by the Coriolis force: the deflection of a parcel of the atmosphere at a right angle to its trajectory because it is on a rotating sphere. (To an observer some distance from the planet it would be apparent that the deflection corresponds to the rotation of the planet's coordinate grid of latitude and longitude below the wind. To an observer in motion with the planet it is the wind that seems to be deflected.) The Coriolis force on a given parcel is proportional to the velocity of the parcel.

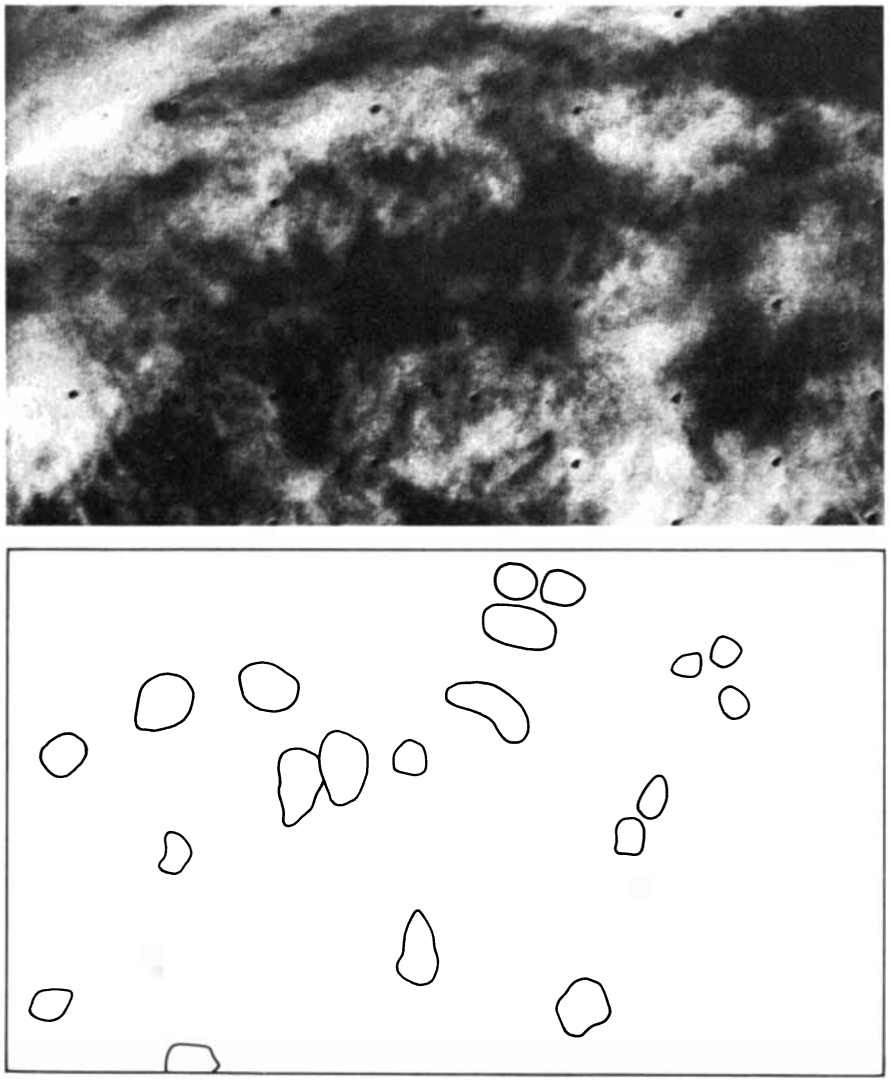
Since Venus rotates slowly, the Coriolis force is negligible. Instead, as was first noted by Conway B. Leovy of the University of Washington, the poleward drive on a parcel of wind that circles the planet from east to west as part of the global superrotation of the atmosphere

is balanced by the centrifugal force that tends to push the parcel away from the spin axis of the planet, and therefore toward the equator. Because the centrifugal force depends on the square of the velocity of the parcel, it is fundamentally different from the Coriolis force.

The ultraviolet photographs of the atmosphere of Venus reveal a variety of atmospheric markings, which presumably will find explanation in accord with the unusual balance of forces there. The largest of the markings are the planet-wide features that first revealed the four-day westward circulation. We noted above the association of such features with planetary waves. Doubtless the waves propagate vertically as well as horizontally; they can therefore carry momentum upward in the atmosphere, and so they may have a role in fostering the four-day circulation. On a smaller scale some of the Pioneer Venus photographs have shown trains of waves. In one case the waves consisted of dark bands about 1,000 kilometers long separated from one another by about 200 kilometers. The waves were found in the southern mid-latitudes of Venus and were arrayed at a large angle to the planet's parallels of latitude.

Polygonal features are apparent in some of the ultraviolet photographs. They are undoubtedly convection cells. Until it is possible to identify all the atmospheric absorbers of ultraviolet radiation that modulate the ultraviolet images, it cannot be said whether the dark core of each such cell is the place where the atmosphere is upwelling and the bright rim of the cell is the place where the atmosphere sinks. The reverse might be the case. In any event the convection cells are seen mostly in the part of the atmosphere that directly faces the sun or else to the west (that is, downwind) of that region. The absorption of solar radiation there is apparently sufficient to induce convection near the tops of the clouds. An analysis of the stability of the atmosphere based on measurements made by the Pioneer Venus probes that entered the atmosphere in the early morning shows that a layer about 10 kilometers below the cloud tops is unstable and capable of convective overturning. This unstable layer must grow thicker throughout the morning, so that by afternoon the tops of the convective cells are visible from above the clouds.

Waves and convection cells are two examples of eddy motions: short-term variations superimposed on the mean state of the atmosphere. Such motions are present at all spatial scales and at every altitude. Evidence of them is found not only in the ultraviolet photographs but also in the records of temperature and wind velocity made by the Pioneer Venus probes. The atmosphere of Venus is also surprisingly variable on time scales as long as months or years. An example of the long-term variability is an apparent change in the dominant westward circulation. Ever since the Pioneer Venus mission entered the atmosphere of Venus in December, 1978, the atmosphere at cloud level has been rotating like a rigid body, but at the time of *Mariner 10* some four years earlier the mid-latitudes were rotating faster than the rest. (This constituted the jet mentioned above.) It is difficult to understand what might be responsible for such a change when one considers the steadiness of the influx of solar radiation on Venus and the gentle undulation that is now known to characterize the surface of the planet.



BRIGHT-RIMMED CELLS on the day side of Venus are visible in an ultraviolet photograph made by the *Mariner 10* spacecraft when it flew by the planet in 1974. Each cell is a few hundred kilometers across, and some cells are distinctly polygonal. They are thought to be convection cells: places where the atmosphere overturns. The overturning is presumably driven by solar energy absorbed near the top of the clouds. Further knowledge awaits the identification of all the trace components of the atmosphere whose absorption of radiation gives rise to the ultraviolet markings. The photograph was provided by Michael J. S. Belton of the Kitt Peak National Observatory. The accompanying map shows the positions of prominent convection cells.

The exploration of the atmosphere of Venus has already revealed its high temperature near the surface, its westward superrotation, its cloud-level Hadley cell, its polar vortexes, its night-side cryosphere and its warm poles above the clouds. A collaboration of Russian and French groups is planning to place a balloon well below the clouds of Venus in 1984. The balloon will be tracked as the east-to-west wind sweeps it across the daylight side of the planet. A mission planned by the U.S. for the mid-1980's may carry instruments to gather data on the upper atmosphere, which the Pioneer Venus orbiter is also continuing to probe. The exploration of the atmosphere of Venus has already been rewarding beyond expectation, and the findings that remain unexplained surely warrant further effort.

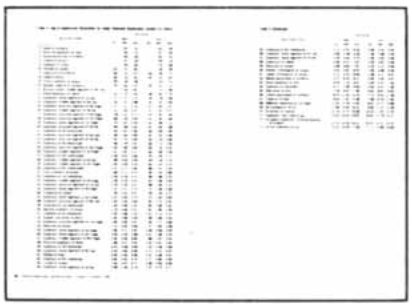
The preceding pages of this magazine best illustrate what this ad is selling

"Don't underestimate the intellectual capacity of the reader" has been the winning strategy of *Scientific American.*

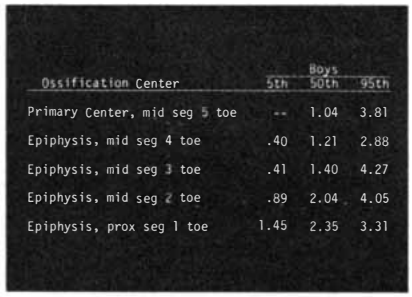
Yet you can get just about everything out of every article—subtleties and all—without the need to read one word of text! It's all in the illustrations.

Our message here to all whose progress depends on the understanding of others about complex matters: *Good graphics are worth your time and effort.* Words are often not enough. The purpose of the slides is not just to get you to use the slide projector. They are not just a formality, like wearing shoes for making the presentation. Slides that make an audience squirm result from treating the graphics as an afterthought.

Your text, let us say, is all ready. Ah yes, slides. Here, in discussing age of ossification, let's have a slide of the pertinent table.
So:



Only when you are in front of the audience does it hit you that this was not a good idea. A better idea would have been to select a part of the table that makes a memorable point and yet contains no more than can be read without optical aid from the slide held in the hand. This calls for retyping within a 2″ x 3″ (or at most a 3″ x 4½″) rectangle, which in itself sets the limit on how much to pack into a single slide. Photograph the typing with KODALITH Ortho Film, Type 3.



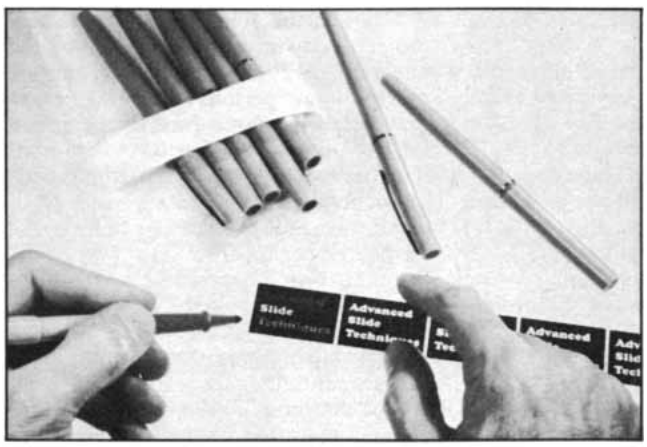
Ossification Center	Boys 5th	50th	95th
Primary Center, mid seg 5 toe	--	1.04	3.81
Epiphysis, mid seg 4 toe	.40	1.21	2.88
Epiphysis, mid seg 3 toe	.41	1.40	4.27
Epiphysis, mid seg 2 toe	.89	2.04	4.05
Epiphysis, prox seg 1 toe	1.45	2.35	3.31

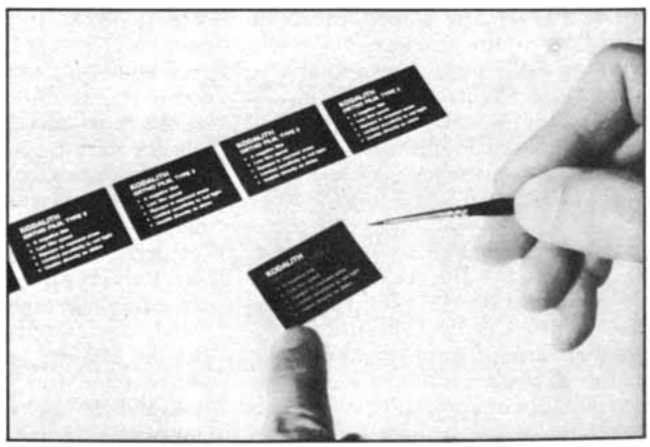
Now those who need to know about the ossification of the human toe can see that males reach the 50th percentile for the third toe at age 1.4 years. (The editors and artists of *Scientific American* would probably have brought it out more vividly with a clever graph, but the rest of us work under sterner limitations.) It's in black-and-white, though. Since most photographs seen on a screen are in color, interposition of a black-and-white slide like this can act as a hidden cue to turn off the audience's attention. Watercolor dye applied to your black-and-white slide with a small brush or cotton swab will help.

There are more directly functional uses for added color in diagrams and tables, as this magazine aptly demonstrates. This is beginning to be widely understood. A large number of the questions we get at meetings of professional societies concern nothing more profound than tips and techniques for preparing more eloquent slides. They are asked by folks whose operating budgets don't allow for calling in professionals who do such work. They find it wise to take time from their own professions to learn the details for doing it well.

Our Publication No. M3-106, "Making Lecture Slides," gets right down to such details as how to avoid those distressing bands on slides made from CRT monitors, how to use the exposure meter on an SLR camera in copying, how to combine several images for side-by-side comparison on a single slide. Single copies free from Department 55-W, Kodak, Rochester, N.Y. 14650.

Kodak





SCIENCE AND THE CITIZEN

Demographic Disadvantage

What price will the children of the baby boom pay for having been born into an exceptionally large cohort? Some cohorts, or groups of chronological peers, are favored by demographic and economic circumstance; some suffer. The members of the relatively small group born in the 1930's and the early 1940's, for example, sought their first jobs in the expansive postwar economy. By the end of their careers the most favored among them, who entered the job market in 1962, will have amassed the largest lifetime earnings of any group since the one that began working in 1940. The huge cohort born in the years after World War II has been less fortunate. When its members entered the labor market in the 1970's, their numbers intensified the competition for scarce jobs. Many remained unemployed for a time; some took jobs in a field other than the one for which they had been trained. The most demographically disadvantaged, who were graduated from high schools and colleges in 1977, will have the lowest lifetime earnings of any group from 1940 to 1990.

In a recent article in *Population and Development Review* James P. Smith of the Rand Corporation and Finis Welch of the University of California at Los Angeles argue that almost all the reduction in earnings can be attributed to the extraordinary size of the baby-boom cohort. This group, flanked by much smaller cohorts, is a pivotal one demographically; how its members fare is important to society as a whole. Smith and Welch suggest that the prospects are not as bleak as they now seem.

The differences in the size of successive cohorts are a result of cyclical fluctuations in fertility. In 1920 the crude birthrate, or number of births per 1,000 population per year, stood at about 25. A steep decline that began soon afterward reduced the rate to less than 20 by the late 1930's. The sharp upsurge in fertility during the postwar baby boom peaked in 1957 at a rate comparable to that of 1920. The decline that followed was steep, and by the mid-1970's it had brought the crude birthrate down to about 15, the lowest level since 1900.

A change in fertility is felt in the labor market about 20 years later. The main impact of the change on recent graduates lies in its influence on the ratio of experienced workers to beginners, Smith and Welch write. The more applicants there are for entry-level positions, the more difficult it is to find such a job. An abundance of applicants also tempts employers to hire a large number of novice workers at low wages. Since much of a job is learned from older, more experienced co-workers, increases in the proportion of younger employees work to their disadvantage by reducing the amount of time senior colleagues can spend with them.

The ratio of older workers to novices has fallen sharply in the U.S. since World War II. In 1954, when the ratio was at its peak, there were nine experienced workers (35 or older) for each one under 25. By 1977 there were fewer than four. The competitive situation has also been altered by an increase in the number of college-educated people in the work force. In 1952 there were fewer than half as many employees with college degrees as there were workers with elementary schooling only. By 1977 workers with college degrees outnumbered those with elementary schooling by more than three to one.

These trends are reflected in the wages of young jobholders. In 1967 the annual earnings of a senior employee who was a college graduate were 1.6 times those of a new graduate. Ten years later older workers were making twice as much as young ones. The earnings of high school graduates have followed a similar course. The effect is more one of unemployment, however, than it is one of reduced wages. The wages of workers with a high school education did not fall as much as those of the college-educated, but the high school graduates worked fewer weeks and so the decrease in annual earnings was about the same.

On the basis of wage rates and data from the Bureau of the Census on the size of age groups Smith and Welch have estimated the proportion of wage reductions that should be attributed to demographic changes. They find that increasing numbers of new job applicants accounted for more than 80 percent of wage decreases between 1967 and 1975. The total reduction in earnings for high school graduates was 12 percent, of which 10 percent was an effect of cohort size. For college graduates the overall reduction was 16 percent, and of that amount 13 percent was a result of numbers alone, independent of other economic trends. (All the monetary calculations have been adjusted to compensate for the effects of inflation.)

Discouraging as this picture is, it is not hopeless, since the effects of being a member of a large cohort apparently do not persist past the beginning of a career. Relying on the correlation between age-group size and earnings, Smith and Welch carried out a simulation to predict the lifetime earnings of all groups who entered the job market between 1940 and 1990. The least favored group, that of 1977, will have lifetime earnings about 10 percent less than those of the best-situated cohort of 1962.

There are several reasons for the effect of large numbers to be concentrated at the start of a career. Unemployment is highest in the first few years after graduation. In addition the wage reductions caused by cohort size are largest during the period of professional apprenticeship, according to the analysis by Smith and Welch. Even in highly skilled professions the learning period is shorter than the period spent as an experienced worker. The pool of novices is thus much smaller than that of older employees. The effect of a large cohort is felt quite sharply by the small group of young workers, but as the members of the outsize group reach higher positions they compete with the larger field of senior workers and their numbers no longer hold wages down as drastically. This mechanism helps to explain why college graduates feel the effect of an increase in cohort size more strongly than high school graduates do: most jobs taken by college graduates require a relatively long time to learn, extending the period when cohort size is critical.

As the members of the large postwar cohort leave their first jobs and begin to fill senior positions, their salaries will improve with respect to those of their older colleagues, with whom they will be in more direct competition. As a group, however, they will never quite catch up with the demographically well-born. By the time they end their careers the next small cohorts, those born during the period of lowered fertility in the 1970's, will be in the middle stage of their working lives. This group, according to the Smith and Welch simulation, will achieve lifetime earnings equal to those of the favored labor-market entrants of 1962 and about 10 percent higher than those of the baby-boom cohort.

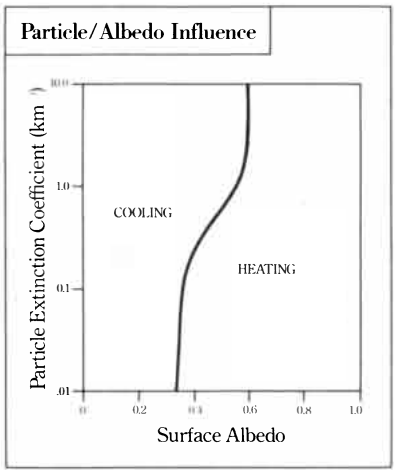
Resonant Explanations

Among the alignments and periodicities in the solar system that seem too precise to have arisen by chance, the one most easily observed is the coupling of rotational period to orbital period that keeps the same face of the moon turned toward the earth. A different kind of resonance keeps the orbital periods of certain satellites close to a ratio of small whole numbers. For example, Titan and Hyperion orbit Saturn with periods that form a ratio of 3 : 4, and Neptune and Pluto orbit the sun with periods that form a ratio of 2 : 3. The most remarkable instance of such a relation is the resonance that keeps the orbital periods of the three innermost satellites of Jupiter at a ratio of 1 : 2 : 4. It is the only resonance known in which three bodies participate. Moreover, each term in the

The Albedo Effect

The Albedo Effect

Mathematical models of the atmosphere are the chief scientific tools for predicting long-term climate and identifying possible climatic changes that may result from man's activities. Recent advances at the General Motors Research Laboratories have revealed new information about the contribution of airborne particles to the delicate thermal balance of the earth's atmosphere.



Regions of heating and cooling determined by particle characteristics and surface albedo.

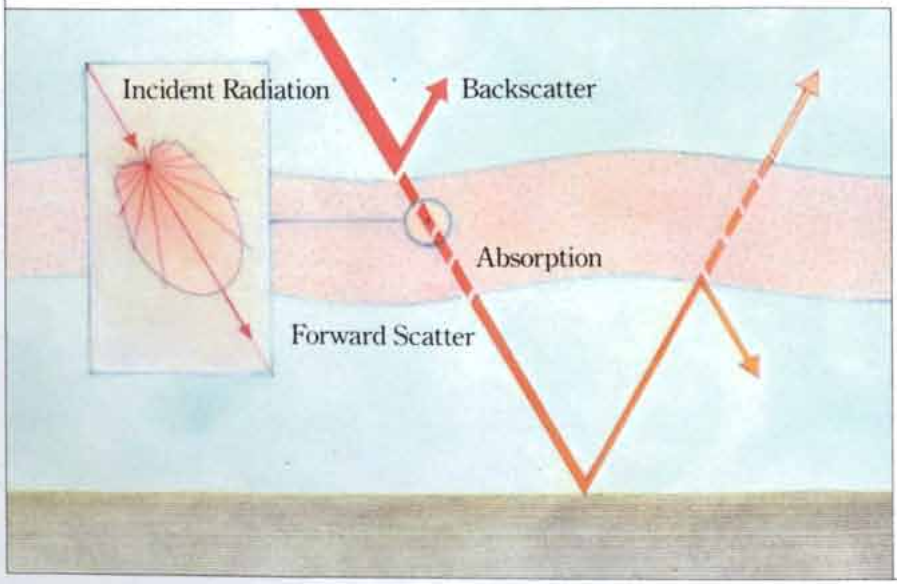
DEVOID of its atmosphere, the bare earth would reach an average temperature of only −1°C. Atmospheric interaction with solar and terrestrial radiation raises the average surface temperature to fifteen degrees Celsius, making life as we know it possible. Small fluctuations in overall temperature can have large-scale effects. It is believed that a drop of a few degrees Celsius lasting for a period as short as four years could trigger an ice age. Fundamental studies conducted at the General Motors Research Laboratories explore the effect of various atmospheric factors, natural and man-made, on the earth's thermal balance.

New knowledge of the influence of airborne particles on the earth's thermal balance has been revealed by investigations carried out by Dr. Ruth Reck. Dr. Reck's work at General Motors integrated for the first time the complex factor of particles into radiative-convective atmospheric models. Her findings help determine under what conditions particles have a cooling influence, and under what conditions they have a heating influence.

Airborne particles have many sources: volcanic issue, wind-raised dust and sea salt, ash, soot, direct and indirect products of combustion and industrial processing, the products of the decay of plant and animal life, the liquid droplets and ice crystals that make up clouds. Particles alter the radiation flow in the atmosphere by the processes of scattering and absorption. Particles differ by size and composition, factors which determine optical properties.

Prior to Dr. Reck's work, models for calculating the vertical temperature profile included layers of clouds and the significant gases—O_2, O_3, H_2O and CO_2—but neglected the particle factor. To establish the thermal effect of particles, later models assumed a uniform vertical temperature change.

Dr. Reck's contribution was to add the particle factor to a one-dimensional model developed at the Geophysical Fluid Dynamics Laboratory at Princeton University. This model divides the atmosphere into nine layers. An initial temperature distribution is assumed, and the model is used to compute the net radiative energy



Radiation scattering exhibited by a layer of particles. The inset shows the distribution of scattering by a single particle of mean size.

flow into or out of each layer. A particle population is input for each layer. Calculated radiation imbalances result in a temperature change for each layer within the model, subject to the condition that change in temperature with altitude not exceed the adiabatic lapse rate. The new temperatures are used to compute a new radiation balance. This process is repeated until there are no further changes in temperature.

The particles of interest, known as Mie-scattering aerosols, are comparable in size to the wavelength of the incident radiation. Dr. Reck models the interaction of these particles with the radiation field in terms of two parameters: the single scattering albedo of the particle, which describes backscatter, and an anisotropic scattering factor, which measures the degree of forward scatter. From these two quantities and the size distribution and abundance of the particles, the transmission, absorption and backscatter of each layer in the model can be calculated.

DR. RECK discovered that whether particles have a heating or cooling influence depends upon the surface albedo, or reflective power, of the earth directly beneath them. Snow (0.6) is more reflective than sand (0.3); water is less reflective than either (0.07). Her results indicate that when surface albedo is small, the net effect of particles is to "shield" the earth from incoming solar radiation, producing a cooling influence. When surface albedo is large, a trapping effect prevails, in which the portion of solar radiation that reaches the earth's surface is "trapped" between the surface and the particles, producing a net heating influence. The competition between these two effects, shielding and trapping, determines the overall thermal influence of particles.

Dr. Reck calculated that for the latitudes between the equator and 35°N, where average surface albedo is low, the current background level of atmospheric particles decreases solar radiation reaching the earth by ~1%, thus producing a net cooling effect. Her findings indicate that heating takes place at latitudes north of 55°N, where average surface albedo is high. Calculations with the model indicate a correlation between the increase in particle abundance due to volcanic activity in 1970 and a subsequent ice build-up in 1971.

"Previous models did not adequately take into account the role played by particles in the earth's thermal balance," says Dr. Reck. "The geosystem is continually changing. It is important for us to understand the elements that affect this evolution, so that we may know how man's activities influence the atmosphere."

THE WOMAN BEHIND THE WORK

Dr. Ruth Reck is a Staff Research Scientist in the Physics Department at the General Motors Research Laboratories.

Dr. Reck received her Ph.D in physical chemistry from the University of Minnesota. Her thesis, on the statistical mechanics of heterogeneous systems, concerned the theory of diffusion-controlled chemical reactions. Prior to joining General Motors in 1965, she was a Research Associate in the Applied Mathematics Department of Brown University.

In addition to global climate studies, Dr. Reck has done research at General Motors in solid state physics and magnetic materials. Over the last seven years, she has participated in several international exchange programs on climate-related subjects.

ratio is precise to nine significant digits. The terms in the other resonances are precise to no more than three digits. The mechanisms that preserve such a resonance are complicated but are well understood. What remains uncertain is how the Jovian resonance arose. Two hypotheses have now emerged.

The resonance of the three innermost satellites of Jupiter has been known since the time of Pierre Simon de Laplace, the 18th-century mathematician whose refinements of Newtonian physics owe much to his interest in celestial mechanics. Laplace's attention was drawn to the three innermost satellites because they periodically lie along a line that passes through the center of Jupiter. At the time of such an alignment two of the satellites are on one side of the planet; the third satellite is on the other side. The arrangement is known as the great conjunction.

By the early 20th century the motions of the satellites could be given with great accuracy. The line along which the satellites periodically assemble is rotating at a rate of .739507 degree per earth-day. In the frame of reference that rotates with the line Ganymede, the outermost of the three satellites, completes an orbit of Jupiter in 7.0509 days. (In a nonrotating frame of reference Ganymede moves 50.317608330 degrees per day, for an orbital period of 7.1546 days.) Europa, the middle satellite, completes a revolution in half the time that Ganymede does, to an accuracy of nine significant digits, and Io, the innermost satellite, completes a revolution in half the time that Europa does, again to nine significant digits. The great conjunctions signify not only that the orbital periods have such a ratio but that the phases of the orbits are likewise locked together.

How did the resonance evolve? One hypothesis is offered by Charles F. Yoder of the Jet Propulsion Laboratory of the California Institute of Technology. Yoder's reconstruction begins with the formation of the three satellites (and the rest of the solar system) some 4.6 billion years ago. The initial orbits of the satellites "are far from either the 2 : 1 commensurabilities or the three-body lock." The only initial constraint is one that requires Io to lie "well inside the orbit of Europa." The interaction of Io and Jupiter is therefore much like the interaction of the moon and the earth. In each case the gravitational attraction of the satellite raises tides on the primary body. On Jupiter the tidal forces raise a gaseous bulge, and the rotation of the planet carries the bulge forward from the point on the planet's surface that lies closest to the satellite. The gravitational attraction between the satellite and the bulge then accelerates the satellite. The acceleration pushes the satellite into an orbit farther from the planet. In response to the same forces the rotation of the planet slows. In effect the tidal interaction

transfers angular momentum from the planet to the satellite.

As the moon slowly spirals away from the earth it approaches no other bodies. Io, however, approaches Europa. The gravitational attraction between the two satellites transfers angular momentum (and thereby rearranges their orbits) until they have entered into a resonance in which the distance between Io and Europa is maximal at conjunction. At conjunction, therefore, Io is at pericenter (its closest approach to Jupiter) and Europa is at apocenter (its farthest excursion from Jupiter). The resonance forces Io to follow an orbit whose eccentricity (a measure of the deviation of the orbit from circularity) is greater than it otherwise would have been.

The tides raised on Io by Jupiter now become important. The tide is in essence a force deforming Io. It arises because the gravitational attraction is strongest on the parts of two bodies that lie closest to each other. The forced eccentricity of Io's orbit makes the deformation greater. The deformation of Io in turn causes an orbital asymmetry in which the conjunction of Io and Europa is attained after Io reaches pericenter. The result is a gravitational torque that Io exerts on Europa. The satellites move outward from Jupiter, still locked in resonance.

Then Europa approaches Ganymede. As the system continues to evolve angular momentum is reshuffled by the gravitational interaction of Ganymede and Europa on the one hand and of Io and Europa on the other. (This interaction was worked out by Laplace himself.) The result is a three-body resonance in which Europa attains its apocenter at the time of its conjunction with Io and its pericenter at the time of its conjunction with Ganymede. The three bodies never line up on one side of Jupiter. Yoder proposes that Io's orbit is driven to a still greater eccentricity as the system coalesces into its three-body cycle.

A different hypothesis is advanced by W. E. Wiesel of the Air Force Institute of Technology. Wiesel's computer simulations of the evolution of the resonance begin with Io, Europa and Ganymede in uncoupled orbits and with Io driven outward by its tidal interaction with Jupiter. The three satellites enter suddenly into a lock step in which the periods oscillate about the strict resonant values and then damp themselves down to those values. An earlier lock step between Io and Europa need not have formed, and the tidal deformation of Io need not be invoked.

However the resonance of the three satellites arose, it turns out to be a crucial influence on the physical state of Io. Specifically, the forced eccentricity of Io's orbit has been calculated by Stanton J. Peale of the University of California at Santa Barbara and Patrick M. Cassen and Ray T. Reynolds of the Ames Research Center of the National Aeronautics and Space Administration; the value they obtain is .0043. In the absence of the resonance the eccentricity would relax to a value of .00001. The value of the forced eccentricity leads Peale and his colleagues to a prediction of the heating of Io by friction as tidal forces deform the satellite. If Io is assumed to be homogeneous, "the current tidal heating rate ... is about three times the current radioactive heating rate estimated for the moon."

From the calculation of tidal heating it seems plausible to assume that Io has a liquid core and a solid mantle. Indeed, runaway melting is conceivable. As described by Peale and his colleagues, "heating in the solid mantle melts material near the liquid core, thereby diminishing the mantle thickness. The thinner elastic shell undergoes greater deformation and hence a greater heating rate." As a result "Io might currently be the most intensely heated terrestrial-type body in the solar system" and "one might speculate that widespread and recurrent surface volcanism would occur." Three days after this prediction was published (in *Science,* on March 2, 1979) the *Voyager 1* spacecraft began transmitting to the earth the images of Io that show eight active volcanoes.

The Concise Mitochondrion

The sequence of nucleotide bases that makes up the genome of an organism constitutes a message in which are encoded all the instructions needed to assemble the organism. In many cases the genome includes much else as well. The genome of a mammalian cell, for example, is extravagantly verbose: functioning genes are interrupted by long stretches of DNA that have no apparent meaning. Now a genome has been found that wastes few words. It is the genome of the human mitochondrion, which turns out to consist of exactly 16,569 nucleotide bases. It has recently been characterized in ultimate detail: every base has been identified.

The mitochondrion is not an independent organism but an organelle of the eukaryotic, or nucleated, cell, responsible for generating the cell's main energy supply. Its origin and evolutionary status have long been subjects of speculation. On the one hand the mitochondrion is closely associated with the eukaryotes, and many proteins essential to its functioning are encoded by genes in the nucleus of the host cell. On the other hand the mitochondrion resembles a prokaryote (a cell without a nucleus, such as a bacterium) in size and in many structural and biochemical details. One hypothesis is that the mitochondrion is descended from an ancient prokaryotic cell that was engulfed by another primitive cell to form the present eukaryotic association. The idea that the mitochondrion was once a free-living organism is consistent with the fact that it has a genome of its own. The genome is a loop of double-strand DNA; if it were fully untangled, it would have a circumference of about a millimeter.

The recent analysis of the human mitochondrial genome has not resolved the mystery of the genesis of the mitochondrion; on the contrary, the new findings may compound it. The sequence of bases has yielded many surprises, not the least of which is that the genetic code of the human mitochondrion is unlike that of any currently recognized class of organisms, either eukaryotic or prokaryotic. Indeed, the genome of the mitochondrion obtained from human cells appears to differ in significant ways from the genome of other mitochondria, specifically those derived from yeasts.

The human mitochondrial sequence was solved by a group of workers at the Medical Research Council (MRC) Laboratory of Molecular Biology in Cambridge, England. The group was led by Frederick Sanger and employed a nucleotide-sequencing method developed by Sanger and his colleagues six years ago. (It was in recognition of Sanger's role in the development of this method that he was awarded a share of last year's Nobel prize in chemistry; it was his second Nobel award.) In the present instance the sequencing method was utilized for the first time to analyze an entire molecule of human DNA.

According to the report of the MRC group's findings, published in *Nature,* the nucleotide sequence of human mitochondrial DNA "shows extreme economy of organization." In the pattern of bases it proved possible to identify genes coding for two molecules of ribosomal RNA, 22 molecules of transfer RNA and 13 assorted proteins (five known and eight unknown). Unlike most other DNA molecules, which have long stretches of noncoding nucleotide bases before and after each gene and in which the gene itself is often split into many pieces, human mitochondrial DNA has genes packed so tightly that there are few noncoding bases between them. Furthermore, the investigators found that the termination codon, or stop signal, normally present at the end of a gene is sometimes not coded in human mitochondrial DNA but rather is created by a special biochemical mechanism that comes into play after the genetic code has been transcribed into messenger RNA and before the code is translated into the amino acid sequence of a protein. The compactness of human mitochondrial DNA is in marked contrast to yeast mitochondrial DNA, which is five times as long but is thought to carry no more genetic information.

Sanger and his colleagues also note that in several instances the genetic code of the human mitochondrion, like that of the previously analyzed bovine mi-

tochondrion, "is different from the so-called universal genetic code and is read in a unique fashion by a minimal set of mitochondrial [transfer] RNA's." On the basis of these findings and others they conclude that "the mammalian mitochondrial genetic system cannot generally be classified as either prokaryote-like or eukaryote-like."

Glueballs

The denial of action at a distance is a tenet of modern physics with considerable intuitive appeal: it holds that physical causality can result only from contact by two bodies or from a disturbance in the space between them. In describing high-energy collisions of elementary particles this principle is embodied in a mechanism. It is assumed that a particle can exert a force on another particle only through the exchange of a third particle. What then are the properties of the third particle? Is it merely a go-between for the force or can it too "feel" forces and combine with other particles into more complex states of matter? In separate submissions to *Physical Review Letters* K. Ishikawa of the Deutsches Elektronen-Synchrotron (DESY) in Hamburg and Michael Chanowitz of the Lawrence Berkeley Laboratory argue that decay products observed in certain high-energy experiments may derive from bound states of the particles that are exchanged in the strong interaction. The exchanged particles are called gluons, after the role they play in binding other particles together. The hypothetical bound states of gluons have been called glueballs.

The existence of glueballs is suggested by the theory of quantum chromodynamics, or QCD, which has been highly successful in explaining the proliferation of particles found in experiments with accelerators. According to QCD, all the particles generically known as hadrons are made up of the more fundamental entities called quarks. The proton and the neutron, for example, are both hadrons; each is made up of three quarks. Each quark bears charges called color charges that combine in such a way that the quark has a net color, say red, blue or green. (Such labels are arbitrary and have nothing to do with ordinary colors.) It seems, however, that bare color is never observed in nature: quarks always bind together in states that have no net color charge. The binding is effected by the strong interaction and so is mediated by gluons.

QCD is analogous in many ways to quantum electrodynamics, or QED, the theory that describes the quantum nature of electromagnetic interactions. The mediating particle in QED is the photon, the quantum of light, which is massless and always moves with the speed of light. Although the photon transmits the familiar repulsion or at-

traction of the electric force, it carries no electric charge itself. Gluons are like photons in that they are massless and move with the speed of light, but some gluons carry color charges. This difference in the nature of the mediating particles has important consequences. Since photons are neutral, there are no direct forces that might bind them together. Since gluons can carry color charges, on the other hand, it should be possible for two gluons to interact with each other by exchanging a third gluon. Under some circumstances the gluons could bind together in such a way that their color charges cancel. If the colorless particle formed in this manner could be observed, it would confirm QCD in a dramatic and unusual way.

Such observations, however, are experimentally taxing to make and theoretically difficult to interpret. The usual experimental procedure calls for injecting a beam of electrons and a beam of positrons into a storage ring, where the two beams circulate in opposite directions and are allowed to collide at carefully controlled velocities. When a positron and an electron collide, they annihilate each other, and the relativistic mass associated with their velocities at the moment of collision becomes available as energy.

The findings that have now been interpreted as evidence for glueballs come from experiments with the storage ring called SPEAR at the Stanford Linear Accelerator Center, with a ring at DESY and with similar devices in other laboratories. In the experiments the energy of each of the two beams is finely tuned to a little more than 1.5 GeV (billion electron volts), so that the total energy is somewhat more than 3 GeV. The energy is exactly enough to materialize a particle called the *J*/psi.

When the *J*/psi decays, several generations of daughter decay particles are formed before the interaction is detected; in one class of experiments the daughter particles that ultimately trigger the detectors are energetic photons. The creation of the *J*/psi is marked by a resonance: a sudden enhancement in the number of photons detected when the colliding beams have an energy exactly equivalent to the mass of the *J*/psi. The possible evidence for glueballs consists of a secondary resonance: a tendency of the photons emitted in a cluster along a certain trajectory to have a fairly well-defined energy level. Such a secondary resonance is evidence for the existence of a particle among the intermediate decay products.

One of the most difficult experimental problems is resolving resonances that lie close to one another in the energy spectrum. Lengthy operation of the storage ring at the energy of each resonance is needed to obtain statistically significant results. The glueball candidate identified by Chanowitz was not noticed earli-

er because a resonance at 1.44 GeV was initially assimilated to a resonance at 1.42 GeV, the mass of another product of J/psi decay called the E meson.

The positive identification of a glueball is further complicated by its resemblance to an ordinary hadron made up of quarks. Indeed, the theoretical description of a glueball shows that it is partially made up of quarks just as a hadron is partially made up of gluons. The main distinction is how energy and momentum are distributed between the gluons and the quarks. A glueball can be thought of as a hadron whose "valence" components are gluons, whereas the valence components of an ordinary hadron are quarks.

The notion of valence components in the structure of a hadron derives from an analogy with chemistry. The valence quarks or gluons are the dominant influence on the properties of a hadron much as the valence electrons determine the chemical properties of an atom. The analogy can be carried further. A periodic table of the hadrons can be filled in by assignment of some of the resonances discovered in accelerator experiments. Any resonances that have no place in the table may well be glueballs.

Addicts' Progress

What is the effect of opiate addiction on longevity? A social experiment of 60 years ago in New Haven, Conn., has provided the basis for a retrospective study that addresses this question. The experiment, begun in 1918, was the establishment of a clinic in City Hall where registered addicts could buy a daily dose of morphine at the nominal cost of four cents per grain. The clinic was set up because of dissatisfaction with the maintenance system that had existed in New Haven under the terms of the Harrison Narcotic Act of 1915. In that system a single private physician had been authorized to write prescriptions for the city's small population of opiate addicts.

David F. Musto and Manuel R. Ramos of Yale University have now compiled an account of the clinic's 25 months of operation. Writing in *The New England Journal of Medicine,* they also give the results of their follow-up study of those clients of the clinic whose history could be traced.

In May, 1920, four months before the clinic was closed, 91 individuals were registered; 47 of them had been among the first to register in 1918. By searching New Haven city directories for the years from 1870 to 1973 Musto and Ramos were able to identify 50 of the 1920 patrons. A review of Connecticut death certificates then enabled them to determine the date and cause of death for 40 of the 50. The sample of clients whose deaths were documented consisted of 35 men and five women, all of them white. For each such client Musto and Ramos compared the life expectancy in 1920 as given by actuarial tables with the actual age at death. All had died earlier than the actuarial tables predicted: the men by 12 years, at a mean age of 56.6, and the women by 18 years, at a mean age of 51.4.

This finding suggests there is a strong connection between addiction and early death, but further analysis of the data indicates that the connection may not be a causal one. Only three of the 40 were described as addicts on their death certificates. Sixteen died of heart disease, five of cancer, four of tuberculosis, two each of pneumonia, venereal disease and alcoholism and one of cirrhosis of the liver. Four died in accidents and one committed suicide. None of the deaths was ascribed to opiate overdose, which, like suicide, is one of the major causes of death among addicts today.

It seems that what can best account for the reduced life expectancy of the addicts is not addiction in itself but the close correlation between addiction and poverty. Musto and Ramos conclude that opiate addiction at the time the clinic operated was only one of many hazards to health associated with low socioeconomic status and that it did not necessarily lead to a quick, catastrophic death or to lifelong addiction.

Oxen of the Sun

Legend has it that in the third century B.C. Archimedes was annoyed with Apollonius of Perga, who had found an approximation of π better than Archimedes' own and who had written a treatise on the multiplication of large numbers that criticized Archimedes' work. For revenge Archimedes devised a computational problem whose solution called for the manipulation of immense numbers, which he sent in a letter to Eratosthenes of Cyrene, the chief librarian at Alexandria. Whether or not Archimedes actually wrote the letter and invented the problem, with which he is linked by Plato and other ancient writers, has been the subject of much debate, but there is general agreement that the problem is at least as old as the beginning of the second century B.C. It was not solved until about 15 years ago. Now it has been solved again, and for the first time the more than 200,000 digits of the smallest solution have been published. Moreover, five new solutions with still larger numbers have been found.

"Compute, O friend," the problem begins, "the host of the oxen of the sun, giving thy mind thereto, if thou hast a share of wisdom. Compute the number that once grazed upon the plains of the Sicilian Isle Trinacria and that were divided according to color into four herds, one milk white, one black, one yellow and one dappled." The problem goes on to state that there are more bulls than cows in each herd. The following constraints must also be satisfied:

1. White bulls = yellow bulls + (1/2 + 1/3) black bulls.
2. Black bulls = yellow bulls + (1/4 + 1/5) dappled bulls.
3. Dappled bulls = yellow bulls + (1/6 + 1/7) white bulls.
4. White cows = (1/3 + 1/4) black herd.
5. Black cows = (1/4 + 1/5) dappled herd.
6. Dappled cows = (1/5 + 1/6) yellow herd.
7. Yellow cows = (1/6 + 1/7) white herd.

"If thou canst give, O friend, the number of bulls and cows in each herd, thou art not unknowing nor unskilled in numbers, but not yet to be counted among the wise. Consider, however, the following additional relations between the bulls of the sun":

8. White bulls + black bulls = a square number (the second power of an integer).
9. Dappled bulls + yellow bulls = a triangular number, a number that has the form $n(n + 1)/2$.

"When thou hast then computed the totals of the herd, O friend," the problem concludes, "go forth as conqueror, and rest assured that thou art proved most skilled in the science of numbers."

It is not difficult to determine that the smallest herd satisfying the first seven conditions consists of 50,389,082 cattle. When the eighth and ninth conditions are taken into account, the problem is much more difficult. It was not until 1880 that significant progress was made in finding the solution. In that year the German investigator A. Amthor concluded that the number of cattle is a 206,545-digit number that begins 7766.

Later in the 19th century A. H. Bell, a civil engineer, and two friends formed a club in Hillsboro, Ill., to take up the computations where Amthor had left off. They quit after four years, having found 28 more of the leftmost digits and 12 of the rightmost ones. (Two digits have since turned out to be wrong.)

The first complete solution was provided with the aid of a computer in 1965 by H. C. Williams, R. A. German and C. R. Zarnke of the University of Waterloo, but it has not been published. The new solution was found by Harry L. Nelson of the Lawrence Livermore National Laboratory, who used the problem to test the performance of the laboratory's new CRAY-1 computer. All 206,545 digits are given by Nelson in *Journal of Recreational Mathematics.*

Nelson comments that the solution of the problem was found with the CRAY-1 in about 10 minutes. This was not long enough to serve as a proper test of the machine, and so the program went on to find five larger numbers that satisfy all the constraints. The largest number has more than a million digits.

"To become more independent from oil is a decisive economic and technical task. For without secure energy supplies, there is no economic and social stability, no growth, no efficient performance at home or abroad."

● Chancellor Helmut Schmidt in his Government Declaration on 24th November 1980. This Special Report by economic and financial writer Andrew Hargrave explores how Europe's strongest economy responds to the challenge of energy.

VEBA

OUR STRENGTH - A BROAD BASE

The VEBA Group is one of Germany's major energy producers. It is at the same time a broadly based industrial and service organization with total sales of more than DM 40 billion and 84,000 employees. The German Federal Republic holds 44% of the equity, the remaining 56% is in the hands of around 800,000 shareholders.

The corporate backbone is electricity, providing power to millions of households and numerous industrial users. VEBA's petroleum sector ranges from exploration on a global scale and refining to a large network of petrol stations. In the chemical field, VEBA offers a widely diversified range of products.

VEBA is also engaged in worldwide distribution and transportation including freight forwarding and handling, ocean and inland shipping.

For further information please get in touch with VEBA AG, P.O. Box 30 03 06, D-4000 Düsseldorf 30.

The economic background

Although the West German economy is showing its customary resilience to the worldwide economic recession, it could not remain immune to it. After a 1.8 percent rise in the Gross Social Product (GSP) last year, a decline of between 1 and 2 percent is predicted for this year coupled with yet another massive balance of payments deficit. (Last year's DM 28* billion compares with only DM 9 billion in 1979).

Nevertheless, the inherent soundness of the economy is reflected in an inflation rate of under 6 percent, among the best in advanced industrial nations, as well as in a continuing high level of investment – though this, too, is expected to fall short of last year's exceptional standards. The major worry is the escalating price of energy which has, for the second time, exposed West Germany's vulnerability.

As Chancellor Helmut Schmidt pointed out in his government declaration last November, the second "oil shock" has left "an as yet unperceivable mark on our economy in terms of rising costs and prices, current account deficits and higher unemployment."

"The increased cost of oil", declared Chancellor Schmidt, "deprives our citizens of a real income of around DM 30 billion a year: such a drop in purchasing power cannot be replaced within the domestic economy."

More recently, Dr. Hans Friderichs, spokesman of Dresdner Bank (West Germany's second biggest commercial bank), estimated that the 170 percent increase in the price of mostly imported oil has transferred about 2.5 percent of the purchasing power from the industrial nations to the oil producers since the beginning of 1979.

As for West Germany, the higher cost of energy – mainly oil – accounted for DM 16 billion of the DM 19 billion rise in the balance of payments deficit last year. Import bills from major suppliers rose steeply: from the OPEC countries by 39 percent, from the United Kingdom (now the second largest supplier) by 33 percent. Increased exports could only partly compensate for these huge rises.

On the other hand, the import bills would have been even higher had it not been for the 11 percent drop in the volume of oil and oil products. Thus, for the first time since 1967, oil's share within the nation's primary energy consumption (down 4 percent) sank below the 50 percent mark last year. (It was in fact 47.8 percent).

With 96 per cent of West Germany's crude oil (mostly from OPEC countries whose volatility is only underlined by the Iranian revolution and Iraq-Iran war) and two-thirds of its gas imported, Chancellor Schmidt needs little proof to ram home the message summed up in the slogan "Away From Oil." The overall aim, he said, was "a more economic and rational use of energy, preference for domestically-mined coal, limited expansion of nuclear energy and the development of renewable energy sources."

Few in West Germany question the objectives: what many people, including those in industry, challenge is what they see as a lack of clear-cut detailed energy strategy. They particularly object to that vague reference to "limited expansion of nuclear energy" in the government declaration in the light of the virtual standstill in the power program for nearly six years.

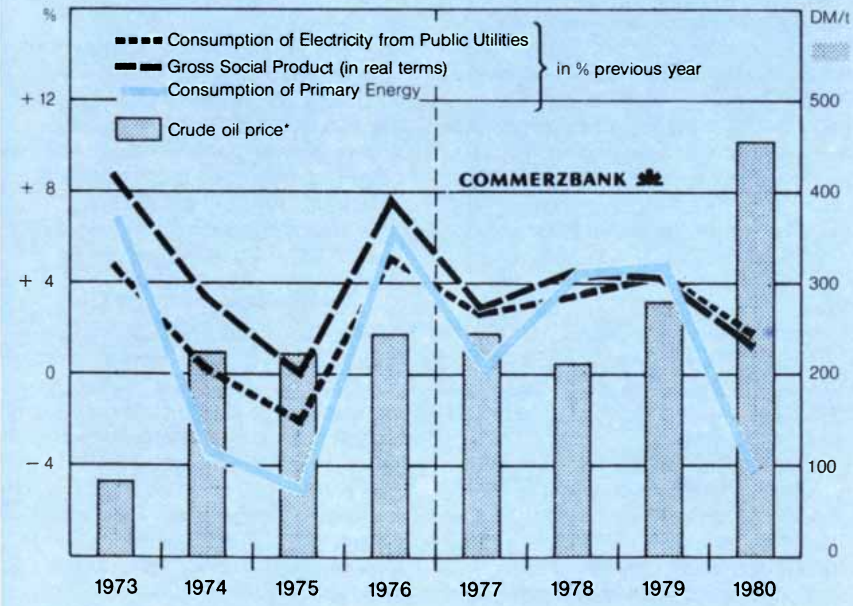
Some foresee a frightening energy gap developing well before the end of the decade unless the present complicated and lengthy approval procedures for power stations, which stretch the period between the first application and final commissioning to up to ten years, are considerably speeded up. They hold up France where the procedure takes up only half the time as an example to follow. Professor Dr. Rolf Sammet, chief executive of Hoechst, one of the world's leading chemical concerns, expresses the feelings of many of his fellow-industrialists when he pinpoints nuclear power as the key issue in any sensible energy policy. Dr. Manfred Lennings, chief executive of the major engineering, energy and transport group GHH, predicts a possible power gap of 25,000 megawatts by 1990 unless the approval procedures are streamlined. "Without nuclear energy we have no future," says Dr. Lennings. His warning is widely echoed by other leaders of industry, commerce and public utilities.

The energy infrastructure encompasses not only safety of supply but the structure of power generation itself. Nuclear energy last year provided less than 3.5 percent of West Germany's primary energy and only 14 percent of the total electricity generated (It was nevertheless still above the European Community average of 12.5 percent).

The time-scale of the threatened energy gap depends on the rate of growth in the economy – though as Chart I shows, less than in the previous years – and on the development of alternative sources of energy and measures for saving and conserving it.

Another incalculable joker in the pack is the presence of increasingly vocal and sometimes violent pressure groups opposed to any extension of nuclear power.

CHART I **Consumption of Energy and Gross Social Product**



Source: Arbeitsgemeinschaft Energiebilanzen, Stat. Bundesamt: courtesy of Commerzbank

* 1 DM represented US $ 0.47 at the time of writing.

They are responsible for much of the delay in the courts and have managed to halt construction on two power stations at Brokdorf in the North and at Wyhl in the South as well as the proposed nuclear waste disposal site at Gorleben, in Lower Saxony.

Yet hopes for renewed economic growth rest on the availability of reasonably-priced and safe energy and a reduction of dependence on imported crude. A medium-range forecast by Westdeutsche Landesbank already speaks of a third "oil shock" in the next few years, with a 50 percent rise in the price of crude which would absorb a further 2–2.5 percent of the GSP of industrialized countries, including West Germany.

Deutsche BP has already predicted a rise of about 15 percent – DM 10 billion – this in spite of a further fall in oil imports.

The silver lining in WestLB's forecast is that such price escalation will give manufacturers of capital equipment for energy saving, conservation and conversion an extra lift in both the home and overseas markets. Hence WestLB's generally shared view that "energy policy will assume a key role in the coming years."

On the other hand, pious statements coupled with economy cuts in the energy research and technology budget have aroused fears that this "key role" is not wholly appreciated by the government. It is believed, for instance, that a projected DM 800 million increase in the Ministry's total allocation of DM 2 billion for energy purposes has been scrapped to keep the government's inflated borrowing requirements within bounds – a false economy in the view of some.

As a result a major program for coal liquification and gasification – cost DM 13 billion at 1979 prices – may be spread over a longer period than anticipated. Some plants may not be built or extended at all. Conflict between future energy requirements and short-term worries over public debt are bound to arise and such conflicts are not confined to the authorities. There seems to be a divergence in views and priorities – indeed between intentions and proposed solutions – in West German industry, finance and the community as a whole.

Such points of conflict include:

- Oil substitution: the major chemical and motor manufacturers in particular are reluctant to contemplate long-term substitution;
- The public's free choice of fuel: district heating schemes using waste heat from power stations, steel mills and factories could certainly cut the use of light fuel oil for space and process heating. On the other hand, this would seem to limit the customer's choice.
- The financing of energy-related projects: should it be left to market forces or treated as a special case? Should undoubtedly long-term and risk-bearing projects be subsidized from the public purse beyond the test and demonstration stage?
- The use of coal for oil substitution: should imports continue to be restricted, even though West German hard coal is among the most expensive in the world?

These and other issues will be explored in some detail in this report.

The worldwide energy imbalance

There is a significant imbalance between the estimated known worldwide stocks of major non-renewable fuels and their uses. Although oil and natural gas comprise only about one-sixth of the world's reserves of non-renewable fuels, they make up two-thirds of total consumption.

The main culprits responsible for this imbalance of use are the industrial nations: the U.S., Japan and Western Europe alone consume almost two-thirds of all oil, gas and coal available on the market.

Although Uranium, the basic material for nuclear energy, is also a non-renewable source, its life-span may be extended by about 50 percent by high-temperature reactors (which can also contribute to better fuel utilization through combined-cycle operations) and no less than sixty-fold by being burnt in fast breeder reactors. The fast reactor is still at an experimental stage (although several industrial nations, including West Germany, have pilot plants in operation): and there is also an acceptance barrier to overcome. Commercial fast reactors are not expected to operate in West Germany before the mid-1990s at the earliest.

Nor are renewable energy sources, such as hydro-electricity (with some exceptions such as Norway, Austria, Canada, etc.), the sun, wind, tides, biomasses (organic and vegetable matter) likely to make significant contributions to the world's energy needs for the rest of this century.

Thus the industrialized world, including West Germany – the main consumers of energy – will have to grapple with the problem of better utilization of conventional resources, including oil, gas, coal and uranium. The worldwide distribution of these resources has, of course, important political implications for the West. For a large proportion of the deposits lie in politically sensitive areas. Around one-third of all coal and natural gas and one-eighth of all oil is located in the Eastern Bloc, mainly the Soviet Union, while two-thirds of all oil deposits are in the volatile Middle East and North Africa.

Distance, too, plays a role, particulary in the case of coal, a bulky commodity: a large part of coal resources lie in Australia, China and the western parts of the U. S.

For West Germany the lesson is clear: an intensification of the effort to upgrade and utilize the country's domestic resources of hard and soft coal, of natural gas, including the modest deposits of crude oil; the re-organization and re-adjustment of the domestic energy market, particularly the heating sector to reduce dependence on imported fuels.

Well over half of West Germany's oil imports originate in the Middle East and, to a much lesser degree, in the Soviet Union. However, supplies from these areas – particularly from crisis-hit Iran – are on the decline while those from the North Sea are rising: the United Kingdom is now the country's second largest supplier of crude oil after Saudi Arabia, followed by Libya and Nigeria.

Since 1973, West German imports from OPEC countries have declined from 95 percent of the total to barely over 75 percent. Trends point to a further reduction in OPEC's share, with the U. K. and Norway filling the gap.

Natural gas, the third largest supplier of primary energy (at 16.5 percent of the total last year) after oil (47.8 percent), hard coal (19.7 percent) and soft coal (9.9 percent) is also becoming a "politically sensitive" issue because of the proposed deal which would almost double the share of Soviet gas by the mid-1980s within the overall gas consumption.

At present, domestic deliveries meet about one-third ot the country's needs; the Netherlands rather more than that; Norway 12 percent and the Soviet Union 17 percent.

A DM 10 billion contract to be financed by a consortium of European countries led by Deutsche Bank would increase the Soviet share to 30 percent, thus replacing to some extent declining supplies from domestic sources and from the Netherlands in the 1990s.

It should be noted that even with expanded gas deliveries and some crude, the Soviet Union would only contribute about 6 percent to West Germany's total primary energy needs – by no means a "blackmail" situation as some opponents of the deal fear.

The home energy market – a forward glance

If a reduction in the share of oil within the overall energy supply is a major priority, the greatest opportunity of achieving it lies in the heating market. For over 70 percent of all energy in West Germany is used for heating – of homes, offices, schools, factories and other premises as well as for process heating. Under one-third of the mostly imported oil – only one-

Each gesture of his is an act of technical perfection

He is a master of technique. The pantomime portrays his ideas, conceptions and observations in an absolutely convincing manner, without big words.

Great technical ability is also necessary when new ideas and discoveries are to be realized for higher performance, greater safety, more independence or for new methods of workmanship.

The engineers and designers at Mannesmann can call upon the knowledge and the experience of all divisions of the company.

New safe high-capacity units were produced for drive and control techniques, module-controlled programming systems for distribution and warehousing, energy saving and environmentally sound processes for metal making and metal shaping.

This is the way the people of Mannesmann carry out their ideas, conceptions and observations to technical perfection.

And, of course, new and improved starting materials for pipe and tubing that have to hold their own under extremely adverse conditions. Fuel element cladding tubes in nuclear power stations, tubes with a high creep rupture strength for the gasification of coal, and large-diameter pipes that bring gas and oil from remote sources to the centres of consumption.

Technology that is convincing, without big words.

Mannesmann AG
P.O. Box 5501
D-4000 Dusseldorf 1
Fed. Rep. of Germany

Ask the man from Mannesmann

5625/8 e MWG

TABLE I **Projected changes in heating market 1977 – 2000, in %**

Source of heat	All heating 1977	All heating 2000	Space heating 1977	Space heating 2000	Process heating 1977	Process heating 2000
Oil	51	19 (+1)*	67	26 (+2)*	35	12
Natural gas	19	34 (+3)*	15	37 (+6)*	24	32 (+1)*
Coal (direct)	18	10	11	1	24	19
Coal (refined)	–	4	–	3	–	5
District heating	3	7	4	11	2	3
Electricity	9	25 (+7)*	3	21 (+11)*	15	28 (+3)*
Others	–	1	–	1	–	1

*In parentheses: additional heat gained through the combined cycle process, recycling and other means of raising the rate of energy utilization.

sixth of all primary energy – goes into motor fuel or chemical feedstock: consequently the major thrust of any oil-saving campaign must aim at the heating market.

A study by the West German electricity supply industry "Wärme 2000" (Heat 2000), published last October, couples heat demand at the end of this century with the industry's own fuel-requirements and tasks ahead. Although the study is based on 1977, the year of the Federal Government's last strategic review of energy needs, the findings are still broadly valid.

The Government's overall objective, restated in Chancellor Schmidt's latest declaration; is to reduce the rate of growth in energy consumption through the economical and rational use of energy and to widen the range of energy sources to cut the share of oil.

The electricity study assumes that within the total volume of primary energy available in Year 2000 – estimated to be 46 percent above the 1977 level – oil could be cut by about a quarter and thus its share reduced from 52 percent to only 27 percent. The share of natural gas, though up in volume, would stay at around 17 percent; hard coal would increase its share from 18 percent in 1977 to 22 percent. Soft coal's share would decline slightly in spite of a volume increase.

But the real burden of filling the energy gap should fall – as other commentators suggest – on nuclear energy whose share within the total primary energy demand should rise from 3 percent in 1977 to 22 percent in Year 2000. This amounts to 120 million tons of Coal Equivalent (C.E.) or equal to 30 percent of last year's total primary energy consumption in West Germany.

The study foresees no major changes in the *structure* of primary energy consumption during the lifetime of this century. Power and light should still use under 30 percent of primary energy, the rest going into space and process heat. As almost two-thirds of the heat market is still dominated by oil, the scope for substitution by alternative sources – nuclear electricity, coal, gas, etc. – shown in Table I – is considerable.

(The study estimates that 90 billion kilowatts of additional electricity could replace nearly 50 million tons of crude in the heating market or half of last year's oil imports. This would, by itself have wiped out the entire 1980 current account deficit).

Electricity, through greatly expanded use of nuclear energy, should therefore meet a quarter of the needs of the heating market by Year 2000 compared with less than one-tenth in 1977: or almost one-third, if the various forms of utilizing waste heat (including district heating) are effectively put into operation.

The study examines in detail the proposed changes in the structure of the heating market. The method of financing is considered in another study (by Dresdner Bank) discussed later in this report.

In industry, which used up 43 percent of all heat in 1977, the share of oil could be cut from two-thirds to one-third in space heating by substituting natural gas and district heating. These, as well as waste heat and coal, could make significant contributions to reducing the share of oil in process heating from one-third to 15 percent.

In the home, 36 percent of the heating sector, storage heating and heat pumps could replace oil-fired central heating in about half of nine million West German homes as well as in the three million homes with individual oil fires. Electricity, heat pumps and sun roofs could provide a quarter of all heat required in Year 2000 compared with only 5 percent in 1977; district heating could treble its share, and natural gas more than double its share. By these means, the share of oil in space heating could be reduced from two-thirds of the total to a quarter, and in process heating from 38 percent to 5 percent.

The same applies to small users (hospitals, churches, schools, offices) which in 1977 took 20 percent of the heating market.

During last year, no doubt helped by the mild winter, considerable strides were made towards these goals. Consumption of light fuel oil dropped by 16.8 percent; of heavy fuel oil by 12.6 percent, of other oil products by 12.3 percent and of diesel fuel by 3 percent. Only gasoline consumption showed a rise, a slight 2.1 percent and even that rise was below last year's increase in the motor vehicle population (3.2 percent).

The power problem

The main reason why industry (as exemplified by Dr. Lennings and Professor Sammet), public utilities, and people at large are becoming concerned about adequate long-term supplies of power NOW is the average ten-year gap between the planning and commissioning of power stations.

Indeed Prof. Dr. Max Ludewig, board member of Kraftwerk Union (KWU) which is a subsidiary of the powerful Siemens group, foresees a possible bottleneck as early as 1986–88, depending on future rates of increases in electricity consumption.

Should electricity consumption rise at the expected average annual rate of 4 percent over the next ten years, major industrial users such as chemicals, steel and aluminium could be the first to feel the pinch.

The public utilities' association, VDEW, has prepared a ten-year plan charting the country's power needs based on slightly higher average rates of increases – 4.6 percent until 1985 and 3.6 percent thereafter until 1991. (This broadly corresponds to the 1973–79 annual average increase of 4.2 percent).

On this basis, the present total generating capacity of 74,000 megawatts (mW) would have to be increased by an average of 3,000 mW per year to reach just over 106,000 mW by 1991.

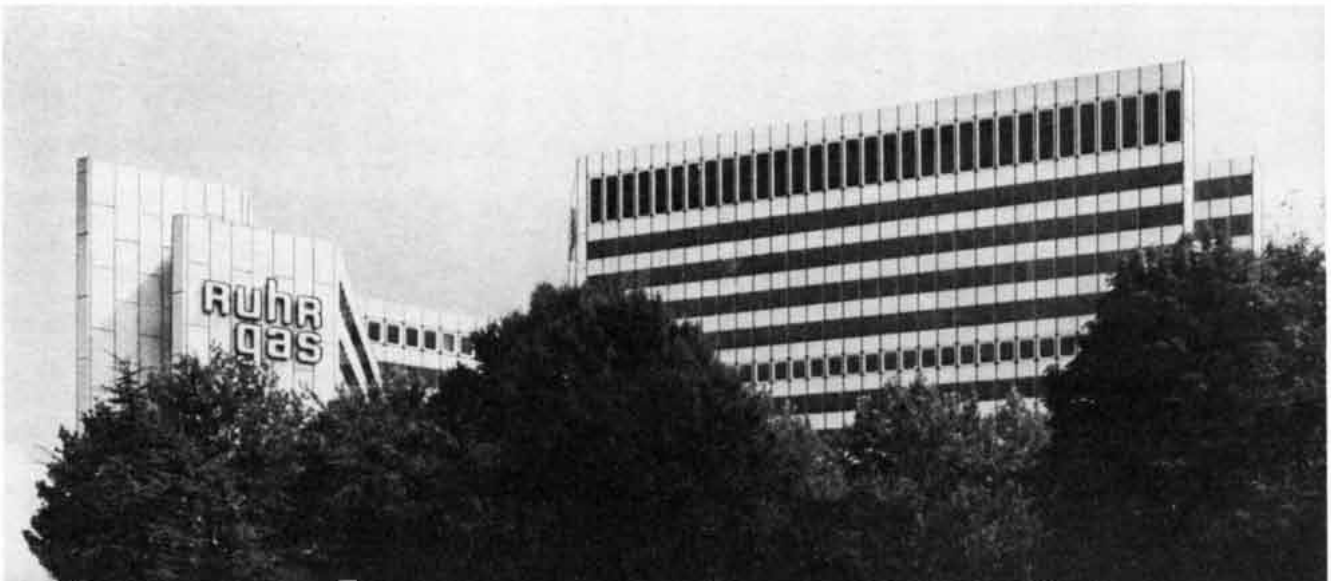
Of this additional capacity of 32,000 mW, about 22,000 mW (20 new power stations) would be based on nuclear power and over 8,000 mW on hard coal. (The balance may be met by other sources, including natural gas).

A further 5,000 mW of hard coal-fuelled capacity would be required to replace out-of-date coal-fired stations.

By that time, too, increased contribution from nuclear power could go some way towards achieving an economic balance in the structure of power stations; ideally 50 percent of capacity should provide base load output, the rest middle and peak load.

In 1979, however, the structure of power station capacity was far from ideal. Table II indicates VDEW's assessment of how it might develop over the next decade.

An important feature of the plan is to rectify the balance between base load output based on "cheap" fuels such as nuclear electricity and soft coal (only 40 percent of capacity in 1979) and the middle and peak load based on the other fuels, mainly hard coal.

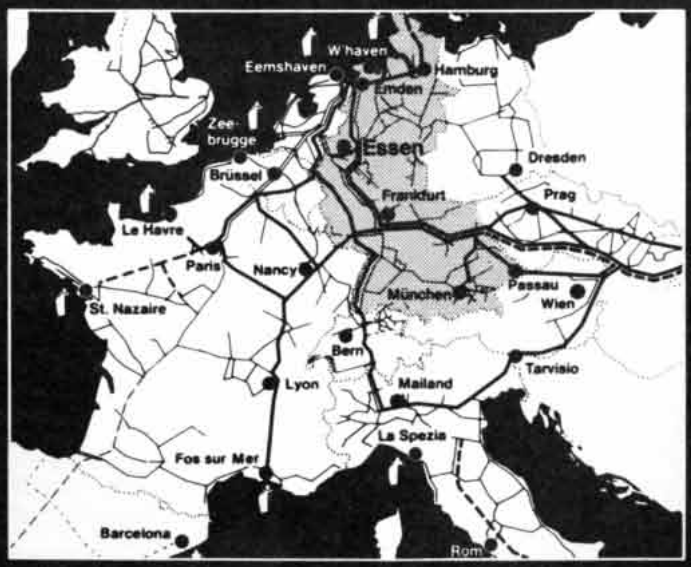


Ruhrgas – Success in the International Gas Trade

Ruhrgas AG is one of the largest gas companies in Europe, and the largest organization of its kind in the Federal Republic of Germany, purchasing gas on the international energy market for supply to utilities as well as industrial and commercial users within West Germany.

Cooperation with its European partners characterizes Ruhrgas operations at international level. Ruhrgas and its partners conclude long-term contracts with exporters both inside and outside Europe, construct and operate international transmission systems linking almost all countries on the continent and supply natural gas to a huge market which represents approx. 75% of total energy and natural gas consumption and some 90% of G.N.P. in Western Europe.

Ruhrgas also devotes considerable time and energy to the advancement of natural gas technology and is particularly involved in energy conservation projects such as the development of a gas-fuelled heat pump.



Backed by its sound financial position, Ruhrgas dedicates its commercial and technical resources and skills to the security of energy supplies in the Federal Republic of Germany.

RUHRGAS AG
Postfach 28 · D-4300 Essen 1 · Federal Republic of Germany

Ruhr gas

Natural Gas is Our Business

TABLE II **Structure of power station capacity 1979 – 1991, in mW**

	1979	1985	1991
Soft coal and hydro-power	14,400	14,800	14,600
Nuclear energy	8,700	17,500	30,800
Hard coal	20,300	26,600	28,700
Natural gas and oil	22,500	23,900	23,800
Others	7,100	7,700	7,900
TOTAL	73,900	91,300*	106,600*

*Planned capacity

How realistic is this plan, especially in view of the 10-year time scale in the approval procedure?

At present, the utilities have, including two experimental reactors (a fast reactor and a hightemperature one) almost 9,000 mW of nuclear capacity at their disposal. Nine more stations of a combined capacity of 9,900 mW are in the course of construction, with one 1,300 mW nuclear station due to be commissioned this year and 3,900 mW, representing three further nuclear stations, in 1982–83.

One or two new coal-fired stations of smaller (700–800 mW) capacity should also be operating by the end of 1983.

To these may be added nine stations (nuclear as well as coal, combined capacity 7,000 mW) in advanced stage of planning with contracts already awarded, plus further stations with a combined capacity also of 7,000 mW awaiting consideration by local courts. Two nuclear stations – Brokdorf and Wyhl – with a combined capacity of 2,700 mW, are still entangled in West Germany's complicated and lengthy legal processes.

Total capacity of all the stations in various stages of construction and planning is 26,600 mW. They would, if allowed to proceed according to plan, go a long way towards averting the threatened bottleneck in power supplies towards the end of the decade.

"Possible solutions to the dilemma would be to standardize and harmonize approval procedures as well as reactor construction throughout the Federal Republic," says Professor Ludewig.

At present, the procedure varies from state to state and is often influenced by local political considerations. For instance, a nuclear power station in Bavaria, Isar, was approved last year without much delay while Brokdorf, in Schleswig-Holstein but influenced by the Hamburg Senate, and Wyhl in Rhineland-Palatinate have been hanging fire for years.

Approval for power stations is also being held up because of procrastination over a proposed nuclear waste disposal site at Gorleben, in Lower Saxony. The state's Minister President, Herr Karl Albrecht, belongs to the opposition Christian Democratic party which demands expansion in nuclear power capacity, but his freedom to act is limited by his party's wafer-thin majority in the state parliament.

The standardization of reactors, KWU's so-called "Konvoi" system, is based on the 1,300 mW pressurized water reactor already approved by West Germany's Reactor Safety Commission.

Uniform approval procedures and standardized reactors would not only speed up the nuclear power program but also, through series production, reduce engineering costs. The Konvoi system, KWU believes, would produce savings of 20-25 percent in both production and time-scale.

Nuclear power stations are admittedly more expensive to build but become more economical with each rise in fossil fuel prices. VDEW argues that even on present form the fuel costs for nuclear power stations amount to only 15-20 percent of the total compared with 50-60 percent for stations run on fossil fuel.

The expansion of nuclear power, of electricity in general, new technologies such as the heat pump, solar power and power-heat combined cycle – all of them replacements for oil and, at the same time, energy savers are basically "substituting capital for energy," says Herr Hartmut Boeck, president of VDEW.

Increased electricity use in space and process heating, he says, would lead to an overall reduction in energy consumption – and cost – which in turn would benefit economic growth: "Although these new technologies require enormous investments, it seems preferable to invest in long-term, job-creating, power-producing plant to investing in non-renewable energy." This year the electricity undertakings intend to invest DM 11 billion in transmission, distribution, new capacity and equipment, by far the largest single private investment in West Germany.

Just how large the required "enormous investments" in future might be has been the subject of a comprehensive study by one of West Germany's leading banks.

A DM 760 billion investment

Dresdner Bank has a traditional interest in energy prognostication. A previous attempt in 1974 was thrown out of gear by the energy price explosion in the wake of the Middle East war, accelerated worldwide inflation and the slowdown in West Germany's own nuclear power program. Undaunted, the bank produced another detailed forecast about the energy investment requirements of the western world, the European Community, and, of course West Germany itself, between now and Year 2000 for the World Energy Conference in Munich last year.

For West Germany the total figure is put at DM 760 billion which at the rate of exchange prevailing at the time (it was DM 1.70 to the U.S. $ compared with DM 2.10 at the time of writing this report – which just illustrates one of the major difficulties in forecasting) or 4.6 percent of the global requirement of DM 16,400 billion. Half of the sum is earmarked for electricity generation and transmission, fuel costs and safety, of which one-half again is the cost of the power stations themselves.

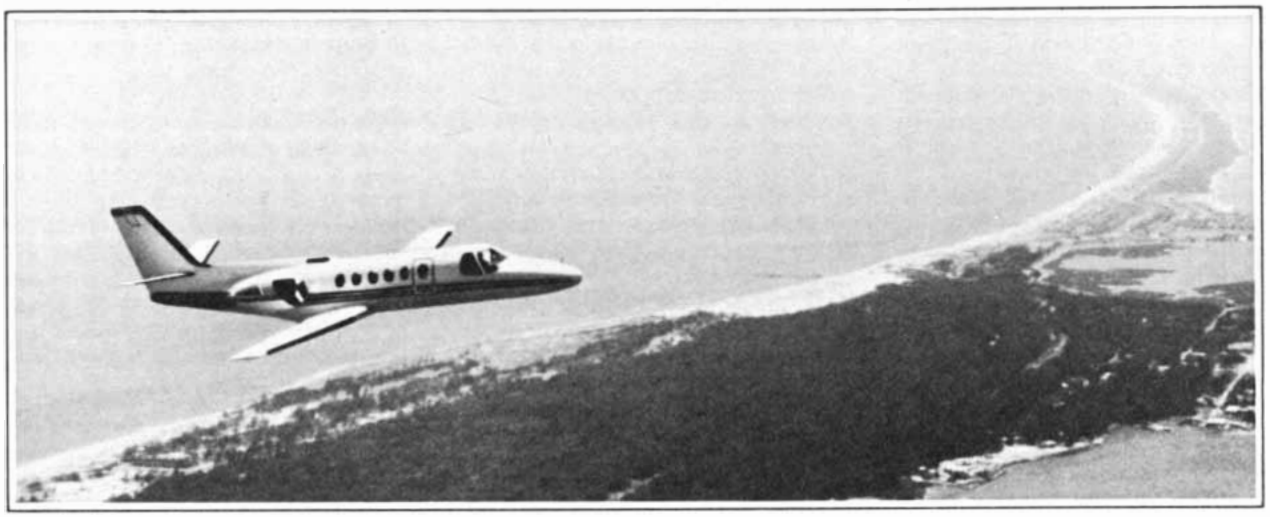
Like other financial institutions, Dresdner Bank believes that the capital market, under free market conditions, would be able to cope with this huge call on funds, provided there is an increased degree of self-financing and "appropriate profitability."

TABLE III **Total investment in energy required 1980 – 2000, in DM billion**

Sector	1980 – 85	1980 – 90	1980 – 2000
Hard coal	22	52	150
Soft coal	6	11	28
Crude oil	15 – 20	30 – 40	75 – 90
Natural gas	17	40	70
Electricity generation/transmission	55	147	300
Nuclear/cycle*	12	28	70 – 90
Renewable sources**	–	7	50
TOTAL, approx	130	320	760

*Fuel, enrichment, recycling, nuclear waste disposal
**Solar, wind, heat pumps, etc.

Hypo-Bank royal client service means speed and mobility in international banking.



Success in international banking is often a matter of speed and mobility. Hypo-Bank, Munich, with consolidated assets of more than DM 83.8 billion, goes to great lengths to provide the quality banking services wherever you need them.

For example, our New York branch is flexible enough to respond quickly to almost any requirement: Corporate finance, foreign exchange and money market transactions, import-export and third country finance through letters of credit, collections, loans and advances, and many other services.

Speed and mobility in international banking are just two aspects of Hypo-Bank royal client service, a tradition since 1835 when we were established in Munich by King Ludwig I of Bavaria.

Through our network of subsidiaries, branches in London and New York, affiliates, representative offices, partnership in ABECOR, and a mobile team of banking professionals, we offer services worldwide.

To learn more about Hypo-Bank's international banking capabilities, get in touch with E.W. Bollinger or W.F.C. Drielsma at Wall Street Plaza in New York.
Tel.: (212) 248 0650,
Telex: 233 604.
Head Office:
Theatinerstrasse 11
D-8000 Munich 2
Tel.: (0 89) 23 66-1
Tx: 05 286 525-27

HYPOBANK
BAYERISCHE HYPOTHEKEN- UND WECHSEL-BANK AKTIENGESELLSCHAFT

Modern Banking in the finest Royal Tradition

The bank's forecast for the global sums required in West Germany is shown in Table III.

The bank has also calculated the proportion of investment to be financed from industry's own resources and estimates a credit requirement of DM 114 billion – or 35 percent of the total investment required – by 1990 and one of DM 350 billion, or 46 percent of the total, required by Year 2000. Assuming that industry achieves further 10 percent internal funding through increased equity and other means, credit requirements are reduced to DM 82 billion (25.6 percent of the total needed by 1990) and to DM 274 billion (36 percent of the total required) by Year 2000.

These are very large sums: but the bank believes that a relatively modest increase in the credit provided by the banks and insurance companies would be sufficient to bridge this gap. In 1978 bank loans to the energy sector represented 3.6 percent of their total lendings and those of the insurance business 4.9 percent of the total. Raising the proportion to 4.5 percent and 6 percent respectively would consequently meet the need for funds, or nearly so.

The study emphasizes that its cautious optimism about finding the means to finance such an ambitious energy program rests on the assumption that industry will, through its own resources and new equity, continue to raise the larger proportion of the funds needed. At the same time, it warns that failure to raise such funds might lead to the energy sector claiming an even higher proportion of the total credit volume which would, in turn, create frictions and distortions in the capital market.

At the same time, the study acknowledges that "considerable technical and financial efforts are necessary for the development of existing and new energy resources." This theme was further developed at the Munich conference by Dr. F.W. Christians, joint spokesman for Deutsche Bank, West Germany's biggest commercial bank. He warned against reliance on "substitutes for self-financing," such as leasing or project financing, which had created "growing scepticism towards these tailor-made forms of finance."

Dr. Christians shares the general view of industry and banking that Government participation in energy financing should be limited to complementing private industry in the search for sensible uses of energy, in the development of existing and new energy resources and in as yet uneconomic projects.

Dr. Günther Radtke, a director of Dresdner Bank and co-author of the bank's study, echoes the opinion that project financing is not primarily a task for banks.

He believes, for instance, that the cost of such experiments as power and heat generation by fluidized bed combustion should be borne at least in part by the public utilities and by the coal companies.

"The banks are not industrial enterprises," says Dr. Radtke. Nevertheless he admits that the time may come when West German banks would have to follow the example of some US and British banks (the Bank of Scotland, for instance) which have set up their own energy departments "to help in the decision-making process."

Herr Guenther Schmidt-Weyland, a member of the DG BANK management board, believes State aid is necessary where the risks in the area of energy are too high for private enterprise: and he does not rule out solutions along the lines of export financing where the Government and the banks (through Hermes) insure exporters against a proportion of the loss they might incur.

Dr. Johannes Voelling, chief executive of Westdeutsche Landesbank, estimates that the proportion of funds companies involved in energy would have to raise in the capital market could be as high as 50 percent of total investments. This would of necessity widen the companies' gearing ratio, lengthen the period of credit and increase risks both for the companies and the lending institutions.

Consequently the requirements of energy financing, Dr. Voelling suggests, might result in significant changes in the structure of the financial market, with the Euromarket playing an important credit-providing role between short-term (up to five years) fixed-interest bank credit and long-term credit raised in capital markets.

Project financing abroad, Dr. Voelling believes, should be, whenever possible, tied to repayment in kind, i.e. deliveries of oil, coal, gas. (The proposed natural gas deal with the Soviet Union, mentioned earlier, is a case in point). The "political risks" of such deals, as for instance the billions of Deutschemarks lost in Iran, might be guaranteed by the State (as in straight export business) through the Hermes insurance scheme.

Technology aids exploration

Not all agree – not even all in West Germany agree – that oil demand can be significantly reduced in the foreseeable future. (A reduction of oil consumption by Year 2000 is forecast in the electricity industry's study cited earlier, but even such an optimistic view envisages a cut of less than 25 percent in oil consumption over the next 20 years).

Dr. Hellmuth Buddenberg, chief executive of Deutsche BP, goes even further: the slogan "Away From Oil!" may be "popular and gripping but is also false and illusory."

Dr. Buddenberg predicts that oil demand in West Germany will be almost as high at the turn of the century as it is at present. He accepts that much of the oil can be substituted or eliminated in the heating and power sectors – for instance last year under 5 percent of the fuel requirements of power stations was supplied by oil – but not in the petrochemical or transport sectors. The slogan should therefore be varied to: "away from fuel oil – up the upgraded oil!"

This, acknowledges Dr. Buddenberg, demands "extraordinarily capital intensive conversion plant." Already an investment of about DM 3 billion has been earmarked by the oil companies in such plant alone (A single hydrocracker may cost anything up to DM 500 million).

BP, with the great US groups and Anglo-Dutch Shell, all multinationals, control 75 percent of West Germany's refinery capacity of around 150 million tons. Of the remaining 25 percent capacity, VEBA, the leading West German energy group, controls just over half.

The oil industry as a whole envisages a slight reduction in overall refinery capacity over the next 10 years, though the proportion of hydrocracker capacity is expected to rise from 21 percent to 27 percent in West Germany, from 30 to 37 percent in the case of VEBA.

(Conversion plant enables the refiner to switch in a more flexible way from heavier oil products such as fuel oil to lighter ones, such as motor fuel and petrochemical feedstock).

The West German oil exploration company Deminex (in which VEBA has a 54 percent holding) has many concessions the world over, including shares in some already oil-producing North Sea fields. And while West Germany's own oil and gas reserves are small by world standards, the companies drilling for oil and gas are investing large sums in upgrading as well as raising the exploitation rate of domestic reserves.

The 1973 oil crisis heightened interest in home-based reserves: BEB-Gewerkschaft Brigitta-Elwerath, the exploration company jointly owned by Esso and Shell which is responsible for around two-thirds of domestic gas and one-third of domestic oil supplies, has raised its annual investment from DM 100 million in 1970 to DM 300 million in 1978 and to DM 500 million last year. This year it plans to invest DM 600 million and according to Dr. Heino Luebben, chief executive of BEB, investment by the company to explore and drill for oil and gas is expected to be maintained at a level of between DM 500 million and DM 1,000 million a year between now and 1985.

Total investment by West German exploration and development enterprises is likely to reach DM 6.5 billion between 1979-83, says the chairman of the industry's association WEG, Dr. Guenther Sassmannshausen, who is also chief executive of the major energy and minerals group, Preussag. By Year 2000, total investment by the industry may reach DM 30 billion.

The ultimate objective is to at least double supplies of home-produced oil and gas, partly through fresh technical advances and partly as a result of steep rises in fuel prices which have made hitherto uneconomic fields more attractive to the operators.

Domestic crude oil contributes just over 4 percent to the country's total requirements. New technology in the form of steam injection and chemical additives could add 70 million to 120 million tons to the 64 million tons of known reserves: new, hitherto uneconomic fields a further 5–10 million tons (To compare: West Germany imported 98 million tons of crude last year).

Most of West Germany's natural gas reserves lie adjacent to the big Dutch land-based gas fields. Through the Frac process widely used in the US but only recently introduced in West Germany (it is based on the injection of liquids into deposits) as well as new deposits, the life of the known reserves of 280 billion cu. m. could be more than doubled. At present the West German resources provide around one-third of the country's total gas needs which have risen steeply in the past decade.

Oil and gas have sulphur as an important by-product: and as Dr. Luebben points out, the company, which is also West Germany's largest sulphur supplier, has to cope with the problem of balancing its sulphur activities with those of oil and gas. For while oil and gas sales are booming, those of sulphur – with recession-hit chemical manufacturers as the main customers – are declining, which certainly hits profitability.

West Germany also has modest oil shale resources, and several major companies, including VEBA, BEB and Preussag, are involved in exploring this high-viscosity and still uneconomical resource.

Experimental drilling for geothermal energy in the Stuttgart area is being sponsored by the Federal Government within an overall European Community research program.

"Urach 3", at a depth of 3,300 meters, is the deepest and biggest hole drilled within the new European project; the heat that it is expected to generate will eventually be used to heat homes as well as water.

Natural gas has the advantage over oil (but not over coal) that around one-third of total consumption is met from home sources. Dr. Sassmannshausen of Preussag is confident of home-produced gas being maintained at the current level of around 20 billion cu. ft. well into the 1990s, though its share of overall gas consumption is to decline as the volume of consumption rises.

All but 17 percent or so of West German natural gas imports originate in Western Europe, primarily in the Netherlands and Norway, the rest in the Soviet Union. The elaborate gas pipeline criss-crossing Western Europe and converging on West Germany will, once natural gas supplies begin to dry up, be capable of being switched to synthetic natural gas and coal gas (both produced from coal) as well as hydrogen.

As it is, Western Europe's present high degree of self-sufficiency (87 percent of the total gas consumed) is expected to decline to around 60 percent by the end of the century.

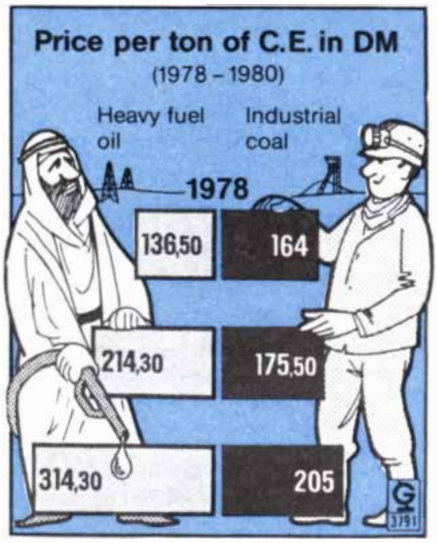
Towards a revival for coal

"We are in a temporary phase between a crisis and revival," says Dr. Karlheinz Bund, president of the association of the West German coal mining industry. He rejects the notion that the revival of coal is already a reality and that no further economic measures are necessary. "This is a false impression racing far ahead of the facts," he maintains.

Dr. Bund is also chief executive of the biggest West German hard coal group Ruhrkohle which accounts for over two-thirds of the country's deep-mined production. Earlier this year it announced an annual investment program of over DM 2 billion for the next few years, more than half of West Germany's total projected investment of DM 3.5 billion in hard coal mining. Ruhrkohle's 1981 investment target of DM 2.3 billion goes mainly into coal preparation and coking plants and associated works: in the coming years at least one-third of the total will be invested in new sinkings.

For all his caution, Dr. Bund is optimistic about the future of coal in the changed circumstances of world energy prices and supplies, and the general economic situation (As Chart II shows, price relationships between heavy fuel oil and coal, interchangeable fuels, has changed dramatically in recent years).

CHART II



"The oil crisis," says Dr. Bund, "has been an inspiration to technological progress. We must get to grips with the problem of alternative energy sources in the industrialized countries over the next ten years. The problem is even more pressing in the Third World. There is now a big future for such technologies as coal conversion, fluidized bed combustion, heat pumps, district heating – and above all, there is now positive thinking."

The hard coal industry plans to increase its output (88 million tons last year) to over 90 million tons by 1985, to over 95 million tons by 1990 and to 100 million tons by Year 2000 at the latest – preferably by 1995 – and allowing for exhausted workings.

Moreover, by Year 2000 West Germany should be importing 40 to 50 million tons of coal per year, instead of the present 7 million tons, with Ruhrkohle alone taking 10–15 million tons. The rate of increase in imports would, of course, depend to some extent on the requirements of the steel industry which at present absorbs around 40 million tons of West Germany's output, including 15 million tons for export; and also on the state of the international coal trade, at present badly affected by the Polish crisis.

Up to an estimated 10 million tons of coal may be required for coal conversion – and these estimates do not include the unspecified amounts of so far unusable coal, dross or coal with very high ash and sulphur content – which may make further supplies available for specific applications.

Dr. Bund himself is keen on adopting the concept, established by West German sponsors such as Deminex, for the exploration and development of oil and gas resources abroad. Dr. Bund's brainchild Dekolex, a parallel organization for coal, may not get off the ground for some time yet owing to the Government's economy cuts.

Meanwhile, Ruhrkohle itself has already taken its first modest steps overseas: following the purchase and subsequent sale of a small mining complex in Virginia, it has now a share in projects in Australia and Canada.

West Germany has the largest resources of soft coal (lignite) in Europe which already makes a substantial 25 percent contribution to power generation. The bulk of domestic resources are owned by Rheinbraun Rheinische-Braunkohlenwerke, a subsidiary of RWE, West Germany's largest public utility: soft coal provides 60 percent of the fuel required in power stations under its own ownership. Further large resources in Halmbach Woods, on the western edge of the country, should ensure that supplies are maintained up to the end of this century, with some to spare for new uses, including synthetic gas and motor fuel.

Old ideas for new technologies

The sheer necessity of offsetting the effect of Allied blockades in two World Wars have forced the Germans to make the most of their existing energy resources, of which coal, both hard and soft, is by far the biggest and longest lasting. So when the first oil crisis burst onto the world in 1973, the Germans had ready-made ideas for a future energy concept aimed at reducing the share of oil within the total primary energy consumption, but they still hard to face the problem of devising and improving technologies to make the converted coal products – oil, gas, gasoline, chemical feedstock – economical in constantly changing circumstances.

This is one reason, perhaps the main one, why despite the anguished cries for greater self-sufficiency in energy, West Germany still has no *commercial* coal-conversion plant, no *commercial* networks of gasoline stations selling alcohol mixtures, no *commercial* power station utilizing very low-grade coal or dross and relatively few heat pumps. Recent Government economy cuts have further slowed down, if not altogether curtailed the program for coal-conversion demonstration plant, the forerunners to commercial production.

An executive of Saarbergwerke (a prominent, technologically advanced coal company) has put it in a nutshell: "West Germany has very little time left for further experimentation with coal conversion. The United States and Britain are already ahead and others are fast catching up."

There is clearly a contradiction somewhere: acute awareness of the problem on the one hand, reluctance to meet it head on on the other. One explanation is that West German hard coal is expensive to mine and is getting more expensive with ever-deepening shafts, narrowing seams and rising investment, yet, as power stations utilize at best 38 percent of the heat produced, some of that huge investment is obviously wasted.

Those engaged in making most of the country's own energy resources – Government scientists, coal technologists, public utilities, coal companies and others (who include the major oil multinationals now showing growing interest in alternative fuels) – are striving to maximize the output of power or heat (singly or combined) while reducing the input of fuel, materials and man-power as well as the damage to or pollution of the environment.

Fluidized bed combustion converts coal directly into useful energy: coal is being pulverized at high temperature under atmospheric or compressed pressure and used to drive a generator. Carried a stage further, the process can be refined to use exhaust gases from the steam generator as well to drive a gas turbine which in turn drives not only the compressor but also generates power. This is called a combined cycle. Finally the need for disposing of a highly toxic by-product – sulphur – is eliminated by the process itself.

A further advantage is the use of the lowest grade coal (with ash content of 40 percent or over) as well as dross, good enough up to now for slagheaps only – eyesores in many coal-mining districts.

There are half a dozen pilot or demonstration plants based on fluidized bed combustion which operate or are in the course of construction in West Germany's coal-mining centers. The first such combined cycle power station to come on stream is at Voelklingen in Saarland, which, according to Herr Rudolf Lenhartz, chief executive of Saarbergwerke, the operating company, may well become "one of the power stations of the future." It incorporates "an entirely new concept," although a power station based on fluidized bed combustion and designed by German engineers has been operating in Spain for over ten years; there, admittedly, environmental regulations are not as stiff as in West Germany.

The 200 mW Voelklingen station, fired by low-grade coal and dross, produces both electricity and heat, some of the latter to be used for district heating. (Plans are already afoot to supply space and water heating to several thousand homes in the nearby city of Saarbruecken).

The station's main purpose is to both prove the concept and to eliminate the remaining technical snags. Apart from firing otherwise unusable coal and being virtually pollution free, the process enables a 40 percent heat utilization, the best of any existing power station so far.

"We are ready to design and build commercial stations of up to 700 mW," says Dr. Werner Brocke, a member of the Saarbergwerke management board.

He reckons the Voelklingen plant will require 3–4 years to prove the technology; thereafter he foresees "quite a boom" in such power units probably in ten years' time. "There are going to be big changes in the generation and use of energy before the century is out, but the course will have to be set in the 1980s," he says.

Another interesting application of fluidized bed combustion is also about to be tested. A pilot plant next to Korf Stahl's Kiel steelworks is to use the process for gasifying low-grade coal (instead of natural gas) for the direct reduction of iron ore pellets.

Coal: a substitute for oil and gas?

Germans have been developing technologies for the conversion of coal, both hard and soft, into liquid fuels and gas for 60 years or more: but it is only under the pressure of circumstances such as a world war or a world oil crisis that these technologies were given the necessary impetus to overcome technical and economic obstacles to their widespread use.

As already noted, the implementation of an ambitious DM 13 billion program announced in 1979 for 14 demonstration plants using up 10-15 million tons of coal will apparently be spread over a longer period than originally intended and may not even be carried out in its entirety. Part of the program involves the extension of pilot plants already operating; for instance, the trebling in size of Ruhr 100 at Dorsten, a synthetic natural gas (SNG) plant based on a high-pressure gasifier by Lurgi (a subsidiary of Metallgesellschaft) which is being operated by Ruhrkohle, its subsidiary STEAG, and Ruhrgas. At the time of the announcement, SNG from coal cost almost three times as much as natural gas, though the gap has probably narrowed since then. The cost of this single project was put at the time at DM 1.8 billion.

Another SNG demonstration plant would be based on the high-temperature Winkler process for soft coal. (A small pilot plant already exists at Frechen near Cologne, owned by the RWE subsidiary Rheinbraun). Soft coal costs only a fraction of hard coal, even considering its lower calorific value and the plant itself would cost only one-third of a hard-coal based SNG plant: it may therefore prove economical for the production of methanol, an important chemical feedstock, in the relatively near future.

The Federal Government would prefer the State authorities, public utilities and private companies to assume greater responsibility for financing these projects. It would also, as already noted, like to see more international projects along the lines of the joint coal-liquification project in Virginia as well as partnerships between multinationals and West German specialists or utilities.

US knowhow is being put to good use among others at the Shell-Koppers coal gasifying plant at Harburg, near Hamburg, built by Krupp-Koppers (a Krupp subsidiary); and at the Oberhausen plant operated by Ruhrkohle and Ruhrchemie (a Ruhrgas subsidiary) which uses a Texaco gasifier.

Selling technology abroad

West German specialists in coal conversion such as Lurgi and Krupp-Koppers whose pedigree as innovators, directly or indirectly, goes back many decades, are naturally loathe to yield the market to foreign competitors whether from the US or Britain.

The largest experimental coal conversion plant in West Germany, a DM 400 million 200 tons-a-day liquification plant at Bottrop, has been funded by the North-Rhine-Westphalia state Government, the operating companies as well as by the Federal Government. The operating companies are VEBA OEL and Ruhrkohle; this, and a smaller plant at Voelklingen based on a similar principle, are using a pre-war German process updated by modern US technology.

There are several other pilot coal conversion plants in operation sponsored by the Ministry, state or municipal authorities, public utilities and private companies, often in combination. That the timing of the next stage – large demonstration plants leading to commercial production – is in doubt, is due partly to the Government's economy cuts but also to reluctance by private operators to commit large funds to long-term, capital-intensive projects. So most of the coal conversion contracts for Lurgi or Krupp-Koppers are for export. Lurgi's biggest contract to date is for Sasol, the South African energy corporation which, in line with South Africa's drive for self-sufficiency, intends to meet around half its total internal motor fuel requirements from domestic coal before the end of the decade as well as produce coal-based chemical feedstock, iron ore reduction gas, etc. from a variety of plants based on the Lurgi technology.

Lurgi is also active in the US, Indonesia, Brazil and, among others, more recently in China and aims to be ready – as chief executive Dr. Dietrich Natus puts it – for "the day after tomorrow" when the escalating prices and scarcity of oil and gas make coal even more competitive, especially in parts of the world where it is in abundant, easily accessible supply.

Lurgi and Krupp-Koppers are among the few companies in the world which can design, build as well as deliver (as engineering contractors) coal gasification and liquification plant.

The Krupp-Koppers gasification plant is based on the Koppers-Totzek principle, another pre-war concept. The company has contracts for building major plants in Poland and Brazil for the production of SNG or methanol. The two plants have a combined value of DM 450 million – DM 550 million. Negotiations to build a similar one in the US recently led to the company supplying a $ 15 million unit for a large demonstration plant in Pennsylvania. The advantage, as Dr. Karl Schmid, a spokesman of the company points out, is that such plants can convert any type of coal, including low-grade and soft coal, into gas or liquid products.

Dr. Schmid calculates that, given a Government investment subsidy of 50 percent, a DM 500 million - DM 600 million coal gasification plant based on imported coal could already produce 500,000 tons

of methanol a year as a motor fuel economically – provided it was taxed on its calorific value which is half of that of gasoline.

Soft coal, of which West Germany has very large reserves but which is mainly used as cheap power station fuel, could also make a bigger contribution to oil substitution. At present, it plays an as yet modest though increasing role in replacing fuel oil: the production of pulverized soft coal, for instance, has increased from 500,000 tons a year in 1977 to 1.8 million tons last year, thus replacing 500,000 tons of fuel oil, mainly in the cement industry.

An expansion of the pilot soft coal gasification plant at Frechen is, as noted above, already being contemplated. A second pilot plant at Wesseling, also in the Rhine area (where Rheinbraun's parent, the Rheinisch-Westfälische Elektrizitätswerke – RWE – has an oil refining subsidiary) uses nuclear process heat for gasification.

According to Dr. Guenther Klaette, a member of the RWE management board, prices of gas from soft coal for chemical feedstock and possibly other purposes may already be competitive with oil-based feedstock prices. Moreover, the Wesseling refinery itself could be adapted to use the gas as a raw material for chemical feedstock and pipe it to chemical works of which there are several in the area.

Indeed the Dresdner Bank study envisages the use of soft coal, in combination with nuclear energy (which would provide process heat) as a base material for gas and gas-based liquids as an alternative to the continued use of soft coal purely for power stations. Such an alternative depends largely on whether RWE is allowed to expand its nuclear capacity and so release soft coal (which at present provides more than 60 percent of its power fuel) for other purposes, and also on general price movements in the energy sector.

Heat pump as oil saver

As individual installations for space and water heating in homes, schools, hospitals, offices, factories, etc., heat pumps are highly capital intensive. They could, however, lead to the biggest individual savings in fuel oil, as space heating absorbs about one-third of all primary energy in West Germany. Scope for heat pumps is considerable, complementing other methods of space heating, such as district heating via waste heat from power stations and from industry.

The Institute of the German Economy (IDW) has estimated that around half the 45 million tons of fuel oil consumed by households and smaller users a year (which is over 70 percent of the total space and water heating market) could be eliminated by the use of heat pumps, equivalent to between a quarter and a fifth of all present crude oil imports.

It would save DM 7 billion in foreign exchange or nearly a quarter of last year's current account deficit.

To achieve this aim, about 3 million of these appliances would have to be installed before the end of this century (at present only 25,000 are in use in the whole of West Germany), at a total cost of DM 35 billion at current prices.

Dr. Kaette, of RWE, believes that with escalating fuel oil prices combined with cheaper heat pumps (as a result of volume production), the breakthrough may come sooner than generally anticipated: certainly within a five-year period during which a 10 to 12-fold increase in the use of the appliances has been predicted.

Dr. Karlheinz Kaske, new chief executive of Siemens, West Germany's top electronic-and-electrical group (a major supplier of heat pumps), believes that the break-even point for consumers as well as for suppliers could be even nearer. Dr. Kaske stresses the mass production effect on pricing and recalls as a parallel the dramatic fall in the prices of electronic appliances and equipment in the past 10–15 years.

District heating cuts heat waste

Like heat pumps, district heating is no new phenomenon in West Germany although so far it has had only limited application and then only in major population centers.

However, the country's largest provider of coal-fuelled power STEAG already supplies, through the public utility RWE, 135,00 homes with space and water heating. There are 14 combined heat-power units in 10 centers of the Ruhr, thus reducing the distances heat has to travel and so minimize heat losses.

Ninety percent of district heating runs on home-produced coal; the system, at present 264 km in length, is being steadily extended. Dr. Hans Kraemer, chief executive of STEAG, emphasizes the heat-saving aspect of district heating. Burnt in a power station, 1 kg coal produces 3 kW of electricity; combined with heat production, it produces 2½ kW of electricity as well as 4 kW of useable heat. "District heating provides the biggest opportunity for oil substitution," claims Dr. Kraemer.

District heating is operated by public utilities as well as by large companies with waste heat to sell, such as Thyssen and Krupp. Thyssen provides heat for thousands of homes around its Duisburg steel works while Krupp has recently announced a similar scheme for the Essen area. Saarbergwerke's plans for the Saarbruecken area have already been mentioned. Other population centers equipped with district heating include the cities of Berlin and Hamburg. There are also talks at Government level to extend the district heating network to countries adjoining West Germany.

Making most of the sun, wind...

The sun is not only a renewable but inexhaustible source of energy. Unfortunately, sun is not easily tapped, least of all in Northern Europe where neither its intensity nor its reliability is sufficient to make it a significant contributor to overall energy requirements.

West German engineers and technologists have been grappling, with some support from the Ministry of Research and Technology and European institutions, to convert solar energy into the best and most economical sources of heat and, where appropriate, power. There are estimated to be around 300 West German companies offering appliances, mostly in the heating market.

In homes, swimming pools, offices, factories, solar collectors normally situated on the roof can be a useful supplement to conventional heat providers for both space and water heating. Some estimates put last year's sales for water heating alone at over 16,000 units. Many tens of thousands of square meters of roof space were added to surfaces already covered.

Many of the leading West German engineering and electrical companies as well as air and spacecraft companies (MBB, Dornier) and public utilities (RWE) pay close attention to this promising segment of the market. Brown Boveri Cie (BBC), a leading supplier of solar collectors and water heaters, has also produced energy storing equipment.

Wider use of solar energy, says Dr. Herbert Gassert, chief executive of BBC, is handicapped by the structural changes required to install the plant. However, the expected rises of fuel oil prices are likely to give solar energy a further boost.

M.A.N., the energy-and-transport member of the GHH group, caters, apart from solar hot-water systems for buildings, for other forms of solar energy. As well, M.A.N. has joined other international companies to install a 500 kW solar farm in Almeido, Southern Spain, as a testbed for process heating, air-conditioning, refrigeration and desalination as well as for power generation.

The real potential of such installations lies, of course, in sun-rich developing countries, as Dr. Friedrich Laussermair, member of the M.A.N. management board responsible for research, points out. It is in the sun belt north and south of the Equator for sparsely populated settlements

that decentralized solar power units of up to 1,000 kW output could make a difference between life and death.

The solar tower, a cooperative experiment by several West German companies, including KWU, M.A.N. and the spacecraft manufacturers MBB and Dornier, is a more long-term project: the eventual target is a power plant with up to 100 mW capacity.

The windmill is one of the oldest forms of energy generation, revived by the energy crisis. There are now around 40 experimental projects in West Germany: the biggest windmill, with a capacity of 3 mW, to be commissioned in Brunsbuettel, Schleswig-Holstein, sometime next year, will probably be one of the largest in the world. It was designed and engineered by M.A.N., with components supplied by Siemens, M.A.N. and several other major companies: like solar devices, windmills have their main use in sparsely populated areas.

A further typical example of advanced technology for decentralized power supply is the solar cell. AEG-Telefunken, the West German electrical group, is a leading supplier of solar cells providing solar generators for most telecommunication and other space satellites in Europe – and for some in the U.S. Now the group is experimenting with "terrestrial" solar generators for a variety of uses.

Dr. Horst Nasko, former management board member and now research consultant to AEG-Telefunken, sees considerable opportunities for the land-based solar generator. A number of prototypes have been installed: a 5 kW unit supplying a radio relay station on the roof of AEG's Berlin headquarters; a seawater desalination plant in Mexico, an irrigation plant in Indonesia; a small (10 kW) thermal power plant in Greece. A larger, 300 kW plant, is planned for an island in the North Sea, at present without any electricity supply.

Solar cells cost DM 25 to DM 35 per watt; their widespread use could bring the price down to as little as DM 2 to DM 5 per watt, according to Dr. Nasko. Land-based solar generators may thus develop into significant business in parts of the Third World in general, at isolated spots not supplied from public power networks.

Total energy concepts by single companies

Since the first oil crisis in 1973, West German companies in all sectors of industry, public utilities and even financial institutions (the Dresdner Bank study is just one example) have developed a new consciousness of energy matters.

Some companies are in business for selling integrated energy concepts to others, to organizations and even to entire countries.

The Siemens group – which includes KWU – for instance emphasizes its close involvement in energy concepts in a recent booklet: and it is of some significance that the new chairman to the management board to succeed Dr. Bernhard Plettner, Dr. Karlheinz Kaske, has spent his whole working life – over 30 years – in the electrical energy industry. The goal, as defined by himself, is "the optimum use of energy," in every area of power generation, distribution and consumption, now reinforced by the new microprocessor technology which enables more accurate controls, regulators and tests.

The Siemens concept, according to the booklet, aims at raising further the energy efficiency of products; improving energy utilization by increased input of electricity; producing energy from renewable sources, especially for the Third World; developing processes for the use and recovery of waste heat; and applying the group's knowhow in steering, measuring and automation technology to rational uses of energy.

The structurally somewhat different M.A.N. group has developed its own concept in four broad fields: extraction and processing (from mine conveyor systems to solar farms and gas centrifuges for uranium enrichment); transport, handling and storage (from diesel engines for ships, trucks and heat pumps to conveyor systems); power and recovery plant; and energy utilization, including buses, trucks, engines and heat pumps running on alternative energy sources.

The aircraft company Dornier (of World War II fame) has become involved in energy projects through lessons learned during aircraft construction and space technology; in effect, an internal technology transfer. Knowhow gained in the development of rapidly rotating turbine parts has led to gas centrifuges for uranium enrichment; satellite systems to solar energy; aircraft wing design to wind-power; the search for high-strength, temperature-resistant materials for aircraft to high-temperature heat exchangers.

The diversification in the steel industry, hit by a painful recession in demand, has contributed to the creation of energy concepts. Companies which are large users of energy put the expertise gained in internal energy economy to manufacturing plant, machinery, conveyor systems and components for a variety of applications in energy saving and conservation.

Industry saves energy

According to the Federal Statistical Bureau's estimates, West German industry's consumption of energy relative to the value of its output has declined by more than 100 percent in the past 30 years. The biggest energy savings have been achieved by the chemical industry which at the end of the 1970s was using only one-third of the energy to achieve the same output as at the beginning of the 1950s. This compares favorably with the iron and steel industry, with savings of about 25 percent over the same period.

Though the industry's own association claims, on a somewhat different basis, better improvements, energy consumption per ton of rolled steel had declined by almost a third while the amount of coke needed to produce a ton of iron dropped by as much as 38 percent.

In terms of oil substitution, the West German steel industry has managed to reduce the share of oil in its total energy consumption from almost 17 percent in 1972 to a mere 5 percent: from 4.4 million tons of fuel oil (largely for blast furnaces) in 1973 to less than 1.5 million tons last year. The leading steel concern, Thyssen has virtually given up oil as a fuel for its blast furnaces.

Another aspect of energy conservation – which is not related to output and is therefore missing from energy saving statistics – is the use of waste heat produced in the iron and steelmaking process and sold to heat the homes of adjoining townships.

The iron and steel industry is, of course, no longer self-contained. Most major companies are diversified groups involved in engineering, plant and machinery manufacture (including turnkey contracts), fuel conversion, transport and trading. The process of diversification has gone farthest at the Mannesmann group, only one-third of whose total sales relate to steel and steel products.

The group is involved in the worldwide energy market: from supplying piping for oil and gas from fields in the great depths of the North Sea to land pipelines carrying gas over thousands of kilometers from Siberia and oil from Iraq's Kirkuk field to a Gulf terminal, to the extraction and transport of fuels, conversion, process heating and steelmaking plant.

West Germany's largest steel producer, Thyssen, too, is diversifying: the last accounts show that only about 40 percent of the group's total turnover originates in steel (including special steels), with capital equipment and manufactures approaching the level of steel sales. Combined trading and services exceed steel sales by a substantial margin.

Thyssen is virtually self-sufficient in power, with sufficient waste heat left to supply space and water heating for 5,000 homes around the Duisburg steelworks.

The group plans to save about one-quarter of the natural gas it uses at present and also reduce the coal and coke input by introducing larger oxygen converters. Energy cost savings from these projects should reach 30 percent of the total over the next few years: 10 percent from conversion from natural to recovery gas and 20 percent from reducing the input of coke and coal per ton of iron and steel produced.

Thyssen's energy concept, says Dr. Karl-August Zimmermann, board member responsible for research and development, is a flexible one which would enable the group to switch back to gas – or even oil – if the price of coal once again exceeds that of gas or oil. (Thyssen already meets 30 to 50 percent of coke requirements from its own coke ovens).

Krupp's contribution to coal gasification through its Krupps-Koppers subsidiary has already been noted: the group's activities, however, extend over virtually the whole energy field. Its overall energy concept serves a multiplicity of aims: giant bucket-wheel excavators employed in opencast soft coal mining are also used for tar sand excavation and are being adapted to hard coal and oil shale stripping.

The tar sand excavator works in extreme temperatures: but so does the ice-breaking ore/oil bulk carrier recently developed to transport fuels and raw materials, and another vessel to carry liquified natural gas (LNG) from the Arctic. (The Bremen shipyard, AG Weser, which builds these ships, is a member of the Krupp group.)

Like Mannesmann, Krupp applies its steel technology to supply parts to nuclear power stations. Like Thyssen, Krupp's recovery gas from a new oxygen furnace will also provide space and water heating for 5,000 homes in the Rheinhausen/Duisburg area.

Chemical waste as power source

Professor Dr. Matthias Seefelder, chief executive of BASF – one of West Germany's Big Three chemical groups – once described the 1980s as "the decade of raw materials." The first year of the decade had confirmed that view: "We believe we have done something towards preparing for it in terms of exploration, promotion, technology and investment," he added recently.

BASF certainly has. Efforts to integrate chemical waste disposal with the provision of process heat and electricity in a closed cycle process have resulted not only in energy economy but have also reduced environmental damage and, at the same time, the cost of toxic waste disposal.

BASF needs power and steam capacity of 750 mW to satisfy its own requirements at the huge Ludwigshafen complex: it provides 600 mW "in house."

Heat left after chemical processing, says board member Dr. Ronald Schmitz, is being used in the factories; and, as a new steam cracker produces 10 percent surplus heat as a by-product, the company

Ich bin ein Berliner.

If you are in a business that is looking to expand in the Common Market, take a lead from Henry Ford.

Become a business Berliner.

Because the same incentives and advantages that inspired Ford Motor Company to build a 76 million dollar factory in Berlin apply to you. No matter what the size of your business.

Unique among these advantages is Berlin's turnover tax rebate which effectively adds up to 10.2% of sales to profit. Every year.

In addition, substantial investment grants, low interest loans, income tax concessions and a highly favorable infrastructure combine to make Berlin an ideal Common Market location.

To show you exactly how these benefits can affect your bottom line, we offer the serious investor a ten year comparative profitability study—using your own data—prepared by an independent Big Eight accounting firm.

Here you will find an economically and politically stable city that hasn't had a major industrial strike in 35 years. A dependable, highly skilled and productive labor force. Free access. No supply problems. A high-technology environment. Plus, the culturally dynamic Berlin.

To Ford Motor Company, Berlin is a major manufacturing location for plastic components needed by their other European production plants. To Mercedes Benz, Berlin is the producer of every Mercedes crankshaft in the world.

Join Ford, IBM, Gillette, General Motors, Philip Morris, Warner Lambert and other business Berliners who enjoy our outstanding financial, economic and political climate.

For more information contact: Mr. Wolfe J. Frankl, the North American Representative, (212) 432-8070. Or fill in the coupon below.

Economic Development

BERLIN

Name _____

Company _____ Tel. _____

Address _____

City _____

State _____ Zip _____

Mail to: W.J. Frankl,
Berlin Economic Development Corporation,
Suite 7800, One World Trade Center, New York, N.Y. 10048

is investigating the possibility of selling it to the town of Ludwigshafen for district heating.

The West German chemical industry has, as already noted, led the way in energy saving, with an energy input per turnover only one-third of its 1950 level.

Although the big chemical groups still regard oil as their principal raw material for the remainder of this century, BASF's steam cracker can use liquid petroleum gas (LPG) for up to 15 percent of its feedstock, thus replacing oil-based naphtha. Both BASF and Hoechst are experimenting with coal-based methanol to produce a wide variety of petrochemicals, plastics and synthetic fibers.

At the same time, the big chemical groups are doubtful about coal altogether displacing oil as a significant source of feedstock, in the foreseeable future at any rate. Professor Dr. Rolf Sammet, chief executive of Hoechst, maintains that the use of coal as a chemicals raw material will remain limited, not only because of the much higher cost of conversion into feedstock but because increase in domestic output (a no more than 15 percent or so growth in coal output is envisaged in the next 15-20 years) itself is limited.

Professor Dr. Herbert Gruenewald, chief executive of Bayer, agrees, and at the same time makes a special plea for the industry: "Crude oil and natural gas resources could be stretched and also reserved for the chemical and transport sectors; it is demand for other forms of energy that should be met from other sources."

As long as the distribution of resources depends on price, Professor Gruenewald is confident that the chemical industry would be able to safeguard its share.

Auto makers face tough test

A recent paper by two research executives of Volkswagenwerke (VW), one of the world's top automobile companies, pointed to the multiplicity of goals which would have to be achieved by the car of the future. It would have to meet demands for less energy consumption, lower exhaust gas emissions and noise levels, more safety and other adjustments to traffic-mix conditions while at the same time maintain at least present performance levels. An altogether tall, and sometimes conflicting, order one may say.

The present climate for such innovative development is not, however, particularly auspicious. Last year, most West German companies suffered sales losses and profit setbacks, reflected among others, in VW's cut of dividends to shareholders. (The notable exceptions were Daimler-Benz which not only increased its share of the domestic market but also its volume sales. And BMW, too, has managed to raise its output, albeit slightly). Export markets were hit: the US suffered the biggest drop in car sales in this century.

All the same, bad times tend to sharpen the survival instinct, and the VW researchers suggest that a 10 percent reduction in the aerodynamic drag alone could reduce fuel consumption by 3-4 percent. They consider an ultimate 30 percent drag reduction attainable.

Weight reduction through steel substitution, narrower tires with greater air pressure, lean-mixture engines coupled with electronic controls, high-compression engines, a modification of gear ratios, "uncoupling" of engines when the driver takes his foot off the accelerator and cutting out groups of cylinders when idling are further devices to aid fuel economy. So are alternative engines such as the gas turbines and the so-called Stirling engine and the battery-driven car.

The time-scale of all these improvements depends heavily on the willingness of the customer to pay as well as on the rise in gasoline prices, says Professor Dr. Ernst Fiala, board member of VW.

Daimler-Benz which, with its larger and more luxurious cars caters for a different clientele, already incorporates many of these features. (Nearly half its cars and virtually all its commercial vehicles run on diesel fuel). Board member Professor Dr. Werner Breitschwerdt predicts, however, that future savings will be on a diminishing scale. Reason: automobile technology is already far advanced.

"The recession demands more intelligent solutions," says Professor Breitschwerdt. "We'll have to accept a change in the worldwide production pattern: 'intelligent' products should be manufactured in the West, 'crude' ones in the less developed countries. The same applies to components."

Herr Eberhard von Kuenheim, chief executive BMW, however sees new emission legislation, particularly in the US, as being in direct conflict with energy saving and tending to aggravate already difficult market conditions. Nevertheless, he was able to point out with some pride earlier this year that BMW was one of only three European automobile manufacturers (the others were Daimler-Benz and Renault) who were able to increase their output in 1980.

All the West German motor vehicle makers participate in a long-range country-wide program for alternative fuels, including coal-based methanol and liquid hydrogen (about one-third of VW's Brazilian-made cars have engines adjusted to alcohol-mixture). In addition various experiments are running which involve among others the "duo bus" (it may switch from gasoline or diesel power to electric and vice versa), the "noiseless" bus as well as vehicles of various sizes running on elevated electric guideways. Companies involved include the automobile, truck and bus manufacturers – VW, Daimler-Benz, Ford, Opel, M.A.N., Kloeckner-Humboldt-Deutz (KHD) as well as MBB, Dornier, Siemens, Mannesmann-Demag, Krauss-Maffei and others.

Brown Boveri has high hopes of their new high-energy car electric battery which, chief executive Dr. Gassert believes, could eventually save millions of tons of motor fuel.

The major motor vehicle component manufacturer Bosch is, of course, in the center of all energy-saving activities. For example, up to 20 percent gasoline savings are claimed from the Motronic – an electronic device controlling fuel injection and ignition; more savings by using the company's engine testing system, or the ALI traffic control and information system which is now on trial in various traffic experiments throughout the country.

Dr. Hans Bacher, Bosch board member, admits that integrated electronic control systems are still relatively expensive: but the Bosch research team is working hard at developing new, cheaper and yet effective systems.

KHD (which has, among others, produced the engine for the "noiseless" truck) has developed a whole range of energy saving engines, turbines and pumps for commercial vehicles, ships and a variety of engineering plant; also devices for the utilisation of waste heat and gases for additional energy input. The company claims up to 30 percent fuel savings in its air-cooled truck engines, most of which are sold in the US.

KHD's comprehensive energy concept as well as strategies in other areas are the result of work by a team of strategic planners engaged in long-range planning as far ahead as Year 2000.

A long term solution

The technological solution of West Germany's long-term energy problem is in the hands of a few thousand – possibly only a few hundred – people working in a variety of Government departments, research institutes, organizations and private companies. Their task is undoubtedly made more difficult by the recession and the financial stringencies associated with it. However, as this report shows, it is being recognized that in precisely such times of stringency must attention be focused on the wise use of scarce resources, if for no other reason than to safeguard the nation's future.

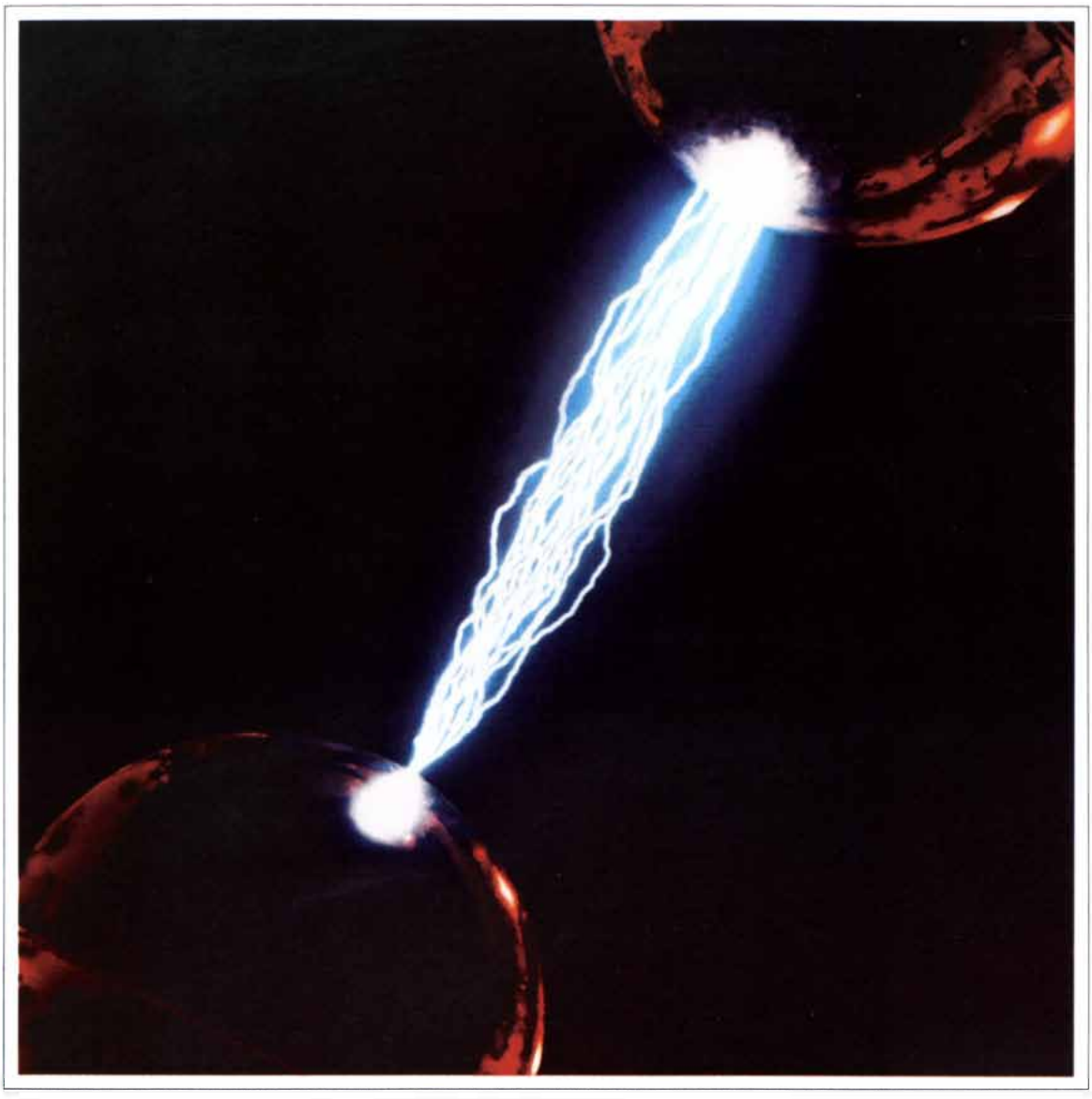
Dr. Bund, of Ruhrkohle, has put it succinctly: "The oil crisis has been an inspiration for technological advance: this is the time for positive thinking."

Turnkey.

The efficient use of available natural resources and the systematic development of manufacturing industries around the world constitute a major challenge to modern technology.
Krupp, one of West Germany's leading industrial groups with worldwide sales in 1980 of DM 15.6 billion and more than 85,000 employees, is in the forefront of technological advances across a broad spectrum of specialized industries. The Group's diversified activities range from plantmaking, steelmaking and mechanical engineering to shipbuilding, trading and services.
For example, Krupp engineers plan, design and construct turnkey plants for world markets: steel works, cement factories, coal conversion plants, water treatment plants, food processing plants, sugar mills, metal and plastic processing plants and many other complex installations requiring vast technical capabilities, know-how and experience.

KRUPP

Fried. Krupp GmbH · Altendorfer Strasse 103 · P.O. Box 10 22 52 · 4300 Essen 1 · West Germany



WestLB

Headquarters:
P.O. Box 1128,
D-4000 Dusseldorf1,
Tel. (211) 8260

Frankfurt Office
Tel. (611) 25791

Branches:
London, Tel. 6386141
New York, Tel. 754-9600;
Tokyo, Tel. 216-0581

Subsidiaries:
WestLB International S.A.
Luxembourg, Tel. 447411;
WestLB Asia Limited
Hong Kong, Tel. 5-259206

Representative Offices:
Latin-America Office
New York, Tel. 754-9620,
Rio de Janeiro,
Tel. 262 4821;
Toronto, Tel. 869 1085;
Tokyo, Tel. 213-1811;
Melbourne, Tel. 6541655

Participations:
Banque
Franco-Allemande S.A.
Paris, Tel. 563 0109;
Banco da Bahia
Investimentos S.A.
Rio de Janeiro,
Tel. 253 9723

WestLB can generate the resources needed for even the most challenging projects.

Big projects often call for financing beyond national borders.

WestLB, one of Germany's top three international banks, has built its worldwide reputation by making big money available wherever capital project needs arise. The Bank's global capacity embraces the broad spectrum of financing packages tailored to clients' needs around the world.

When next evaluating your international or domestic financing, talk first to the wholesale bankers from WestLB.

Westdeutsche Landesbank

A strong force in wholesale banking

Who will be first to sweep the skies of the world?

It could be you and Hughes Ground Systems.

We're pioneers in 3-D radar, and a world leader in deployed automated air defense systems. Throughout the NATO countries, the Far East and across the United States, our systems form protective rings around the world.

And there's more. Hughes Ground Systems' sonar, radar, communications, computers, software and displays form total interactive systems for whole countries, and we're there from inception to completion.

From undersea systems to the depths of space, there's really no limit to your career at Hughes.

In fact, today, Hughes is one of the nation's largest employers in virtually every electronic, scientific, computer and technical discipline — with 1,500 projects and a backlog of over $6 billion.

Who will be first with the land, sea and air security of the free world? It could be you and Hughes.

At Hughes Fullerton, we'll introduce you to people, ideas and jobs that could change your world. And maybe ours.

It could be you and Hughes **Ground Systems**

Call or send resume to:

Hughes Ground Systems
P.O. Box 4275, Dept. AS-7
Fullerton, CA 92634
(714) 732-7623

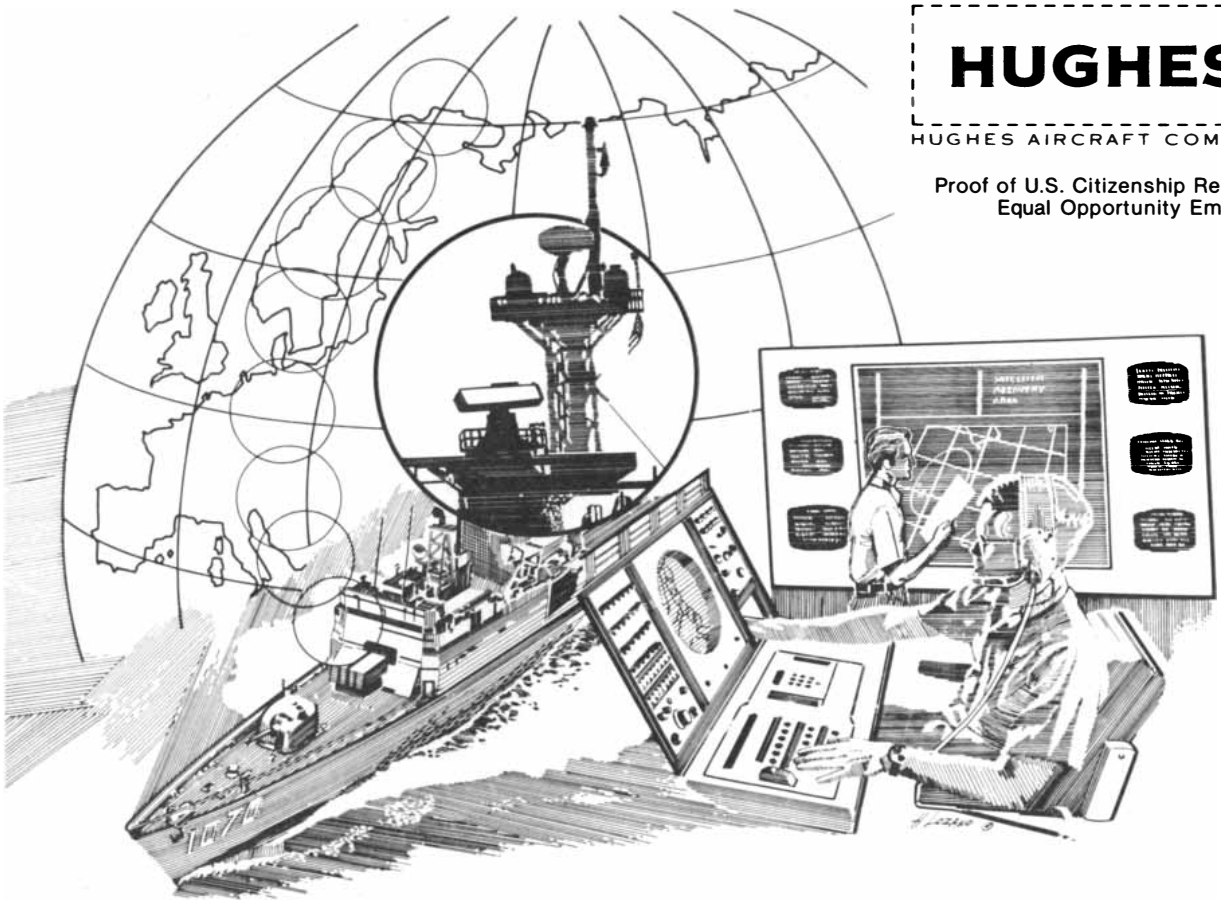
Current openings:

ASW/USW Systems Engineers
CAM/Engineers/Programmers/CAT
Circuit Design Engineers
Communication & Radar Systems Engineers
Communications Systems Engineers
Data Processing & Display Engineers
Field Engineers
Manufacturing Systems Analysts (Operations)
Programmers — Sonar, Real Time
Radar Systems Engineers
Signal Processing Engineers
Software Systems/Test Engineers
Test & Integration Engineers
Torpedo System Engineers

HUGHES

HUGHES AIRCRAFT COMPANY

Proof of U.S. Citizenship Required
Equal Opportunity Employer



Genetic Engineering in Mammalian Cells

One goal of recombinant-DNA technology is the cure of human genetic diseases. As a step toward the goal a hereditary defect is corrected by microinjecting a single gene into a mammalian cell

by W. French Anderson and Elaine G. Diacumakos

More than 2,000 human genetic diseases have been described: diseases caused by the inheritance of a defective gene coding for a defective protein. Some of these diseases can be treated but none can be cured. A cure might be effected either by correcting the mistake in DNA structure that is responsible for the disease or by transferring a normal, functional gene into the defective cells. Genetic-engineering technology makes it possible to isolate a single gene from an organism's total DNA, to "recombine" it with a carrier molecule of DNA and to introduce the recombinant DNA into a bacterial cell. From the beginning it has been clear that if genetic engineering could be extended to mammalian cells, one of the exciting applications might someday be gene therapy in human patients.

Readers of this magazine are familiar with the major recombinant-DNA techniques involved in the molecular cloning of foreign DNA in bacteria. DNA, the genetic material, is a long, very thin double helix, each strand of which is a chain of the subunits called nucleotides. Each nucleotide includes one of the four chemical groups called bases, and the sequence of those bases encodes the genetic information; a gene is a stretch of DNA encoding the information that is transcribed into messenger RNA and then translated to make a protein. In molecular cloning the DNA to be cloned is spliced into a vector molecule, usually a plasmid (a small circular piece of bacterial DNA) that carries a gene for resistance to some antibiotic. The recombinant plasmids are incubated with bacterial cells that have been treated to make them permeable. Some of the cells take up the plasmid; such "transformed" cells can be selected by the investigator because they grow in the presence of the antibiotic. Transformed cells that incorporate the desired DNA fragment are further selected by a variety of techniques. These cells can be grown in large quantities either to produce a protein encoded by the fragment [see "Useful Proteins from Recombinant Bacteria," by Walter Gilbert and Lydia Villa-Komaroff; SCIENTIFIC AMERICAN, April, 1980] or to study the structure and expression of the fragment [see "Split Genes," by Pierre Chambon; SCIENTIFIC AMERICAN, May].

Genetic engineering in mammalian cells calls for additional steps. Ways must be found to introduce the desired gene, purified and amplified by cloning in bacteria, into a mammalian cell. Then ways must be found to express the gene: to get the DNA of the gene transcribed into messenger RNA and to get the messenger RNA translated into protein. Techniques for inserting a gene into a mammalian cell have been developed. In the case of at least one gene normal expression has been achieved. Here we shall describe these results and outline some of the difficulties, largely having to do with the expression of cloned genes in mammalian cells, that must be surmounted before gene therapy for human patients is feasible.

The Transfer of Genes

Even before the advent of recombinant-DNA technology a procedure had been developed for transferring genes from one mammalian cell to another: somatic-cell hybridization, or the fusion of two body cells (as opposed to germ cells). Two lines of cells are incubated with an agent, such as inactivated Sendai virus or polyethylene glycol, that promotes fusion between cells of the two lines. At first a fused cell has two nuclei, each containing the chromosomes of one of the parent cells; in the course of cell division the nuclear membranes disintegrate and a new single nucleus is formed that contains chromosomes from both parent cells and expresses some genes of both parents. Such hybrid cells are isolated by incubation in a "selective" medium in which only hybrids can survive.

The study of hybrid cells makes it possible to associate particular characteristics with particular chromosomes and thus to map the location of genes on chromosomes. It also provides information about how the expression of genes is regulated in mammalian cells. As a hybrid cell divides it often loses chromosomes from one parent cell or both of them. Sometimes a gene that has not been expressed for a number of generations is turned on again, implying that the cell has lost a chromosome carrying a negative regulator for the gene in question. Sometimes a gene that is normally not expressed in a specialized parent cell is turned on in a hybrid. For example, the genes for the alpha and beta chains (of amino acid units) in the protein hemoglobin, which are not expressed in fibroblasts (connective-tissue cells), are turned on when a fibroblast is fused with an erythroid cell (a hemoglobin-forming cell or one of its precursor cells); some regulating factor in the erythroid cell apparently activates the genes in the fibroblast chromosome for globin: the protein chains of hemoglobin.

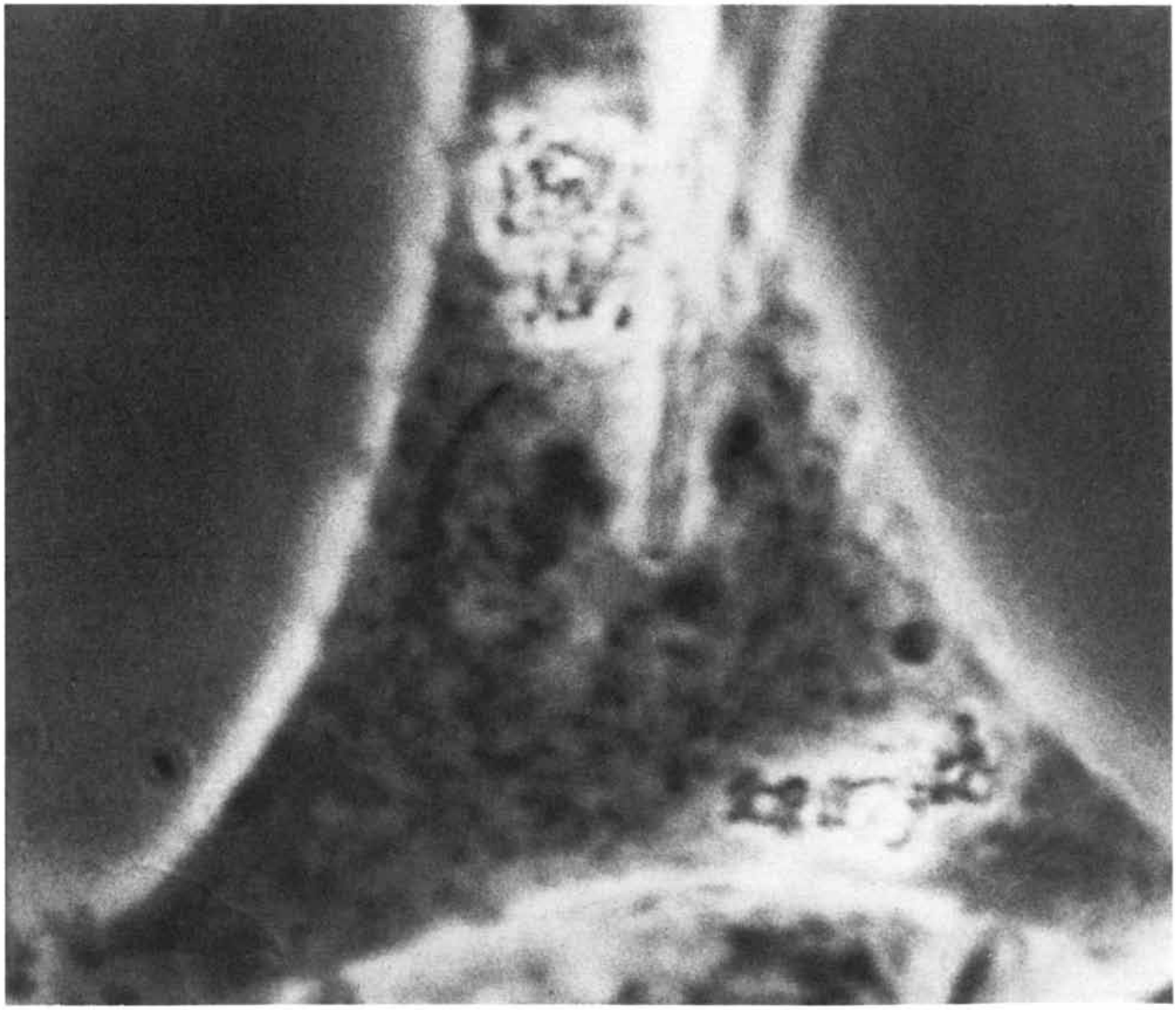
Cell fusion continues to be an effective procedure for studying gene expression, but other methods are necessary if one is to capitalize on recombinant-DNA techniques and insert a particular DNA fragment carrying a single gene into living mammalian cells. Two procedures have been developed for transferring individual genes into populations of cells. One method calls for linking the gene DNA to the DNA of an animal virus and then infecting the target cells with the virus. For example, the rabbit beta-globin gene has been inserted into the DNA of the monkey virus SV40. A laboratory line of monkey cells infected with the recombinant virus expressed the gene: they made both its messenger RNA and the protein it encodes. The human beta-globin gene is of particular interest because mutations in that gene are responsible for both sickle-cell anemia and beta thalassemia, two

hereditary disorders of red blood cells that afflict many thousands of people worldwide.

The other method is DNA-mediated gene transfer. A purified DNA fragment carrying the desired gene is mixed with a carrier DNA and is precipitated out of solution with calcium phosphate. The target cells are incubated with the precipitate, and some of them are transformed by the desired gene. By this method Richard Axel, Angel Pellicer, Saul J. Silverstein and Michael Wigler, who were all then working at the Columbia University College of Physicians and Surgeons, were able to correct a genetic defect in mouse cells. They worked with a purified fragment of DNA (from the herpes simplex virus) that carries the gene for the enzyme thymidine kinase (*TK*). A small amount of the purified DNA, mixed with several milligrams of salmon-sperm DNA as a carrier, was precipitated onto a culture of mouse *L* cells, which lack the *TK* gene; they are TK^-. About one *L* cell in a million incorporated a *TK* gene; these transformants grew and divided normally in a selective medium in which TK^- cells die. In effect they had been "cured" of their genetic defect. It is possible to transfer essentially any gene of interest either by mixing it with, or linking it to, the *TK* gene or some other gene that confers a selective advantage on transformed cells.

Microinjection

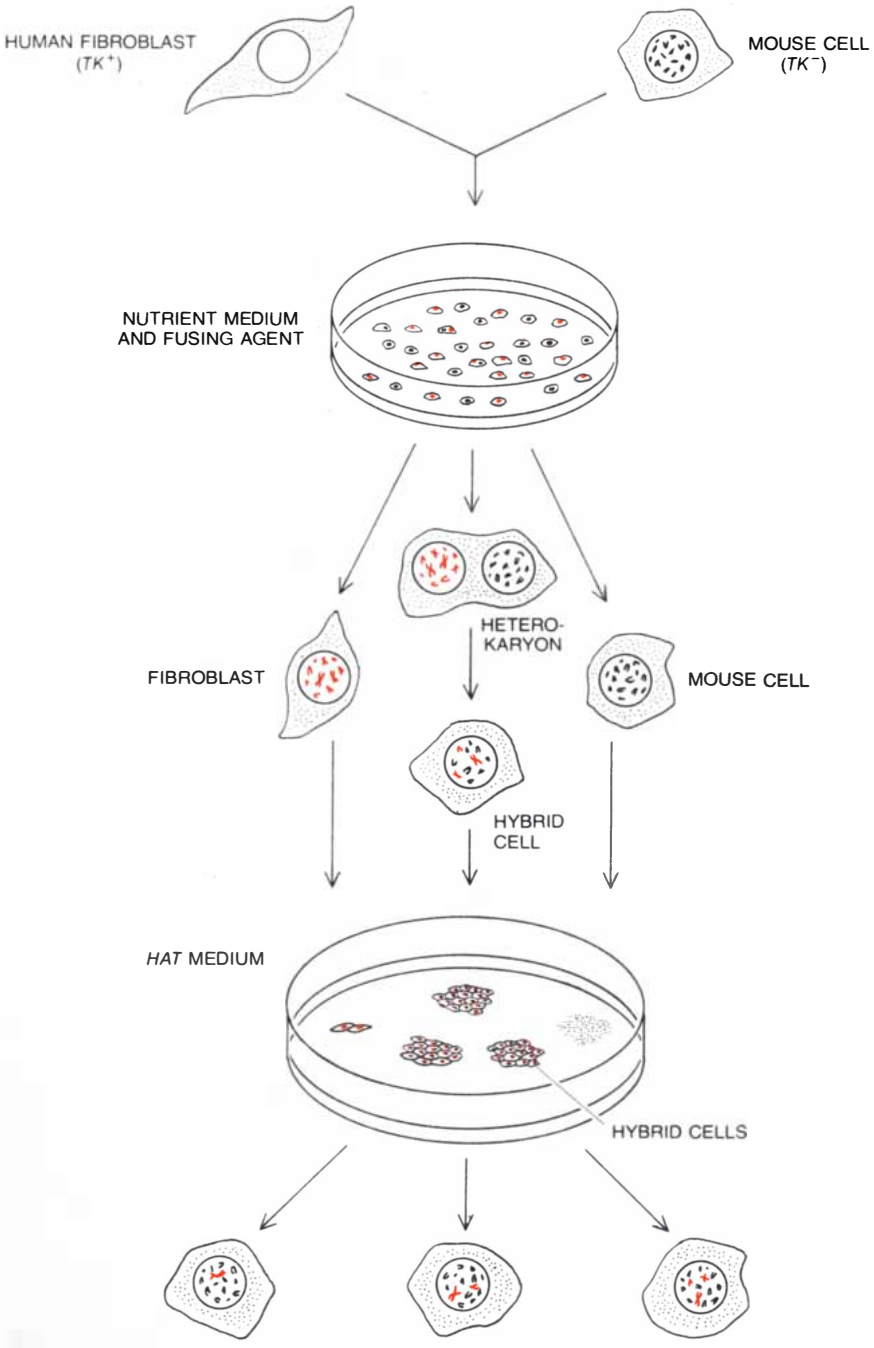
Both in the recombinant-virus technique and in DNA-mediated gene transfer the DNA of interest must be incubated with a large population of target cells in order to obtain a few cells that stably incorporate the desired gene. It seemed to one of us (Anderson) that it would be valuable to develop a procedure whereby a single copy of a specific gene could be inserted directly into the nucleus of a single cell, which could then be propagated into a large clone of identical cells, each incorporating the inserted gene. Such a procedure might begin with the injection of the desired gene into a cell nucleus through a micropipette. J. B. Gurdon, who was then working at the University of Oxford, and others had demonstrated the great investigative power of his procedures for microinjecting either entire nuclei, purified genes or specific messenger RNA into denucleated frog eggs or into oocytes, which are immature egg cells. A frog oocyte



SINGLE MAMMALIAN CELL being microinjected with DNA is enlarged 8,500 diameters in a phase-contrast photomicrograph made by one of the authors (Diacumakos). The cell, a mouse fibroblast, is suspended from a glass cover slip. The micropipette enters the field from the top; its tip is in the cell nucleus. The pipette's outer diameter is one micrometer (thousandth of a millimeter); the internal diameter is .5 micrometer. The pipette is filled with a saline solution of a recombinant plasmid carrying the gene to be transferred to the cell. An amount of solution calculated to contain (on the average) one copy of the gene is injected into the nucleus. A microinjected cell appears to recover from the procedure immediately. When the cell is put in a culture medium, it divides and proliferates to form a colony.

is about a millimeter in diameter, however; its volume is 100,000 times larger than that of a mammalian body cell. Very different techniques would therefore be required for microinjection into a mammalian-cell nucleus.

The necessary techniques had been developed by one of us (Diacumakos) in 1970. Careful examination of cells microinjected with DNA showed that they recovered immediately from the injection and subsequently grew and divided perfectly normally. A few years later, when it became possible to isolate and clone specific genes, we decided to collaborate in an effort to microinject single genes into individual mammalian cells. Adolph Graessman of the Free University of Berlin has developed a similar method; his procedure is particularly effective for analyzing the gene products of a large number of injected cells, since it can be done faster (although with less precision) than our procedure.



SOMATIC-CELL HYBRIDIZATION is one technique whereby genes can be transferred from one cell to another. Parental cells from two sources are fused by mixing them together in the presence of a fusing agent such as inactivated Sendai virus or polyethylene glycol. Fusion of two cells results first in a heterokaryon, with two nuclei (which contain the chromosomes) and a mixed cytoplasm. After cell division a single new nucleus is formed that contains the chromosomes from each of the parent cells. Fused cells are selected by incubation in a selective medium, *HAT*, in which only they survive and multiply; in the case illustrated the mouse cells, which lack the gene for the enzyme thymidine kinase (*TK*), die and the human fibroblasts divide only slowly. In cell division some chromosomes are lost; in mouse-human hybrids it is the human chromosomes that are lost. Hybrid cells can be studied to associate particular characteristics (and thus their genes) with specific chromosomes or to learn how genes are expressed (transcribed into messenger RNA, which is translated into protein) in a foreign environment.

We decided to transfer a copy of a *TK* gene and a copy of a beta-globin gene into a TK^- mouse *L* cell (which does not make globin) and see whether the cell's *TK* deficiency could be corrected genetically and whether the globin gene could be replicated and expressed. We chose this combination because we knew that the *TK* gene had been transferred to TK^- cells by DNA-mediated gene transfer and that other DNA fragments could be cotransferred with the *TK* gene. There should be a reasonable chance of success if we made a solution in which both the *TK* gene and the globin gene were present at a concentration of one molecule per 10^{-11} milliliter (one hundredth of a picoliter) and then injected 10^{-11} milliliter of the solution into each nucleus. (Actually, of course, a given injection dose could contain anywhere from no DNA molecules at all to several molecules.)

The plan was to insert a *TK* gene from herpes simplex into a copy of the plasmid designated pBR322 and to insert a human beta-globin gene into another copy of the same plasmid. This particular vector plasmid has genes conferring resistance to two antibiotics: ampicillin and tetracycline.

The Experiment

To make a recombinant plasmid one exploits one of a group of enzymes known as restriction endonucleases. Each restriction enzyme recognizes a particular short sequence of bases and cleaves the DNA at a particular site in that sequence. DNA's cleaved by the same restriction enzyme therefore have matching ends and can be ligated. Lynn W. Enquist of the National Cancer Institute, William C. Summers of Yale University and their colleagues isolated a herpes simplex DNA fragment, 3.5 kilobases (thousands of bases) long, carrying the *TK* gene by cleaving the viral DNA with the endonuclease *Bam*HI. In our laboratory at the National Heart, Lung, and Blood Institute, Peter J. Kretschmer cut open the circular plasmid pBR322 with the same endonuclease. Then Enquist and his colleagues ligated the two molecules to create a recombinant plasmid designated pX1. Because *Bam*HI cleavage disables one of the plasmid's two antibiotic-resistance genes, pX1 confers resistance to ampicillin, but not to tetracycline, on any bacterium in which it replicates.

A recombinant plasmid incorporating the globin gene was prepared in a sim-

The all-new Toyota Cressida. Lavish. Extravagant. Uniquely European in look. This is a sedan different from any Toyota you have seen before.

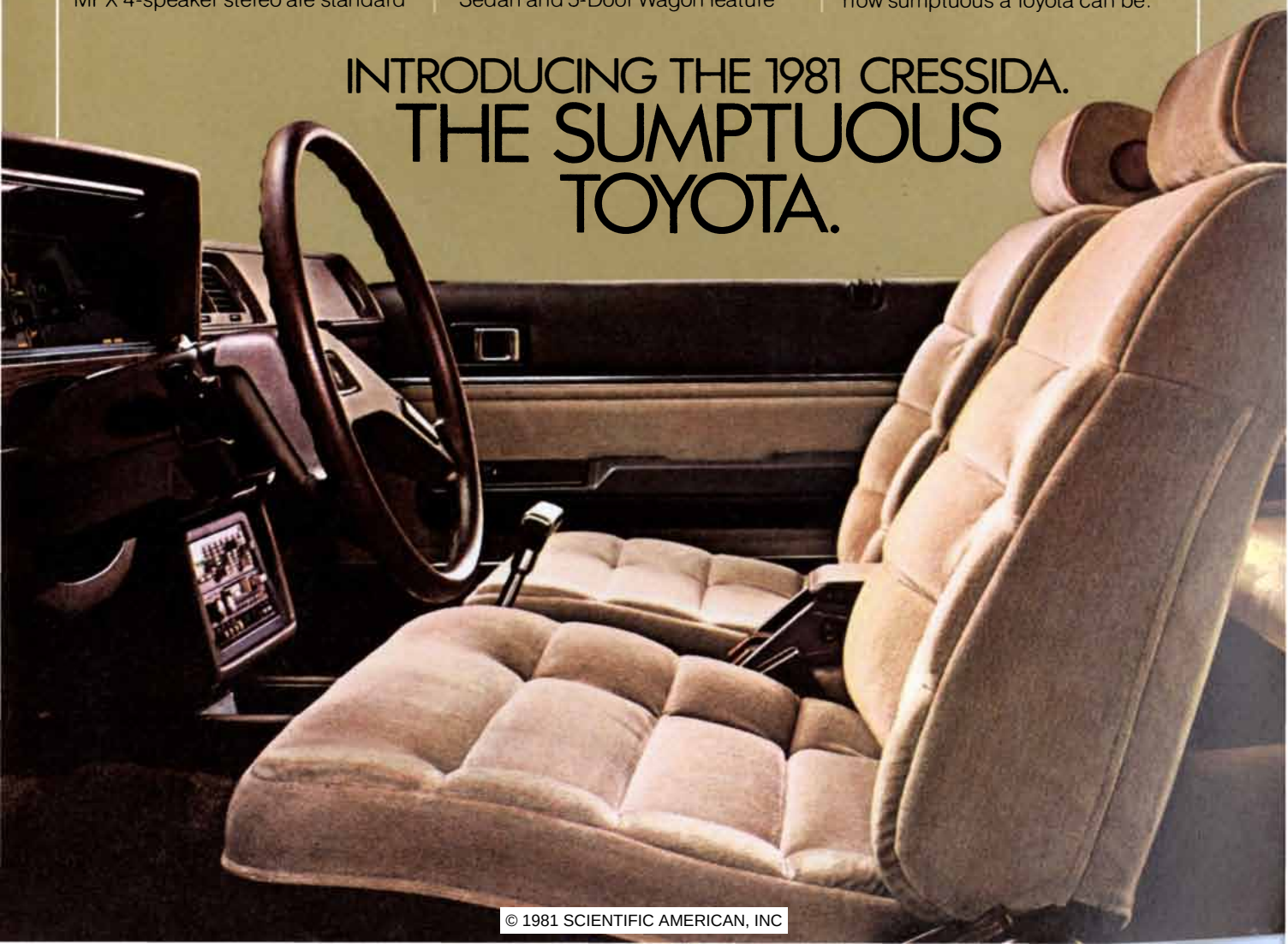
Inside, the new Cressida offers every comfort and convenience feature you could expect in the world's highest-priced automobiles—and then some. Power windows, cruise control, climate control and AM/FM MPX 4-speaker stereo are standard equipment, of course.

But Cressida's state-of-the-art engineering includes much more. Like a unique automatic shoulder-seat belt system. The belt is guided around you by an electric motor! And an optional cassette player with amplifier and graphic equalizer, for the ultimate in automotive sound.

Performance of the Cressida is also impressive. Both the Cressida Sedan and 5-Door Wagon feature electronic fuel injection and an innovative 4-speed automatic overdrive transmission, to enhance the new 2.8 liter engine's efficiency. MacPherson strut front suspension assures a soft, controlled ride.

The Toyota Cressida. If you're wondering why you should buy a luxury car from a company famous for economy cars, you need do only one thing. Drive a Cressida, and feel how sumptuous a Toyota can be.

INTRODUCING THE 1981 CRESSIDA.
THE SUMPTUOUS TOYOTA.



New, improved Allied.

By now, you probably know our name is new (it used to be Allied Chemical).

So what else is new?

We lost our losers.

Some of our operations just weren't profitable. So we sold them. And some operations lacked potential for leadership or growth. We sold them, too. Sometimes, you have to shrink to grow

Fewer operating parts.

We turned nine dependent divisions into four self-sufficient operating companies. That move let us trim our corporate staff from 1750 to 450 people. And saved us $30 million a year.

Saving is good. Earning is better.

1980 was a year for setting earning records. Our net profit hit a record $289 million. Our return on assets jumped one-third. And we wiped out our entire domestic short-term debt. All of which boosted our earnings per share 37% to a record $8.15.

And while we're on the subject, guess what our first quarter earnings did in 1981. They set another record.

"Our research is our future."

Every corporation talks like that. But how many corporations double their research programs in two years? *We* did. From $70 million in 1979 to $145 million this year. During the next five years, we'll spend $900 million.

What are we spending all that money on? Bright ideas like Metglas,® our new metal alloy that can cut energy losses in electrical equipment by 75%. And our synthetic alexandrite lasers. (We're developing them for NASA for possible satellite weather forecasting.) And genetic engineering for new agricultural products.

The future's the future. What about now?

It can take ten years for an invention to become a product. So, to keep growing right now, we're expanding old markets and opening new ones for current products.

Our new Anso IV™ carpet fiber, the first with built-in soil and stain resistance, has taken its market by storm. Our new Linotron™ 202E phototypesetter sets over 6,000 Japanese characters, opening a vast market for us in the Far East. And our big new offshore Louisiana gas field has just gone into production.

What we do with what we earn.

This year, we're spending $600 million of it. On capital expenditures. Because long-term growth means more to us than short-term profit.

Our master plan.

It happens to be one of our most valuable assets. Because it keeps us moving straight toward our goals:

To grow our present businesses. To diversify with inventions and acquisitions. And to increase our profits at a steady rate.

We're going to achieve those goals the only way we know how:

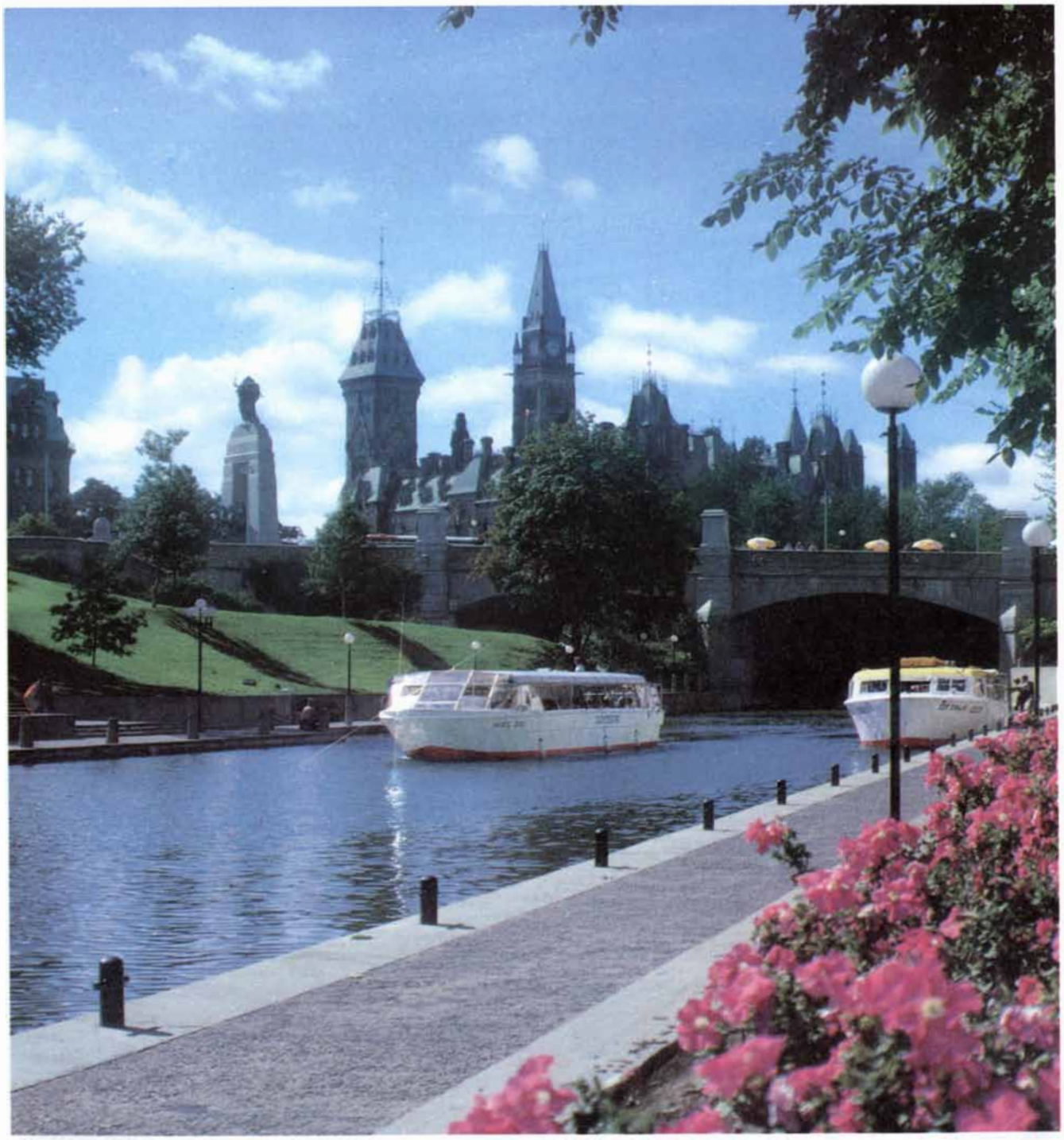
By rolling up our sleeves and getting down to business.

That way, we'll stay new and improved.

ALLIED Corporation

We mean business.

"capital sights"



Savour the old world ambience of Ottawa, Canada's capital city, the splendours of our heritage in art, architecture, pomp, pageantry and spectacular Changing the Guard ceremonials on Parliament Hill. The pleasures of riverside parks and promenades, canals and winding waterways, bicycle paths and garden walks – yours to discover this summer in Ottawa, Ontario/Canada.

Call us COLLECT (416) 965-4008 or write: Ontario Travel, Dept. C.S., Queen's Park, Toronto M7A 2E5.

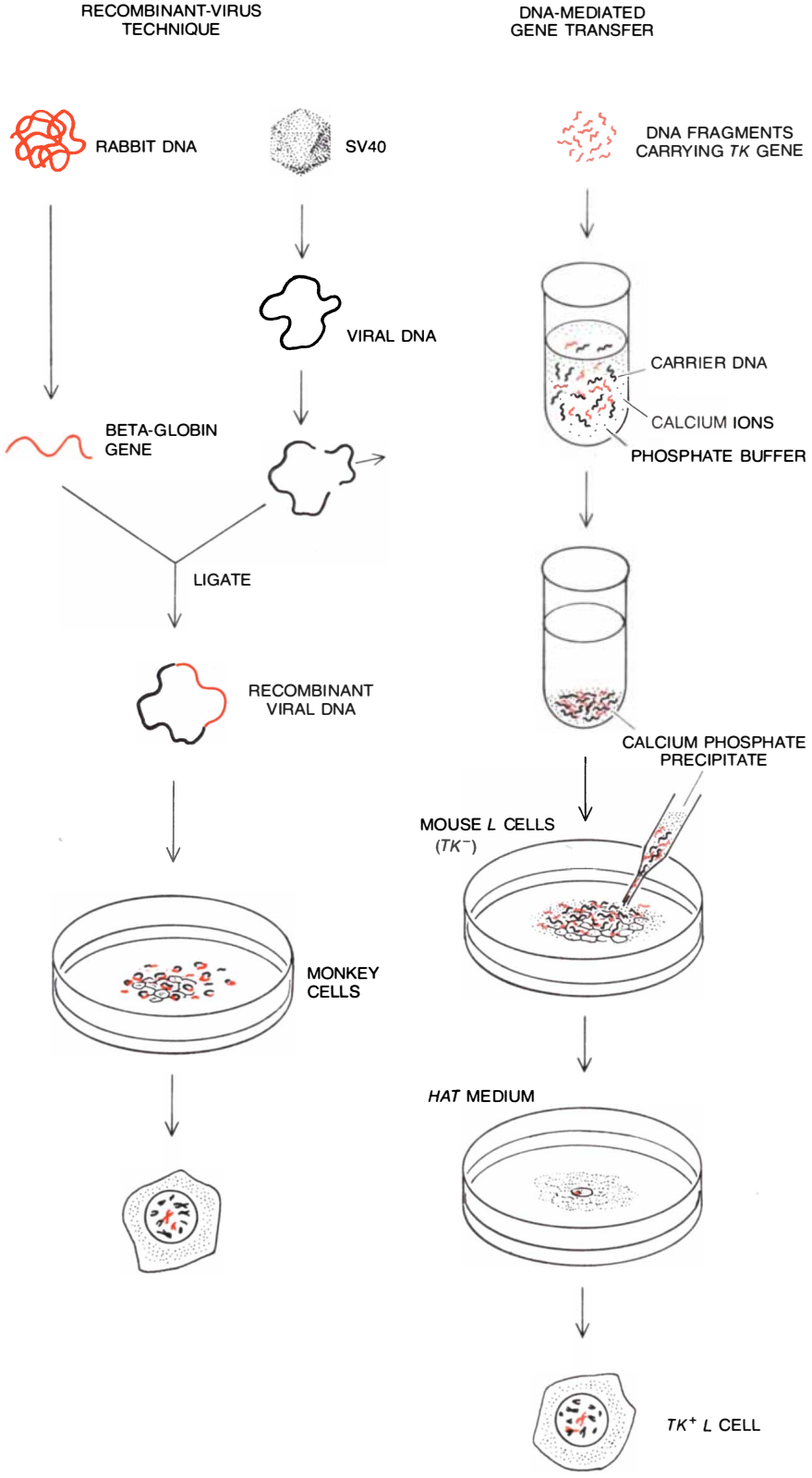
ONTARIO
yours to discover!

ilar way by Russell E. Kaufman, Arthur W. Nienhuis and Kretschmer. They started with a 15.9-kilobase segment of human DNA that included the beta-globin gene, which had been isolated by Thomas P. Maniatis of the California Institute of Technology and his colleagues. They cleaved from it, with the enzyme *Pst*I, a smaller 4.4-kilobase segment carrying the gene, and they inserted the segment into the *Pst*I-cleaved opening of another copy of the plasmid pBR322. The resulting recombinant plasmid, pRK1, confers resistance to tetracycline but not to ampicillin. The two recombinant plasmids were cloned separately in the bacterium *Escherichia coli,* isolated and purified. That supplied us with two pure preparations of plasmids, one carrying the *TK* gene and one carrying the gene for beta globin.

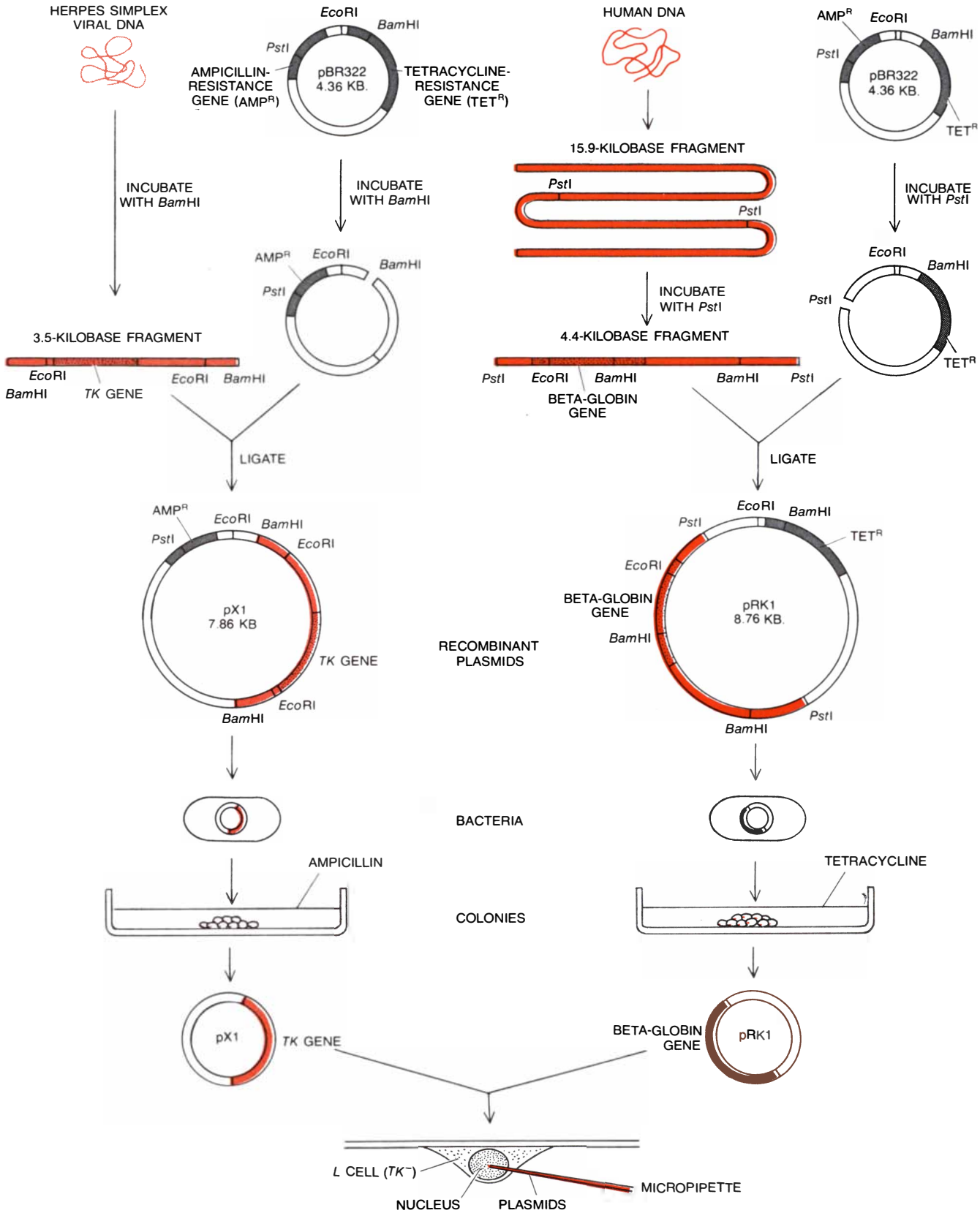
Microinjection was done by Lillian Killos, a George Washington University graduate student working in our laboratory. She seeded several dozen *TK* mouse *L* cells onto a glass cover slip, which was placed upside down on a glass slide and separated from the slide by small glass supports. The chamber thus formed was filled with a culture medium, its edges were sealed with silicone oil and it was placed on the stage of a binocular phase-contrast microscope under 2,000-fold magnification.

A micropipette whose tip had an internal diameter of .5 micrometer (five ten-thousandths of a millimeter) was made on a specially built "microforge." Several cells in each of a number of chambers were injected with (on the average) one of each kind of plasmid. The cover slips were removed from the chambers and the cells on them were incubated first in a normal medium and then in a standard medium, called *HAT*, in which only cells making thymidine kinase can survive. Three cover slips from the first series of injections yielded seven dividing colonies, which were transferred to individual Petri dishes and grown for analysis.

The fact that these cells had survived in *HAT* medium showed they had a functional *TK* gene. Was it a *TK* gene that had arisen by a "back" mutation in a normally TK^- *L* cell or (as we hoped) was it a microinjected herpes simplex *TK* gene we had injected that was replicating and being expressed? Linda Sanders-Haigh, another George Washington graduate student, established that the vital *TK* gene was physically present in the *L* cells by "Southern blot" analysis, a procedure devised by E. M. Southern of the University of Edinburgh. The total DNA of one of the successfully microinjected colonies, C2B, was digested with the restriction enzyme *Bam*HI. The resulting fragments were subjected to gel electrophoresis, which separates DNA fragments according to their size. If the 3.5-kilobase fragment of herpes simplex virus containing the



TWO ESTABLISHED TECHNIQUES for transferring a particular gene into a mammalian cell are illustrated. In the recombinant-virus technique (*left*) the genetic material DNA is extracted from a virus such as the monkey virus SV40. A segment of the DNA is cut out and is replaced by the DNA fragment of interest, in this case a fragment, cleaved from rabbit DNA, that includes the gene coding for the beta chain of amino acid units in the hemoglobin molecule. When monkey cells are infected with the recombinant virus, the gene for globin (the protein of hemoglobin) replicates with the viral DNA and is expressed. In DNA-mediated gene transfer (*right*) DNA fragments carrying the gene for *TK* (and perhaps another gene to be transferred) are mixed with a carrier DNA in a phosphate buffer solution. Calcium ions are added, forming a calcium phosphate precipitate that traps the DNA. The precipitate is pipetted onto a culture of TK^- cells, which lack the gene for *TK*. About one cell in a million takes up the *TK* DNA and expresses the gene; only these "transformed" cells survive in *HAT* medium.



RECOMBINANT PLASMIDS carrying the *TK* gene (*bottom left*) and the gene for human beta globin (*bottom right*) are prepared for microinjection. In each case both a source of the genes to be injected (herpes simplex virus Type 1 DNA or human DNA) and a vector plasmid (pBR322) are cleaved with a restriction endonuclease, an enzyme that cuts the double helix of DNA at a unique site in a particular short sequence of bases (the chemical groups that characterize the subunits of the DNA strand and whose sequence encodes genetic information). Cleavage sites and other restriction-enzyme sites are indicated. Lengths of molecules are given in kilobases (thousands of bases). The vector plasmid carries genes conferring resistance to two antibiotics; cleavage disables one of those genes. Since a gene fragment and the plasmid into which it is to be inserted are cut with the same endonuclease, they have matching ends; the ends are ligated, forming a recombinant plasmid that incorporates the desired gene: for *TK* (*left*) and for human beta globin (*right*). These plasmids are introduced separately into the bacterium *Escherichia coli;* transformed bacteria are selected by their resistance to an antibiotic. Each transformed cell gives rise to a clone: a colony of cells each of which houses the identical plasmid. The plasmids are extracted for microinjection.

TK gene was present and was replicating in the C2B DNA, its many copies should form a discrete band on the gel.

The Southern procedure for finding such a band depends on the fact that two single strands of DNA with complementary base sequences will hybridize, that is, match up and bind to each other. The electrophoresed DNA is "blotted" onto cellulose nitrate paper and a known DNA, labeled with a radioactive isotope, is applied to the paper as a probe. All the DNA is then denatured, or separated into single strands. Probe strands hybridize with any DNA having a complementary base sequence, and their radioactivity reveals the location of the sought-after band.

When Sanders-Haigh used as a probe a labeled sample of the original 3.5-kilobase *TK*-containing fragment, it revealed a radioactive band at the position on the paper corresponding to 3.5 kilobases, indicating that the injected *TK* gene was present in the C2B DNA and was being replicated in the colony's cells. As a further check the C2B DNA was digested with another endonuclease, *Eco*RI. The fact that two *Eco*RI sites within the 3.5-kilobase fragment are 2.4 kilobases apart meant this digestion should produce a 2.4-kilobase band, which would be revealed by hybridization with part of the radioactive probe DNA. Such a band was found. Finally double digestion of the injected colony's DNA with both *Bam*HI and *Eco*RI showed the same 2.4-kilobase band, indicating that (as we expected) there was no *Bam*HI site within the 2.4-kilobase *Eco*RI fragment. These results and others established that the herpes simplex *TK* gene was present and was being replicated in the microinjected cells.

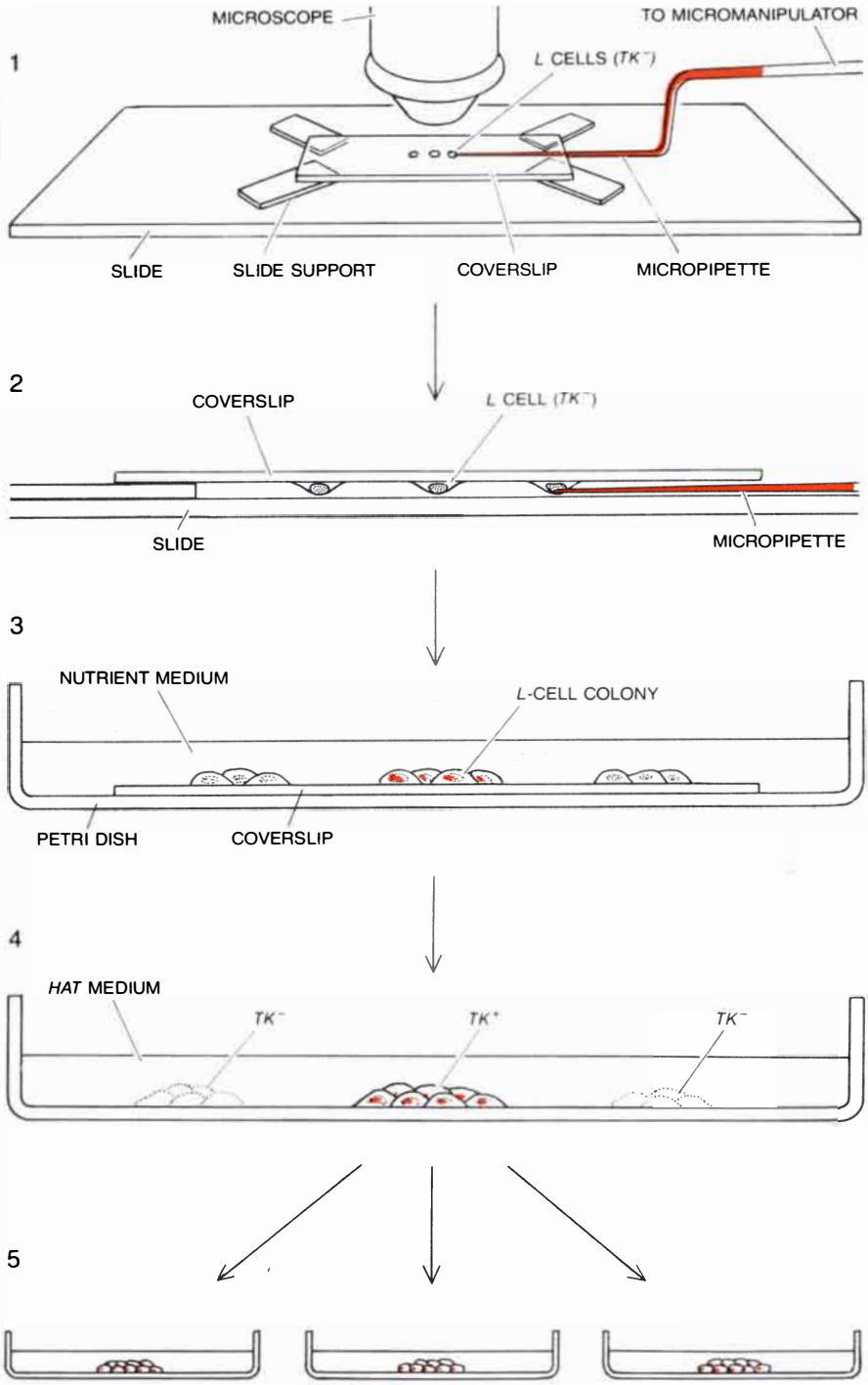
A similar study established the presence in the same cells of the human beta-globin gene; both genes had been successfully microinjected, and they were being replicated as the injected cells proliferated. Were they being expressed? Were they being transcribed into messenger RNA that was being translated to make functional herpes simplex thymidine kinase and normal human beta globin?

Testing for Expression

William Summers and Wilma P. Summers at Yale have devised a way to show that the thymidine kinase specified by the herpes simplex *TK* gene is present in a cell and to distinguish it from the same enzyme made by the mouse cell's own DNA. The viral enzyme can convert a precursor molecule into the nucleotide cytidylic acid; mouse thymidine kinase cannot. Mouse cells do, however, have another enzyme, a deaminase, that can make the conversion possible. There is an inhibitor, called *THU,* that blocks the mouse deaminase. The conversion of

the nucleotide precursor into the nucleotide in the presence of *THU* is therefore evidence for the presence of the herpes simplex thymidine kinase. Two of our colonies were tested at Yale. Both colonies were shown to be expressing the microinjected herpes simplex *TK* gene, not the mouse gene. In other words, the TK^- defect in a mouse *L* cell had been corrected by the injection into its nucleus of a functional viral gene. This constituted successful genetic engineering at the level of a single mammalian cell.

Perhaps that was not surprising; after all, we had specifically selected cells that



MICROINJECTION is carried out under the phase-contrast microscope (*1*). A cover slip carrying mouse *L* cells lacking the *TK* gene is inverted over a slide to form a chamber, which is filled with a culture medium and sealed with silicone oil. A micropipette filled with a solution containing the recombinant plasmids is inserted into the nucleus of a cell and an amount of solution containing about one plasmid of each kind is injected (*2*). The cover slip is placed in a culture medium and the cells divide (*3*). Successfully injected cells survive when the colonies are transferred to the selective medium *HAT* (*4*). Each surviving colony is transferred to an individual dish and is expanded to provide a large number of cells for subsequent analysis (*5*).

Picking options for your new car?
Don't forget Ford's Ultimate Option. The only one that could pay for itself.

FORD'S ULTIMATE OPTION—E.S.P.

Ford EXTENDED SERVICE PLAN

Ford Motor Company's Extended Service Plan is designed to offer you long-term protection against rising service costs.

"How long is long-term protection?"
Just about as long as you want. There are three plans to choose from. You can purchase a maximum coverage plan of 3 years/36,000 miles or 5 years/50,000 miles. The third plan is the power train plan of 5 years/50,000 miles.

"Protection on what?"
Protection against the cost of major repair bills on your Ford-built car or light truck. Your Dealer can give you specific information about each plan.

"How does E.S.P. work?"
You pay just a $25 deductible per eligible repair visit under either of the two maximum coverage plans, or $50 under the power train plan...regardless of the cost of all labor and parts needed.

"What if they tie up my car?"
If your car needs to be kept overnight for repair under the original new vehicle warranty or the repair of E.S.P.-covered components thereafter, you're provided a rental car reimbursement allowance of up to $15 per day (excluding mileage) for up to five days.

"What if I sell my car while it's still covered?"
For a small fee, the remaining coverage is transferable to the new owner. This can be a strong selling point.

"Suppose I'm out of town..."
The Ford Extended Service Plan is the one and only plan that's honored by more than 6000 Ford and Lincoln-Mercury Dealers throughout the U.S. and Canada. That can mean real peace of mind wherever you are.

"Okay...but is it expensive?"
The cost of E.S.P. is surprisingly low. And it can be added to your monthly payment if you finance your new vehicle. When you consider the low cost of protection on your initial investment, and how long you'll probably keep your car, it just makes good sense to buy Ford's Ultimate Option...E.S.P.

Ask for Ford Motor Company's Extended Service Plan by name.

carried a *TK* gene by growing them in the selective medium *HAT*. When we looked for beta globin in the colonies grown from microinjected cells, the news was not as good. We did not find any of the protein. We therefore looked for the messenger RNA specifying globin, the presence of which would be evidence for at least partial expression of the beta-globin gene. Sanders-Haigh identified RNA sequences coding for human beta globin by doing a hybridization test. The results indicated that only between two and 10 molecules of the globin RNA were present in each mouse *L* cell, which is not a significant amount biologically.

The result of our microinjection of the globin gene was roughly the same as what Axel and his colleagues had reported after inserting a rabbit beta-globin gene into mouse *L* cells by means of DNA-mediated gene transfer; they too had found only a few messenger-RNA molecules per cell that coded for rabbit globin. Moreover, both Axel and Charles Weissmann of the University of Zurich, who did similar experiments, found specifically that one end (the so-called 5′ end) of the globin RNA molecule was sometimes missing in the mouse cells, so that no complete globin molecules could be synthesized.

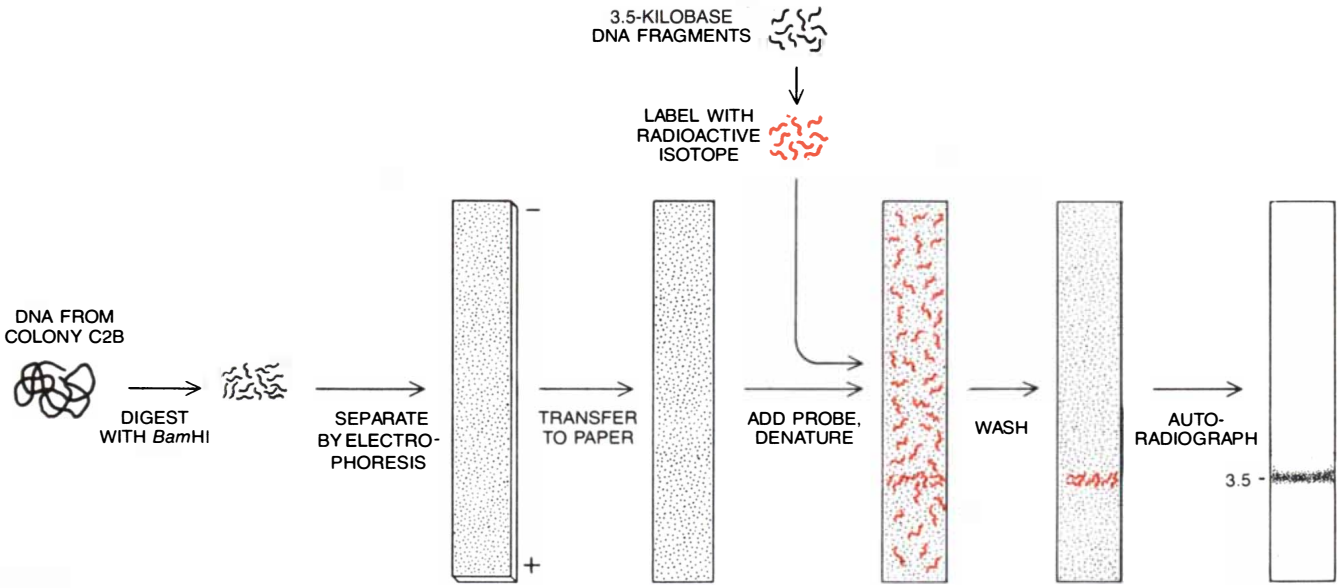
Plasmid Rescue

From the time of the first Southern-blot analysis of the injected cells' DNA we had been puzzled by an unusual observation. We had assumed that the microinjected *TK* and globin genes would probably be integrated into the mouse-cell genome (the total complement of genetic material, divided among several chromosomes) at several sites; in that case the various probes would reveal an array of radioactive bands of unpredictable size. It turned out, however, that most of the bands we saw (no matter which colony was analyzed, which restriction enzyme had generated the electrophoresed fragments and which radioactive probe was used) could be accounted for by DNA fragments from one or the other of the two original recombinant plasmids, pX1 or pRK1. The pattern of the bands differed somewhat from colony to colony and some of the bands clearly did not come solely from the microinjected plasmids. Why, however, should the bands apparently arising from the original plasmids be so prominent?

We could think of at least two explanations. One would be the presence in the colonies of injected cells of multiple copies of each gene-carrying plasmid in a long tandem-repeat pattern, so that digestion with an endonuclease would generate the same bands that would have been generated by the original circular plasmids; such tandem repeats might be integrated into the mouse genome or might be present extrachromosomally as concatamers (linear chains). The second explanation would be that some or all of the plasmid DNA (pBR322 with the inserted gene) was present in the circular extrachromosomal plasmid form, replicating (or at least existing) autonomously in the *L* cells.

We did a number of experiments designed to determine the structure of the plasmid DNA in the mouse *L* cells. One line of investigation led to unexpected results. We wanted to know if the original plasmids pX1 and pRK1 could be rescued, or recovered, from the DNA isolated from microinjected *L* cells. This could be done by adding the *L*-cell DNA directly to *E. coli* and looking for transformed bacteria; these could be distinguished by their resistance to an antibiotic (tetracycline in the case of pRK1, for example). As a control Kretschmer and Anne H. Bowman first incubated *E. coli* cells, which were sensitive to both tetracycline and ampicillin, with DNA from mouse *L* cells that had not been microinjected. They grew the exposed cells in a medium containing tetracycline and looked for resistant cells. None were found. Then they treated the same antibiotic-sensitive *E. coli* with DNA from injected colony C2B and again screened for resistant transformants. They found 42 tetracycline-resistant colonies. Each of these transformed colonies was expanded and the plasmid responsible for the antibiotic resistance was isolated. Each isolated plasmid was thereupon digested with several restriction enzymes. The pattern of fragments thus cleaved was analyzed to prepare a restriction-enzyme map of each plasmid, which could be compared with the maps of the original plasmid, pRK1.

The plasmid isolated from 30 of the transformed colonies turned out to be



FOREIGN DNA replicated in colonies grown from microinjected cells is detected by the "Southern blot" procedure. DNA from the microinjected colony C2B is digested with the restriction enzyme *Bam*HI. The fragments are separated according to size by electrophoresis on an agarose gel; the smaller the fragments are, the farther they migrate toward the positive pole of the gel. The fragments are "blotted" onto cellulose nitrate paper. The *TK*-bearing 3.5-kilobase *Bam*HI fragment of herpes simplex DNA, labeled with a radioactive isotope, serves as a probe for detecting any fragments having a base sequence complementary to that of the probe. The probe is applied to the paper and the paper is boiled to denature the DNA, so that the strands of the double helix are separated. Now the probe DNA binds to fragments that have the complementary base sequence; the rest of the probe DNA is washed away. Autoradiography reveals a radioactive band on the paper at a position corresponding to a molecular size of 3.5 kilobases, showing that 3.5-kilobase fragments homologous to the viral-DNA probe are present in colony C2B either in the original circular-plasmid form or as long tandem repeats called concatamers.

identical with pRK1; in other words, the original plasmid pRK1 could be rescued. The other 12 tetracycline-resistant colonies yielded surprising results, however. There were five different kinds of recombinant plasmid molecule, four of which appeared to be dimers or trimers of pRK1: circular molecules made up of two or three linked copies of the original plasmid. They were not complete dimers or trimers, however. In each case there was a deletion: part of the plasmid was missing. The fifth kind of unusual plasmid was still more complicated in structure. The reader will remember that both pX1 (carrying the *TK* gene) and pRK1 (carrying the globin gene) had been microinjected into the same mouse *L* cell. Could we recover from the transformed colonies any DNA molecules that were recombinants of the two original plasmids? Kretschmer and Bowman had identified pRK1 transformants by their ability to grow in the presence of tetracycline. They went on to screen those tetracycline-resistant colonies for resistance to ampicillin (which would have been conferred by pX1). They found four colonies that were resistant to both antibiotics. Each colony contained the same plasmid, which appears to consist of one complete copy of pRK1 and part of pX1.

Examination of all five kinds of recovered plasmid shows that the characteristic deletion begins at or near the same position in the original vector plasmid pBR322. The site is near the end of the region encompassing the tetracycline-resistance gene. Apparently this region has become a "deletion hot spot," a site whose base sequence somehow favors deletion. There had been no report that any such specific deletion-susceptible region is present in the original vector. The implication is that the susceptible region arose in the DNA's passage through the mouse *L* cells.

In the course of that passage was the plasmid DNA integrated into the mouse-cell chromosomes or was it free as extrachromosomal material? Michael H. Huberman carried out a procedure, devised by Bernard Hirt of the Swiss Institute for Experimental Cancer Research in Lausanne, that was likely to identify extrachromosomal forms of the injected DNA that might be present in the microinjected *L* cells. The *L*-cell DNA was fractionated by spinning in a centrifuge, so that large (chromosomal) DNA molecules were sedimented into a pellet, whereas molecules smaller than about 100 kilobases in length (extrachromosomal DNA) remained suspended in a supernatant liquid. The supernatant DNA proved to be more effective (microgram for microgram) in transforming *E. coli* than the pellet material. This shows that at least some of the pRK1 DNA in the injected cells is extrachromosomal.

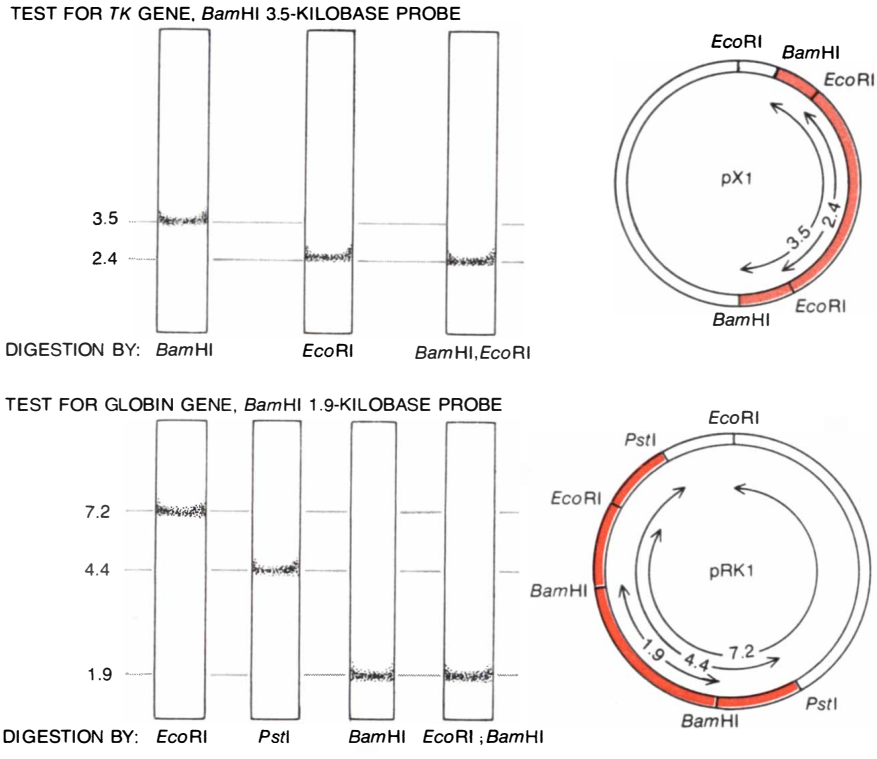
Our present feeling, based on several lines of evidence from our laboratory and others, is that pRK1 and pX1 molecules combine and are replicated within the mouse cell, probably as linear tandem repeats (concatamers). Some concatamer material may be integrated, but at least a portion of it is extrachromosomal; a few circular plasmids may also be present. The deletion events giving rise to the unique recombinant plasmids probably take place in the bacterium in the rescue phase of the procedure.

Regardless of where these events take place, however, clearly something is happening to the DNA of the original vector plasmid while it is in the mouse cell such that a region highly susceptible to deletion is generated. If the deletion turns out to be effected in the mouse cell, there must be some mouse-cell enzyme that promotes deletion in the susceptible region. If it takes place only later, in the bacterium, then presumably a mouse enzyme has modified the site in such a way that it is later recognized by a bacterial enzyme as a site for deletion. This is an intriguing possibility. It implies that DNA-modifying enzymes in mammalian cells are able to recognize particular sites in bacterial DNA (or perhaps in any DNA) and to alter those sites in very specific ways (without changing the base sequence) that affect subsequent recognition of the sites by other specific enzymes. Some such process might be exploited in the developing embryo for controlling the expression of genes in differentiating cells.

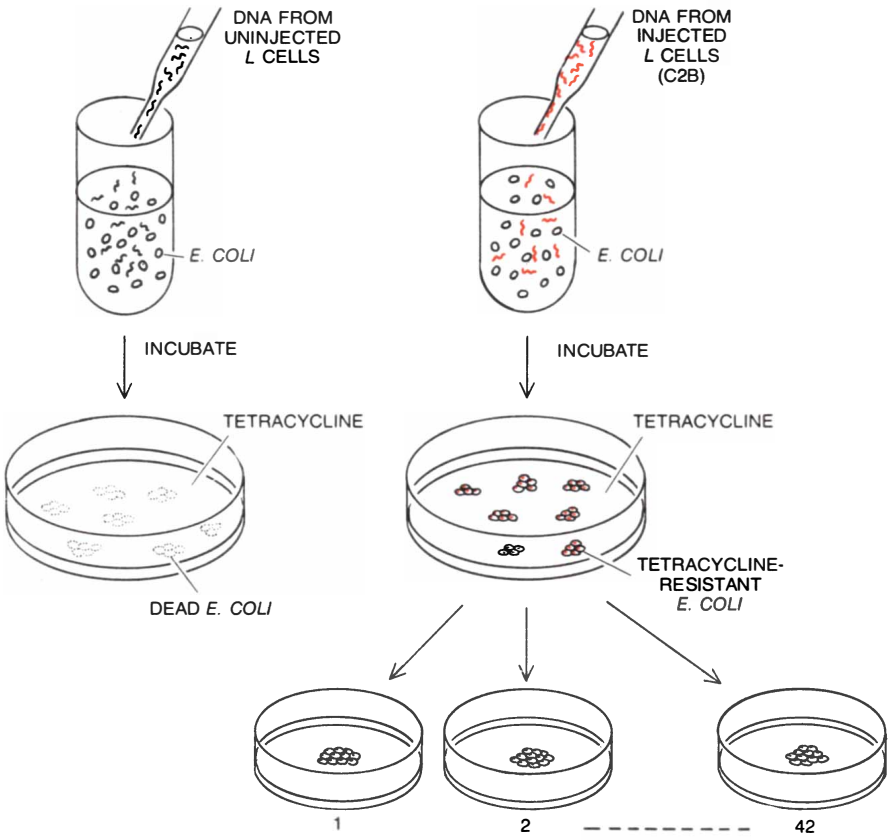
Globin-Gene Expression

Even if the form in which the injected genes are present in the *L* cells is in doubt, the fact remains that the *TK* gene appears to be expressed normally. Why is the globin gene only minimally expressed? The first answer might be that *L* cells are not erythroid cells and therefore may well lack certain regulatory factors necessary for the induction of globin-gene expression. Yet even when the human beta globin is transferred into a mouse erythroleukemia (*MEL*) cell, which can make 10,000 molecules of mouse-globin messenger RNA, the human globin gene is not induced. Where is the obstacle?

Thymidine kinase is active in every cell because it is involved in the synthesis of DNA. Perhaps its regulation is very simple: perhaps it is always "on." The human beta-globin gene, on the other hand, is part of a large region of the genome that codes for a number of different globin chains. In erythroid cells various globin genes are switched on and off sequentially as the organism develops from an embryo to a fetus to a newborn infant. Perhaps various regulatory regions are scattered throughout the 65-kilobase (or larger) globin region, all (or at least most) of which are required for proper regulation of the beta-



SOUTHERN-BLOT RESULTS are displayed (*left*) with maps of the injected plasmids (*right*) that show the source of the fragments detected. In each case the probe and the digesting enzyme are indicated. The results demonstrate that both of the injected genes, the *TK* gene (*top*) and the gene for human beta globin (*bottom*), are represented in cells of the injected colony.



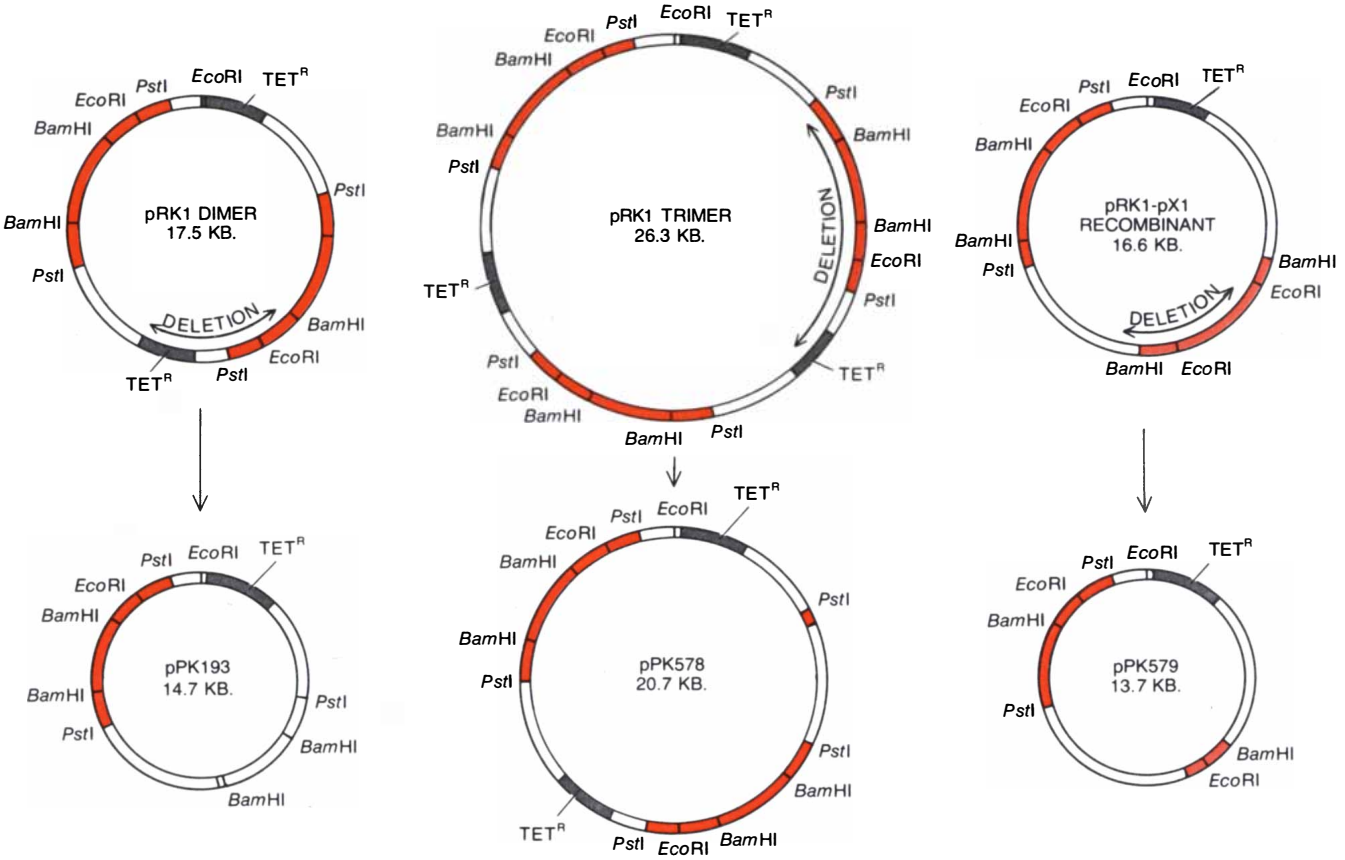
PLASMIDS WERE RESCUED, or recovered, from the DNA of an injection-cell colony by transforming *E. coli* bacteria with the DNA from the colony. The plasmids were detected by their ability to confer on the bacteria resistance to tetracycline. As a control the DNA of uninjected cells was used to transform bacteria (*left*); no resistant bacteria were recovered. Incubation of *E. coli* with injected-cell DNA (*right*) produced 42 tetracycline-resistant colonies. The plasmids responsible for each colony's antibiotic resistance were isolated and characterized.

globin gene. If that is the problem, one solution would be to inject not just a small DNA fragment incorporating the beta-globin gene itself but rather the entire globin region; at present, however, such large fragments cannot be cloned. An alternative might be to find and isolate all the necessary regulatory regions, hook them together and attach them to the beta-globin gene; the transfer of such a recombinant molecule might then achieve normal expression of a foreign globin in an erythroid cell.

Recent studies with somatic-cell hybrids suggest the disconcerting possibility that such manipulations may not be enough. Albert B. Deisseroth, now at the National Cancer Institute, and Marcia C. Willing of our laboratory found that when a mouse erythroid cell and a human erythroid cell are fused, the hybrid synthesizes both mouse and human hemoglobin chains. Even if the mouse erythroid cell is fused with a human lymphoblast (which does not make hemoglobin but is descended from a stem cell that can also give rise to an erythroid cell), Deisseroth found that both mouse and human hemoglobin chains are synthesized by the hybrid. When the mouse cells are fused with human skin fibroblasts (which are far removed from the blood-cell line), however, the outcome is different. Messenger RNA coding for human globin is transcribed, but no human globin chains are synthesized; mouse hemoglobin is synthesized.

These results suggest that the developmental history of the cell from which the globin DNA is obtained may be important. Perhaps the early stem cell somehow modifies the globin gene so that when the gene is "turned on" at a later stage it can be transcribed into normal messenger RNA. In that case a fibroblast globin gene, even when it is put in an erythroid environment (by cell fusion) and is transcribed into globin messenger RNA, does not make a normal RNA that can be translated into protein. Perhaps there are enzymes that modify sites in the vicinity of a gene so that those sites can later be recognized by regulatory factors. The modification might be accomplished, for example, by the addition to particular bases of a methyl (CH_3) group or some other group; a correlation has been demonstrated between the degree of methylation and the activity or lack of activity of some genes, including globin genes.

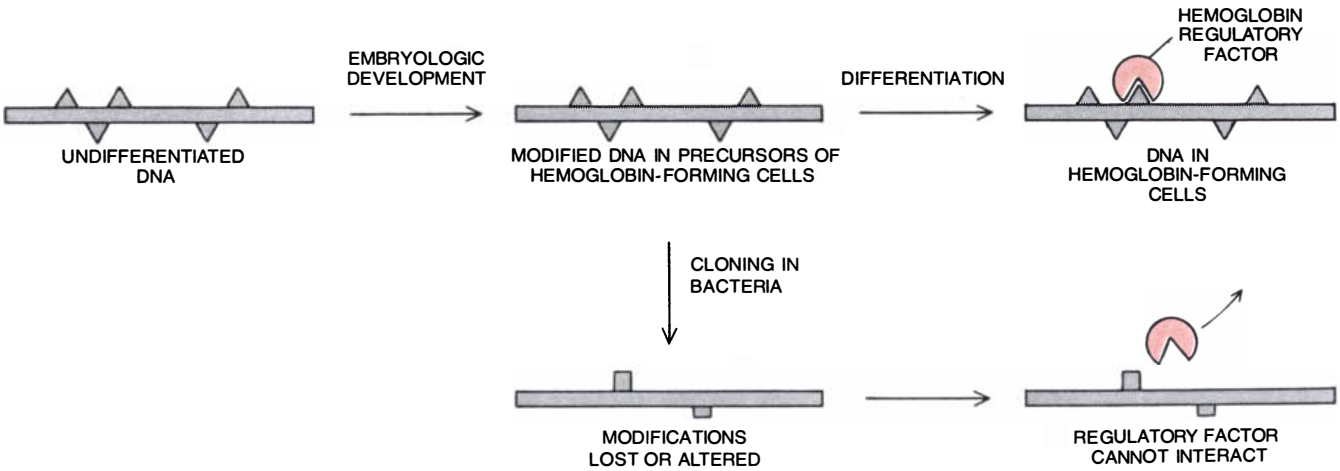
If the developmental history of a gene is important, it may account for the relative inactivity of mammalian globin genes cloned in bacteria and then transferred to mammalian cells. Perhaps the passage of a globin gene through bacteria alters or eliminates modifications that had previously been made in the gene. If that is the case, cloned genes will only be regulated normally, and so expressed normally, when the nature of those early modifications has been dis-



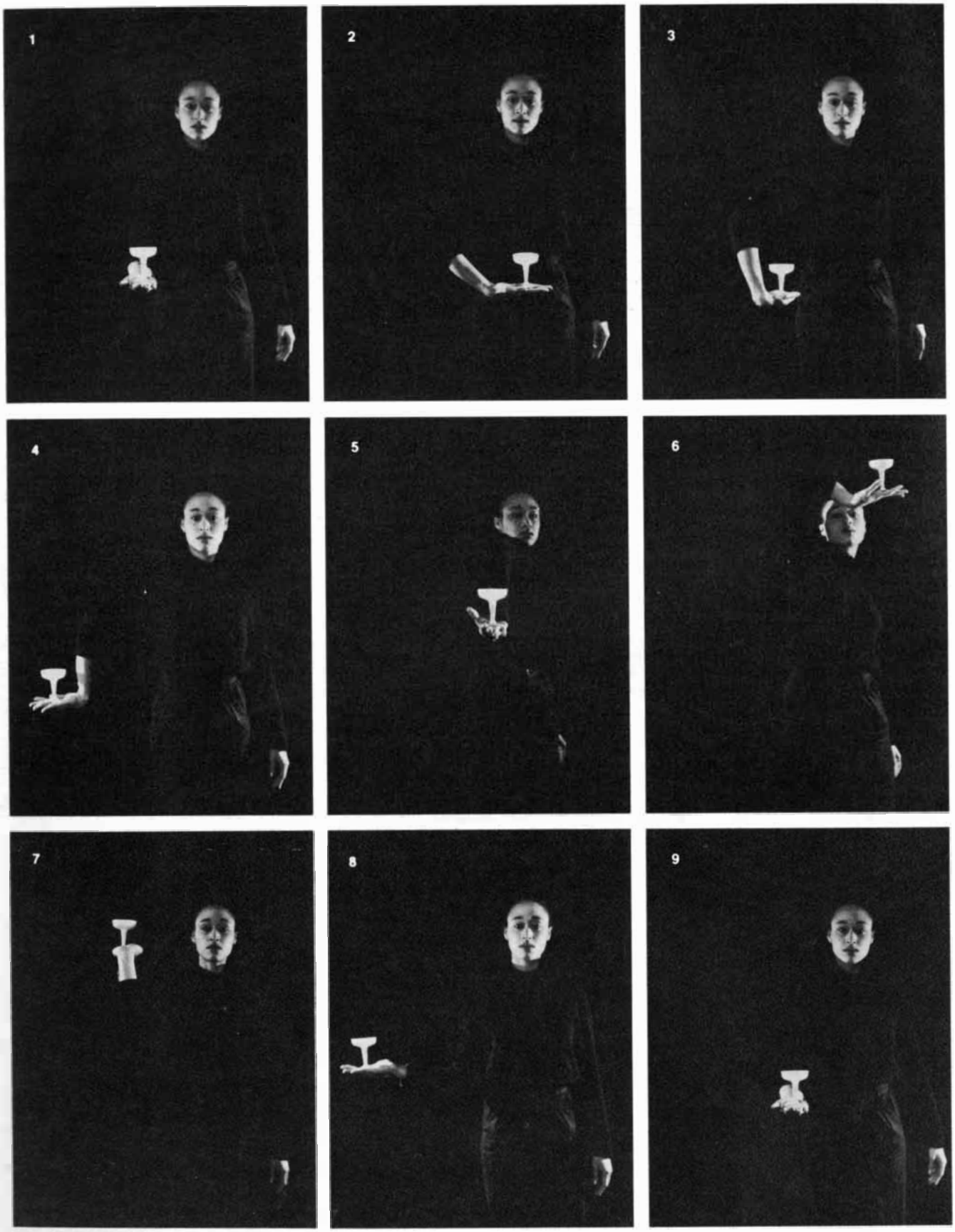
RESCUED PLASMIDS were digested with a battery of restriction enzymes and were mapped. Here three of the plasmid maps are shown (*bottom row*). Each mapped plasmid seems to have arisen from an earlier structure (*top row*), or from a concatamer, through deletion. Plasmid pPK193 can be explained as the result of the deletion of 2.8 kilobases from a pRK1 dimer (a plasmid composed of two pRK1 plasmids linked end to end); another deletion from a pRK1 trimer could have produced pPK578. Source of pPK579 (*right*) appears to have been a plasmid in which the two injected plasmids were combined (a pRK1-pX1 recombinant), from which 2.9 kilobases were deleted.

covered and the modifications can be made in the laboratory. Genetic engineering in mammalian cells still has a long way to go.

The results reported in this article demonstrate that genetic engineering is feasible in mammalian cells at least in the case of simply regulated genes such as the one for *TK*. The next step toward gene therapy for children suffering from inherited diseases is to learn how to correct defective cells not in a laboratory culture but in genetically defective experimental animals, first to replace "simple" genes and then to replace those whose regulation is complex, such as the globin genes. The recent development of lines of mice that have alpha thalassemia, an important disease in human beings, will facilitate the necessary studies. Genetic engineering should be feasible in experimental animals in the near future; ultimately gene therapy in human beings should be possible.



LACK OF EXPRESSION of a transferred globin gene might be explained in a number of ways. One possibility is that the DNA of a cell designated for a particular specialization (such as the manufacture of globin) is modified in the embryo; then, after the cell has differentiated to its specialized end stage, some regulatory factor might recognize the specific modifications and turn on the synthesis of globin (*top*). In other words, the globin gene (or its surroundings) would carry information other than that encoded in the base sequence alone. Perhaps the cloning of such a gene in bacteria abolishes or alters the embryonic modifications so that regulation is impossible (*bottom*).



FIBER BUNDLES can express an exact mathematical analogy between the movements of the Philippine folk dance Binasuan ("wine dance") and the quantum-mechanical effect of a magnetic field on the spin of a neutron. In the wine dance the orientation of the wineglass in space as it is rotated about the vertical axis cannot be identified with the relative rotation between the wineglass and the dancer's body. If the dancer's feet remain in place, a single 360-degree rotation of the hand and the wineglass introduces a twist in the dancer's body. Another complete rotation of the hand in the same direction restores the dancer's hand and body to their original configuration. The generalized phase shift of a neutron in a magnetic field exhibits the same relation to the orientation of its rotating spin vector.

Fiber Bundles and Quantum Theory

A branch of mathematics that extends the notion of curvature to topological analogues of a Möbius strip can help to explain prevailing theories of the interactions of elementary particles

by Herbert J. Bernstein and Anthony V. Phillips

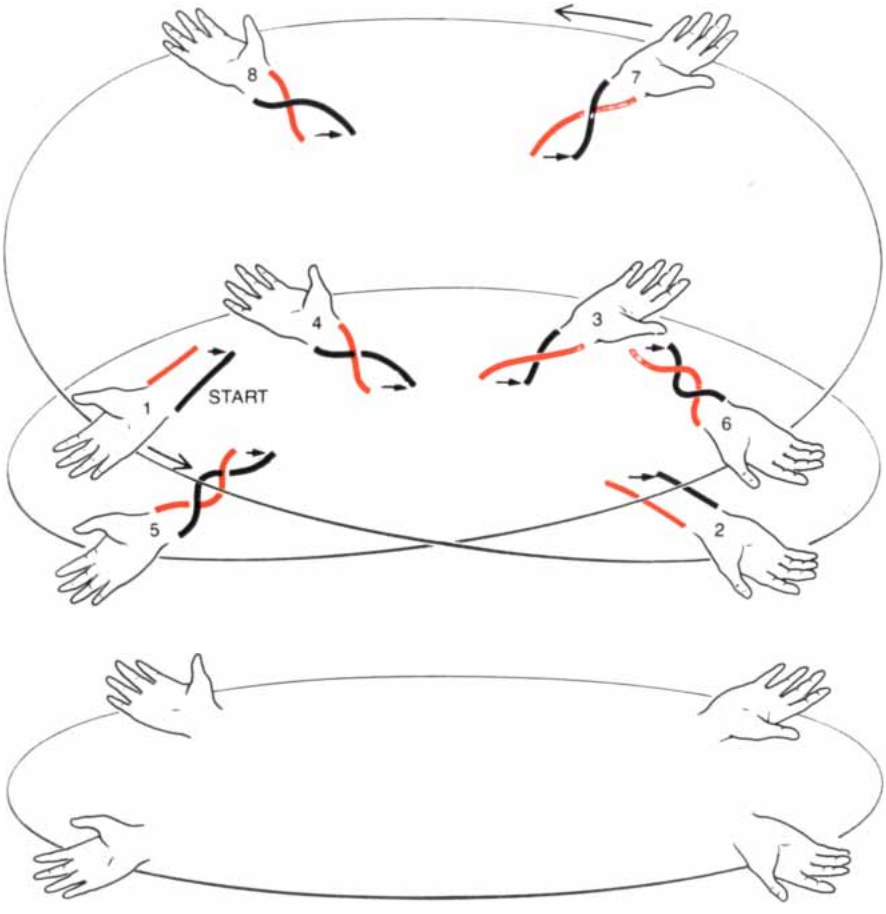
The intimate relation between mathematics and physics may seem surprising to the layman, but to someone well acquainted with either field it is a natural evolutionary development. Physical problems have stimulated mathematical thinking at least since the Egyptians introduced geometry as a means of accurately measuring land. Newton's invention of the integral calculus was in part his practical-minded response to a difficulty in formulating the law of universal gravitation. Infinite trigonometric series were devised to study the flow of heat. The abstract patterns investigated by present-day mathematicians are still based, albeit sometimes remotely, on the real patterns exhibited by the physicist's universe.

Mathematics has not failed to repay the debt in kind. A theory invented by mathematicians to settle mathematical questions often turns out to be exactly what physicists need to advance their analyses and predictions of the ways of nature. Tensor calculus, the product of almost 100 years of development by such mathematicians as Karl Friedrich Gauss, Bernhard Riemann and Tullio Levi-Civita, was essential to Einstein's formulation of his general theory of relativity. The most recent beneficiaries of such mathematical research are the physicists who study the forces and quantum-mechanical fields that mediate the interactions of elementary particles.

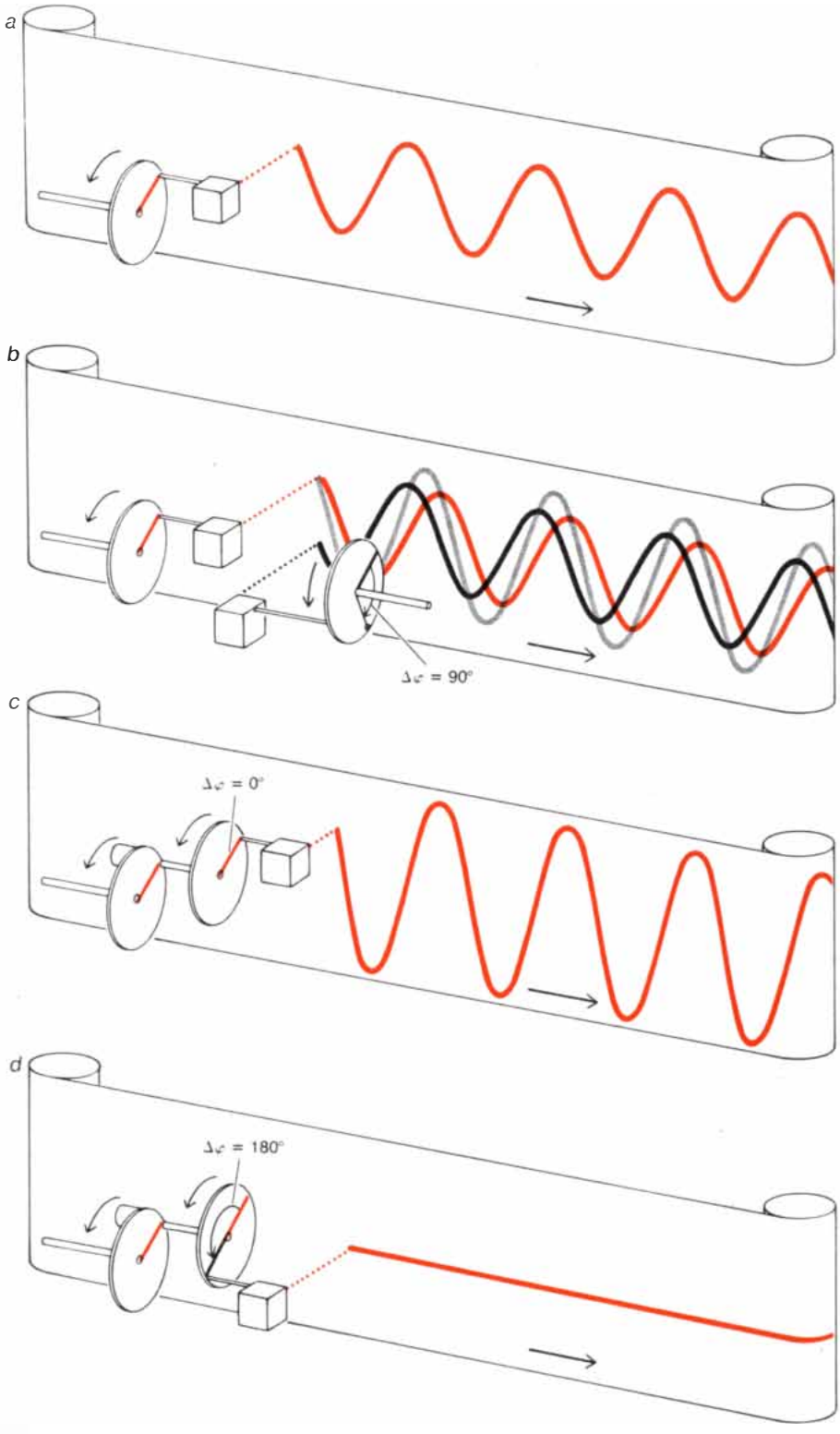
The fields that are most promising for this purpose are called gauge fields. Their utility lies chiefly in their ability to express underlying relations among forces that appear superficially to be quite distinct. Gauge fields have played a major role in recent attempts to devise a unified theory of three of the four basic forces known in nature, namely the strong, the weak and the electromagnetic forces. For the fourth basic force, gravity, there is as yet no quantum-mechanical theory, but the general theory of relativity suggests that gravitation too may ultimately be described by a gauge-field theory.

The idea of a gauge field was introduced by Hermann Weyl in the 1920's, but the present development of gauge-field theory began in 1954. In that year C. N. Yang and Robert L. Mills, who were then working at the Brookhaven National Laboratory, applied the gauge-field concept to nuclear forces. After almost 20 years of further refinement physicists were able to express the concept of a gauge field in such a way that it could be recognized as an instance of more abstract structures known to mathematicians as connections in fiber bundles. The discovery of this equivalence has made it possible to apply mature and exceedingly powerful mathematical concepts to the description of physical reality.

What is a gauge field? What are fiber bundles and how do they enter into physics? What does it mean for a fiber bundle to have a connection, and how



DOUBLE-COVERING FIBER BUNDLE represents the set of relative rotations between the dancer's hand and the rest of her body (*see illustration on opposite page*). Arrows represent the fixed orientation of the dancer's feet. The circle at the bottom represents the orientations of the hand. Motion along the circle induces motion along one of the segments of the twisted curve directly above the circle. The topology of the latter curve shows that a counterclockwise relative rotation of 90 degrees is equivalent to a clockwise relative rotation of 630 degrees.



PHASE OF A WAVE, which is usually expressed as an angle, can be detected only as a difference between the phases of two waves. For sine and cosine waves the waveform can be traced by projecting a point on a uniformly rotating circle onto a screen moving uniformly at right angles to the line of projection. Any position on the circle can be chosen as zero degrees. The angle of rotation away from the arbitrarily chosen position then labels the phase (*a*). The relative phase between two waves is well defined in that each crest and trough on one wave is the same number of degrees ahead of the corresponding feature of a second wave. When waves interfere, their amplitudes add at each instant; the resulting maximum height depends on the relative phase (*b*). The addition can be carried out automatically by centering the second circular motion on the perimeter of the first. If the interference is constructive (*c*), crests coincide with crests and troughs with troughs, and the maximum height of the resulting pattern is equal to the sum of the heights of the original waves. If the interference is destructive, corresponding to a phase difference of 180 degrees (*d*), crests coincide with troughs and waves cancel exactly.

are the concepts of a connection and of a gauge field related? We shall attempt to answer these questions by analyzing two physical experiments. One experiment shows what happens when neutrons are rotated 360 degrees by a magnetic field and the other shows the effect on two partial electron beams of a magnetic field in the region between them. Each experiment demonstrates in a different way how a fiber bundle can arise in quantum theory. In the neutron-rotation experiment the global structure of the fiber bundle is significant, whereas in the electron-beam experiment the central feature is a connection, an intrinsic local structure that can be imposed on the bundle. The connection gives an elementary but fundamental example of a gauge field.

The study of fiber bundles is part of the branch of mathematics called topology, but bundles have also been investigated in differential geometry because of their relation to the geometric concept of curvature. The idea of a connection in a fiber bundle grew out of attempts to generalize the notion of the curvature of a two-dimensional surface, such as the surface of the earth, to the curvature of a space with three or more dimensions. Hence another way of expressing the mathematical difference between the two experiments we shall describe is to note that the neutron-rotation experiment concerns the topology of a fiber bundle whereas the electron-beam experiment concerns the geometry of a fiber bundle.

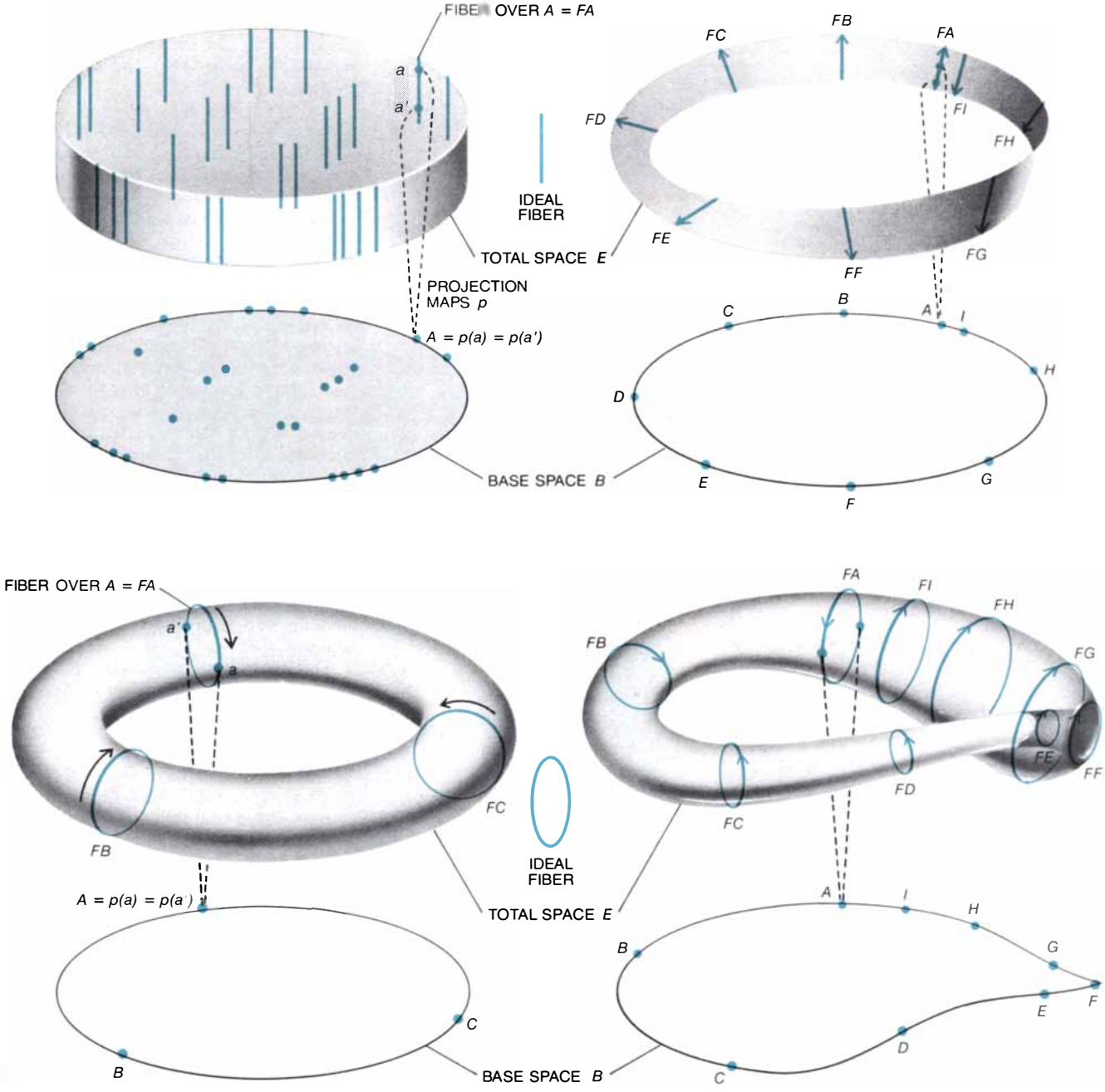
The 360-degree neutron-rotation experiment was proposed by one of us (Bernstein) in 1967. A similar thought experiment was described almost simultaneously by Yakir Aharonov and Leonard Susskind of Yeshiva University. The experiment demonstrates a highly counterintuitive effect whose mathematical equivalent is the one-sidedness of a Möbius strip. At issue is the spin, or intrinsic angular momentum, of a subatomic particle. According to quantum theory, a neutron or another particle with spin does not return to its initial state when its orientation is rotated through 360 degrees. Instead it takes two full turns, a 720-degree rotation, to restore the state of the particle to its initial condition.

To understand the experiment it is necessary to know something of the quantum theory of spin. Physicists have extrapolated the notion of intrinsic angular momentum from the spin of a top or a gyroscope to the realm of elementary particles. In each case spin is a vector quantity, which means it has both a magnitude and a direction. By convention the vector points along the spin axis in a direction determined by what is called the right-hand rule. If the fingers of the right hand are curled as if to grasp

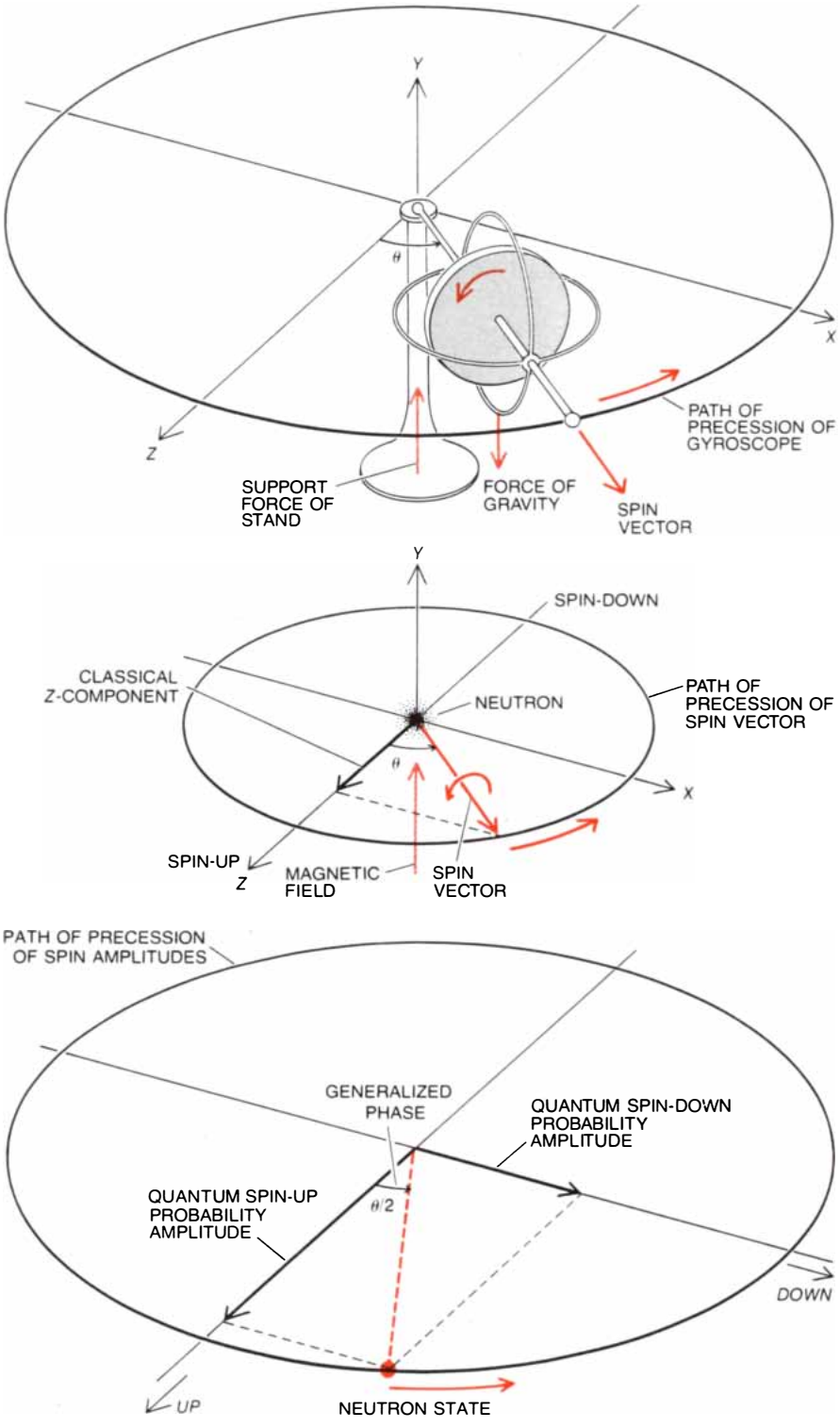
the axis of a spinning object with the fingers wrapped around the axis in the same sense as the spin, the direction of the thumb along the spin axis gives the direction of the spin vector.

Unlike the spin of a top or a gyroscope, the spin of an elementary particle is quantized: its magnitude can have only certain discrete values, which are multiples of the smallest possible quantity of spin. Moreover, for any given particle the magnitude of the spin can never change; it is one of the intrinsic properties that determine the identity of the particle. The electron, the proton and the neutron (and a number of other particles) have the smallest allowed nonzero quantity of spin. This minimum spin magnitude is equal to $\hbar/2$, where $\hbar$ is a form of Planck's constant with a value of about 10^{-27} erg-second.

The constraint on the magnitude of the spin vector in quantum theory reflects a still more curious experimental finding. Like all vectors, the spin vector has components along the axes of any chosen coordinate system. An experimenter must select a particular axis along which to measure a spin component. No matter what direction is chosen as a reference axis, however, the only values ever found are $+\hbar/2$ and $-\hbar/2$. No intermediate values are observed. In



FIBER BUNDLE consists of a base space, a total space and a map that projects each point in the total space onto a point in the base space. The set of all the points in the total space that are mapped onto the same point in the base is called a fiber (*colored lines and circles*). The total space can resemble a sheaf or bundle of fibers. Every fiber in a fiber bundle must have the same topological structure, and so all the fibers can be represented by a single ideal fiber. For the solid cylinder (*upper left*) or the Möbius strip (*upper right*) the fibers are straight lines above the points in the base, whereas for the torus (*lower left*) or the Klein bottle (*lower right*) the fibers corresponding to each point are circles. Motion in the base may induce a change in the orientation of the fibers, depending on the topological structure of the total space. A single circuit along a path in the base changes the orientation of the fibers of the Möbius strip and the Klein bottle, but the orientation of the fibers of the torus and the cylinder is not altered by motion along any such path. In moving from fiber to fiber on the Klein bottle, the bottle appears to pass through itself. The self-intersection cannot be avoided in a three-dimensional drawing.



PRECESSION OF THE SPIN VECTOR of a neutron in a magnetic field resembles the precession of a gyroscope in a gravitational field. The magnetic torque on the spinning neutron causes it to precess at a rate proportional to the strength of the field and independent of the orientation of the neutron. If the initial direction of the neutron's spin vector is called "up," and if the field is perpendicular to it, precession through an angle of 180 degrees will make the spin vector point down. At intermediate angles classical physics predicts that the component of the spin vector measured along the z axis is equal to the perpendicular projection of the vector onto the axis. In quantum mechanics the component of the spin measured along any axis can have one of only two values, $+1/2$ or $-1/2$ times Planck's constant $\hbar$. What changes during precession is the probability of detecting a neutron in the spin-up state ($+\hbar/2$) or the spin-down state ($-\hbar/2$). Each probability is determined by squaring a probability amplitude whose value can be either positive or negative. Hence the quantum-mechanical precession of a neutron can be graphed as a rotation of the neutron state in an abstract space whose coordinate axes are the probability amplitudes of the spin-up and spin-down states of the spin vector.

spite of this nonintuitive property of particle spins the picture of a spin vector remains useful for describing neutrons. It is possible to spin-polarize a series of neutrons so that ideally all the spin vectors point in the same direction. The experimenter can determine the direction by turning the axis of measurement to maximize the probability of detecting the value $+\hbar/2$. If all the neutrons are in the same state, all of them will give this result. For convenience the reference axis along which the spins are polarized can be labeled the z axis and the two possible states of the spin can be designated spin up ($+\hbar/2$) and spin down ($-\hbar/2$).

Suppose the spin vector of each neutron in the series rotates away from the fixed reference axis. The component along the z axis cannot change continuously, since measurements yield only the two discrete values $+\hbar/2$ and $-\hbar/2$. Instead the probabilities of finding spin-up and spin-down neutrons begin to change. In the initial state of the spin vector the probability of finding a spin-up neutron is 1 and that of finding a spin-down neutron is 0. After the spin vector has rotated a quarter of a turn, or 90 degrees, from the z axis, the classical (non-quantum-mechanical) model predicts that the z-axis component will vanish. The spin vector would then be oriented so that it points neither up nor down along the z axis. For neutrons, however, a spin of zero is never found. According to the quantum-mechanical model, a 90-degree rotation changes the state of the neutron to one in which both the spin-up and the spin-down probabilities are .5.

This outcome reconciles the quantization of spin with the classical description. If the z-axis components of a series of neutrons are measured while their spin vectors are oriented perpendicular to the z axis, half have spin $+\hbar/2$ and half have spin $-\hbar/2$. Thus the sum of all the measured components is zero and so is the average value, in agreement with the classical result.

In a similar way a rotation of 180 degrees orients the neutrons with their spin vectors down. This makes the spin-up probability 0 and the spin-down probability 1. After the neutron has swept out a full circle the spin-up probability (the probability of finding the z component of the spin equal to $+\hbar/2$) returns to 1 and the spin-down probability again becomes 0.

Physicists have taken the distinctness of spin states and the need for probabilities as being basic postulates of quantum theory. Probabilities are calculated from the mathematical description of each state of a particle as a wave function. At any point in space there is a number called a probability amplitude

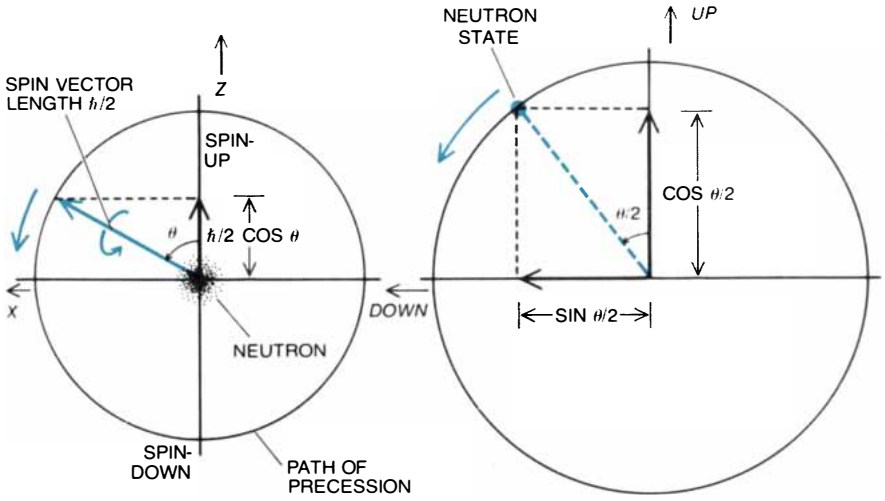
of the neutron wave function for each distinct state. The term amplitude refers to the waves with which quantum theory describes material particles. A probability amplitude can be positive or negative, which reflects the observed capacity of waves to add constructively or destructively. The probability of finding a particle in a given state is the square of the corresponding probability amplitude. Squaring the amplitude ensures that the probability itself is always a positive number.

Since the neutron has two distinct spin states, the complete description of a neutron at a point in space consists of two numbers, namely the probability amplitude for spin up and the probability amplitude for spin down. One way to show the amplitudes that respects the distinctness of the two states is to plot them on perpendicular axes marked "up" and "down" in an abstract space called neutron-state space [*see illustration at right*]. The state of the neutron can be represented by a point on this graph. If the initial state is spin up, the spin-up probability amplitude is +1 and the spin-down probability amplitude is 0. This combination of values corresponds to a point one unit from the origin along the "up" axis.

As the orientation of the neutron spin is changed the two probability amplitudes each vary continuously. The sum of the two probabilities must always add to 1, however, because spin up and spin down are the only possible states. Hence the sum of the squares of the spin-up and spin-down probability amplitudes must be 1 and the point representing the state of the neutron must lie on a circle of radius 1. Every possible state of the neutron can then be labeled by the angle from the "up" axis to the corresponding point. The angle is called the generalized phase of the neutron state.

After a 90-degree rotation of the physical spin vector away from the z axis, the spin-up and spin-down probability amplitudes must be equal in absolute value since the corresponding probabilities must each be .5. The point in the neutron-state space therefore lies halfway between the "up" and the "down" axes. It follows that the generalized phase of the neutron state has changed by 45 degrees as a result of a 90-degree rotation of the neutron spin vector.

The half-angle relation continues. When the physical spin vector has rotated 180 degrees, it points down: the spin-down probability amplitude is +1 and the spin-up probability amplitude is 0. The corresponding point in the neutron-state space is 90 degrees from the initial direction. After a 360-degree physical rotation both probability amplitudes change their sign. It takes two full turns, or 720 degrees of rotation, to restore the probability amplitudes to their initial values.



SPIN VECTOR OF A NEUTRON can precess, but the geometry of the precession cannot be pictured consistently by the classical model (*left*), since a spin measurement along any given axis yields only the values $+\hbar/2$ or $-\hbar/2$. In quantum mechanics precession is manifested as a change in the probability of finding a neutron with spin $+\hbar/2$ (spin up) or with spin $-\hbar/2$ (spin down). The two amplitudes that determine the probability can be considered coordinates in an abstract space with axes labeled "up" and "down" (*right*). After a precession of 90 degrees from the z axis the spin vector points neither up nor down. If one measures the z-axis component of the vector, one finds spin up half of the time and spin down half of the time. The average value of the spin is therefore zero, in agreement with the classical result. Because the probabilities are equal the probability amplitudes can be chosen to be equal; the corresponding point in neutron-state space is rotated 45 degrees from the "up" axis. Physical precession through any angle θ causes a generalized phase shift $\theta/2$, represented as a rotation in neutron-state space.

This feature of quantum theory may at first seem paradoxical. When an ordinary object makes a complete rotation in space, it returns to the same state from which it started. A person's body or a spinning gyroscope is unchanged by a 360-degree rotation about any axis. This fact is so deeply ingrained by common experience that although the theory of neutron spin is now some 50 years old, before 1967 even most physicists thought a 360-degree rotation could not have directly measurable consequences. Quantum theory implies that one cannot measure a probability amplitude in any direct way, and the change in the sign of the amplitude caused by such a rotation disappears when the amplitude is squared to compute the probability.

On the other hand, there are some circumstances in the realm of macroscopic objects in which a 360-degree rotation has an observable effect. For example, if two objects are attached by a flexible ribbon, it is obvious that a full turn of one object does not restore the system to its original state: the ribbon ends up with a twist in it. What is not so obvious is that a second full turn in the same direction can bring such a system back to its initial state: the ribbon can be untwisted even though the relative rotation of the two objects undergoes no further change. The effect can also be demonstrated by holding a wineglass in the palm of the hand and rotating the glass about its vertical axis (without moving the body as a whole). After a 360-degree rotation the glass returns to its original orientation, but the arm is twisted; a further 360 degrees restores the glass and the arm to their initial position [*see illustration on page 122*].

All these phenomena, both macroscopic and quantum-mechanical, can be represented by the properties of a single fiber bundle. A fiber bundle is a mathematical structure that consists of two distinct sets of points, called the base space B and the total space E, and a rule p called the projection map that associates a point in B with every point in E. In the fiber-bundle model of the rotation of the wineglass the points in the base space represent the possible orientations of the glass and the hand. The points in the total space represent the rotation the hand has undergone with respect to the rest of the body. The projection map defines an association between each rotation and the relative orientation determined by the rotation.

In the fiber-bundle model of neutron spin rotation the points in the base space represent the orientation of the spin vector. In the part of the effect we are describing in detail the vector rotates in a fixed plane, and so any orientation can be described by the angle it makes with the z axis. (Other possibilities can be treated analogously.) The points in the total space represent generalized phases of the neutron state, and so they

correspond to points on the unit circle in the neutron-state space with "up" and "down" coordinates. Each of these points can thus be described by its angular distance from the "up" axis. The projection map in the model assigns to each point in the total space a point in the base space according to the rule $p(\phi) = 2\phi$(modulo 360). The application of the rule is equivalent to wrapping the circle of generalized phases twice around the circle of orientations, so that two phases ϕ and $\phi + 180$ degrees are possible for each orientation.

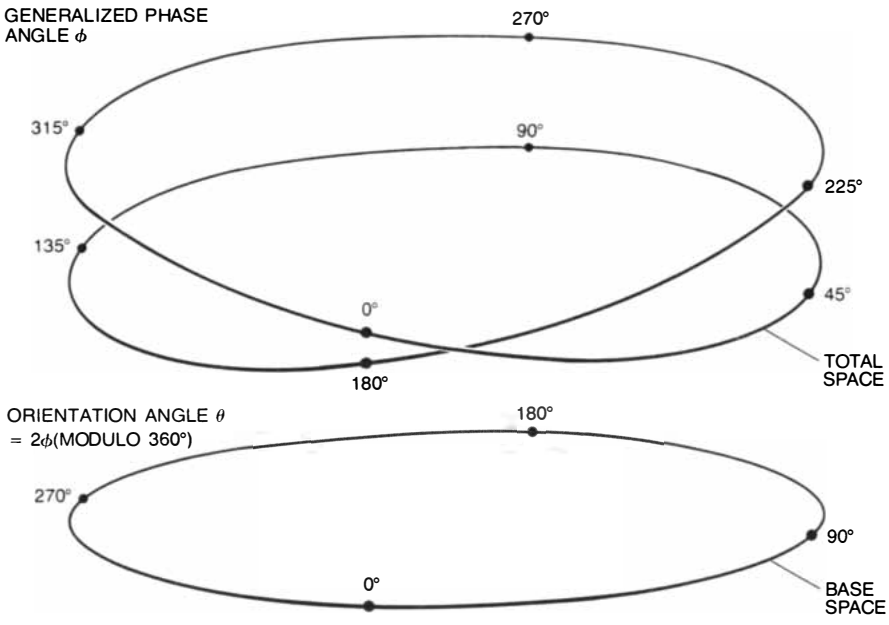
The correspondence between points in the total space and points in the base is generally expressed by regarding the total space as being "over" the base. In this representation the point or points in the total space that the projection map associates with a point in the base lie vertically above the base point. The set of points in the total space over a base point is called a fiber. Another part of the definition of a fiber bundle requires that the fibers over any two points be topologically equivalent, so that the topological structure of the fiber does not vary from one point in the base to another. In some instances the fiber over each point in the base is a line, and it is this appearance that has given rise to the name fiber bundle; when each fiber is a line, the total space looks like a bundle of fibers. In general, because the fibers are all topologically equivalent they can all be described as copies of a single fiber, F, the ideal fiber of the fiber bundle.

The ideal fiber of the bundle representing neutron spin rotation is a space that consists of two distinct points. For example, over the point labeled 0 degrees in the base there are the two points in the total space that correspond to the generalized phases 0 degrees and 180 degrees. Hence the fiber over 0 degrees is the set that consists of 0 degrees and 180 degrees in the total space. Similarly, the fiber over 90 degrees is the set that consists of 45 degrees and 225 degrees in the total space. In this bundle both the total space and the base space are topologically equivalent to a circle. The projection map corresponds to the way the edge of a Möbius strip would project onto a circle at the center of the strip.

How does the fiber-bundle model of neutron spin rotation represent the relation between a rotation and a generalized phase shift? Suppose a neutron starts with its spin vector pointing along the positive z axis ($\theta = 0$), so that its spin-up probability amplitude is $+1$ and its spin-down probability amplitude is 0. (This describes the point $\phi = 0$ on the unit circle in the abstract space of neutron spin states.) Now, if the spin vector is physically rotated 90 degrees from the z axis, ending up at $\theta = 90$ degrees, our preceding discussion shows that the point on the generalized phase circle moves to $\phi = 45$ degrees. If the spin vector rotates another 90 degrees to $\theta = 180$ degrees, the point on the phase circle moves to $\phi = 90$ degrees. The correspondence between rotations and generalized phase shifts can be described by saying that the point on the phase circle moves continuously in such a way that it always remains above the point on the orientation circle. This geometric principle, together with the topological structure of the bundle, accounts for the sign change of the neutron state as an effect of a phase reversal. One complete rotation in the base must shift the generalized phase to the opposite of what it was [*see illustration on this page*].

How might one rotate the spin vector of a neutron 360 degrees? Current experimental designs take advantage of the magnetic properties of electrically neutral particles. A neutron has not only spin angular momentum but also a magnetic moment, which makes it resemble a bar magnet spinning about its north-south axis. Suppose the spin vector of the neutron is initially aligned with the z axis and a magnetic field is introduced at right angles to that axis. The torque that aligns a bar magnet with an external field makes a spinning magnet precess about the direction of the field. The spin vector of the neutron will precess in the plane at right angles to the magnetic field just as a spinning gyroscope precesses in response to the pull of gravity. Hence to rotate the spin vector of a neutron away from the z axis one can take advantage of its magnetic moment. Actually even if the magnetic field is not perpendicular to the initial spin, the neutron precesses at a rate that is proportional to the strength of the magnetic field and that does not depend on the orientation of the neutron. Thus all the neutrons in an unpolarized beam passing through a magnetic field precess at the same rate; the rate is called the Larmor frequency.

Having a way to rotate neutrons 360 degrees is not enough, however. One must be able to compare the probability amplitudes for a rotated neutron with the amplitudes for the original state. The amplitudes for both the rotated and the unrotated states have the same magnitude but opposite signs. The difference in sign can be detected because in quantum mechanics it is possible to make a particle arrive at the same point in space along two different paths, in the sense that there is a nonzero probability of detecting the particle along either one of the paths. As always, the probabilities are given by the squares of the probability amplitudes at any point along the paths. At a point where two paths contribute to the probability of detecting the particle the probability amplitude is the sum of the amplitudes for each of the paths. The addition is done before the amplitude is squared to get the probability. This rule, which embodies the phenomenon of quantum interference, provides a method for demonstrating the sign change associated with a 360-



FIBER BUNDLE of phase shifts shows the relation between the angular precession of a neutron and the shift in the generalized phase of the neutron spin state. Points in the base space of the bundle represent the orientation of the spin vector of a neutron. Points in the total space represent the relative phase shifts in neutron-state space that correspond to a given orientation. For instance, the projection map of the bundle assigns the points 45 degrees and 225 degrees in the total space to the point 90 degrees in the base. This means that generalized phase angles of 45 degrees and 225 degrees both correspond to an orientation of the spin vector 90 degrees from the z axis. The topology of the total space, however, shows that attaining a phase shift of 225 degrees requires a precession of 450 degrees in the base, a one-and-a-quarter turn.

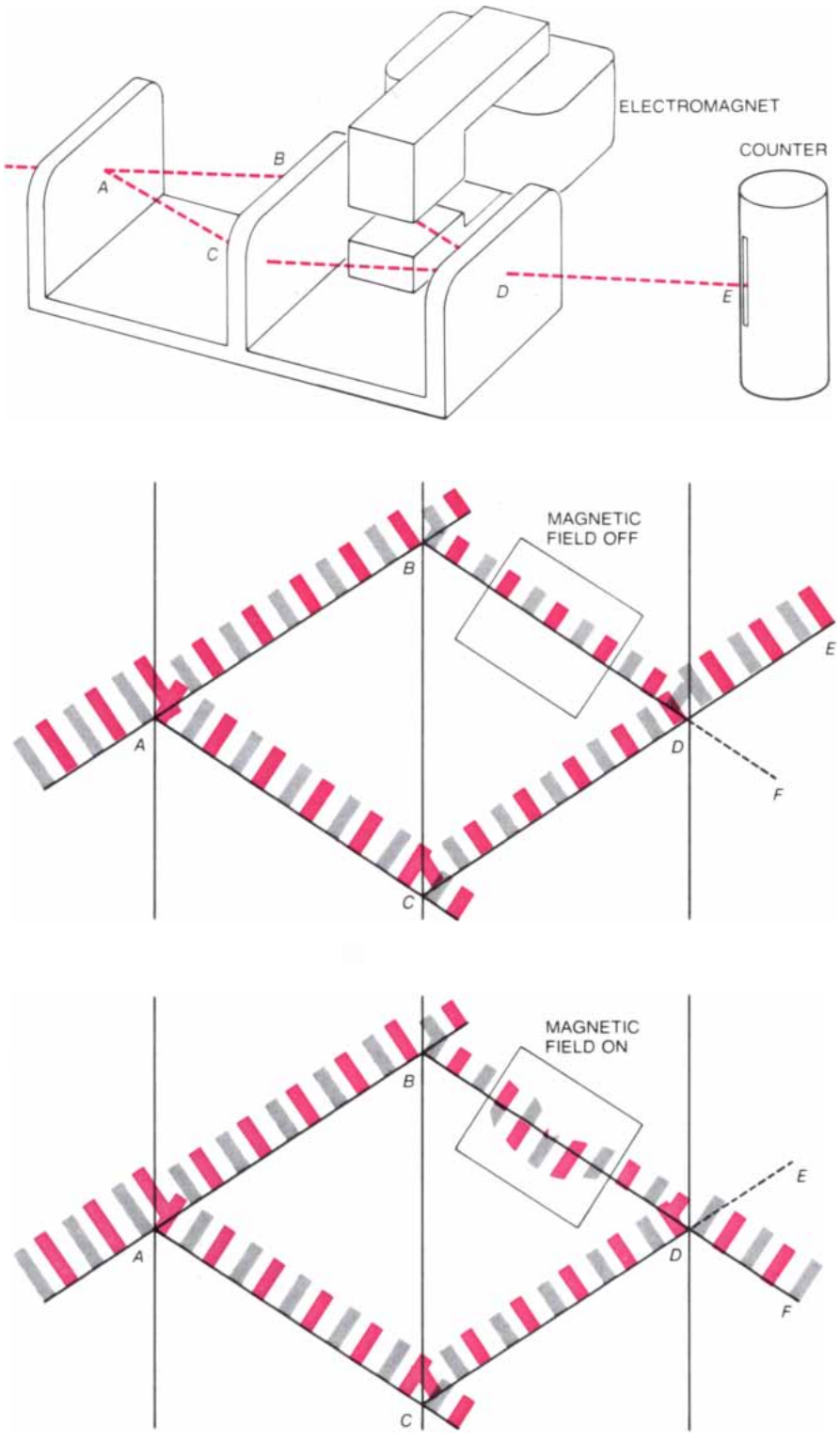
degree rotation. The recent development of a neutron interferometer has made it possible to split a beam of neutrons so that the particles follow two paths and to recombine the partial beams. Moreover, it is possible to rotate the spin vectors of the neutrons in one beam but not in the other. When the relative rotation is 360 degrees, the resulting sign change manifests itself through destructive interference of the two partial beams.

Experiments to detect 360-degree neutron rotations have been done by several groups. Helmut Rauch, Ulrich Bonse and their colleagues first demonstrated the effect in 1975 at the Institut Laue-Langevin in Grenoble. An American team headed by Samuel A. Werner of the University of Missouri carried out a similar demonstration at about the same time. In 1976 Anthony Klein and G. I. Opat of the University of Melbourne employed a novel Fresnel diffraction technique to show the effect of neutron rotation, again at the Institut Laue-Langevin.

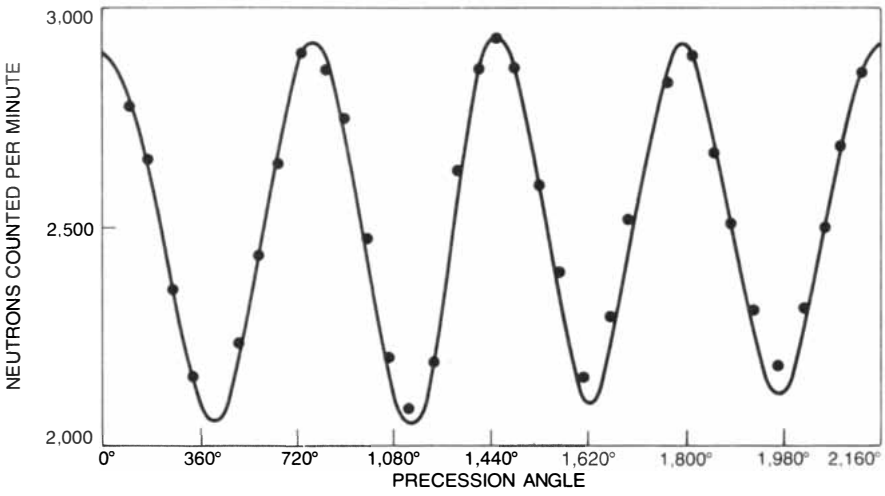
The heart of a typical neutron interferometer is a perfect cylindrical crystal of silicon, cut away so that three polished "ears," or projections, stand up from a single intact base. When a beam of neutrons strikes the first ear at a suitable angle, it divides into a transmitted partial beam and a diffracted partial beam. At the second ear each of these beams divides again. At the third ear two partial beams recombine. The two beams interfere constructively or destructively depending on their generalized phase. The recombined beam then splits again, and counters, or detectors, placed beyond the third ear record the number of neutrons in each of the two partial beams.

The probability of a neutron's arriving at a counter in the path of the recombined beam changes as the generalized phase of the rotated partial beam changes. The probability is measured by counting the number of neutrons arriving per second. If the two partial beams are exactly in phase when there is no rotation, the initial counting rate is high. This situation results from constructive interference: the two amplitudes have the same sign, and the square of their sum is at a maximum. Indeed, whenever the generalized phase difference is zero or a multiple of 360 degrees, there is constructive interference. After one full rotation of the spin vector the phase difference is 180 degrees and the amplitudes have opposite signs. Ideally the sum would be zero; in practice the counting rate is at a minimum. This is destructive interference.

Rotation of the neutron spin vectors is accomplished by placing an electromagnet in the region between the second and the third ear. One of the partial beams passes through the field of the magnet



NEUTRON INTERFEROMETER, cut from a single perfect crystal of silicon, has three projecting "ears." Each ear divides a beam of neutrons into a diffracted partial beam and a transmitted one. In this experiment one partial beam passes through a magnetic field, where the spin vectors of the neutrons are rotated and the generalized phase of the neutron state is shifted. When the two partial beams are recombined, they interfere according to the phase difference between them. If the magnetic field is zero, there is no phase difference and the beams interfere constructively. If the magnetic field rotates one of the beams 360 degrees, the phase difference between the two paths is 180 degrees and the beams interfere destructively at counter *E*. The crests of the neutron waves are represented by colored stripes and the troughs by gray stripes. The probability amplitudes of the spin states are indicated by the relative lengths of the stripes. Because the effect is illustrated for an incident spin-up neutron, a nonzero spin-down probability amplitude is shown below the black line only in the region of the magnetic field. Negative probability amplitudes are shown by converting crests into troughs and vice versa. The scattering of the neutrons at each ear is represented as a 90-degree phase shift (skip one stripe) each time a neutron beam is diffracted and as no phase shift when it is transmitted.



SEVERAL CYCLES OF INTERFERENCE can be generated even with a moderate magnetic field. The total precession angle θ of the rotated beam can be calculated from the strength of the magnetic field. When the intensity of the recombined beam is plotted against θ, the experimental curve closely resembles the graph of the cosine of $\theta/2$. Background neutrons raise the minimum count above zero. Eight full rotations of the beam give only four intensity peaks.

but the other beam does not. Hence the field rotates spin along one partial beam, leaving the other beam unchanged. The angle of rotation is proportional to the strength of the magnetic field. As a result the generalized phase of the first partial beam increases continuously as the experimenter increases the current in the electromagnet from zero to its maximum value. As the phase shift increases, the counting rate at one counter at first declines; the interference changes smoothly from constructive to destructive. After reaching a minimum count the interference tends back toward the maximum as the generalized phase approaches one full cycle. The resulting cycle of variation in the counting rate repeats as long as the current continues to rise. Since the rotation angle does not depend on initial spin orientation, the experiment does not require a polarized neutron beam.

The total rotation angle induced along the path through the magnetic field is equal to the Larmor precession frequency multiplied by the time the neutrons spend in the field. The angle can therefore be calculated from measurements of the velocity of the beam, the intensity of the field and the distance across the field. In the version of the experiment done by Rauch, Bonse and their colleagues the neutrons travel through a magnetic field 1.5 centimeters wide at a speed of 2,170 meters per second, so that each neutron spends a little less than seven microseconds in the field. When the electromagnet is operating at maximum current, the strength of the field is 400 gauss, which corresponds to a Larmor frequency of 433 million degrees per second. At this rate, in seven microseconds the spin vector of each neutron rotates about eight full turns. If each 360-degree rotation of the spin vector restored a neutron to its original state, one would expect to observe eight cycles of maximum and minimum counts. The actual result is significantly different. As the magnetic field increases from zero to its maximum the number of neutrons detected at the counter passes through only four cycles.

The outcome of the neutron-rotation experiment shows in a sense that fiber bundles exist in quantum mechanics and can be observed. The fiber bundle associated with neutron rotation, however, is an extremely simple one, because both its base space and its total space are one-dimensional. (To repeat, the base space and the total space are both circles with the total space wrapped around twice like the edge of a Möbius strip.) The correspondence between rotations and phase shifts is embodied in the rule that the point in the total space should move continuously so that it always lies above the position of the point in the base. Because there is only one degree of freedom in the total space the rule unambiguously specifies a phase shift for each rotation.

In more general fiber bundles things are not so simple. For example, if each fiber is a line rather than a pair of points, motion in the base implies motion from fiber to fiber, but it does not specify which point in each fiber is to be traversed. To determine an unambiguous path through the total space in such a fiber bundle, additional structure is needed. A procedure that gives a path through the total space lying directly above a path in the base, once a starting point in the total space has been specified, is called a path-lifting rule.

The study of fiber bundles grew out of attempts to make more accessible to analysis the complexities of curvature in a manifold: an abstract topological space with an arbitrary number of dimensions. The idea of a fiber bundle was implicit in the work of the French mathematician Élie Joseph Cartan, but it was first explicitly stated in about 1935 by Hassler Whitney, who now works at the Institute for Advanced Study. The concept of path lifting was developed as a systematic way of comparing the effects of curvature at different points of a manifold and was extended to fiber bundles by the French mathematicians Charles Ehresmann and Henri Cartan and by others in about 1950.

Consider what is perhaps the simplest curved manifold: the two-dimensional surface of a sphere. An observer standing on the surface has a complete circle of directions in which to look along the surface. (Assume that the observer looks only along the horizon, never up or down.) There is such a circle of directions for each point on the surface. The circles taken together form a fiber bundle in a natural way, called the bundle of directions on the surface of the sphere. The base space of the bundle is the surface itself. The fiber over a point in the base represents the set of all the directions on the surface of the sphere that can be surveyed from that point. Hence each fiber is a circle.

It is not possible to draw a faithful picture that encompasses all the total space of the bundle of directions on the sphere. One reason such a drawing cannot be made is that it is topologically impossible to assign a reference direction to every point on the surface of a sphere in a continuous way. This fact is expressed whimsically by the statement, "You can't comb the hair on a sphere." At least one point on the sphere will always have a "cowlick." On the other hand, it is possible to comb any patch of hair on the sphere, even if the patch covers the entire surface with the exception of one point. For such a patch it is possible to assign a continuous set of directions and so to draw a topologically accurate picture of the total space of the bundle of directions over the patch. For example, on a flat map of the northern hemisphere one can draw a continuous set of directions, say all pointing down and to the left. The set of directions can then be transferred back to the northern hemisphere of the sphere in order to comb its hair.

Combing the hair on the hemisphere gives a reference direction at each point. The total space of the bundle of directions on the hemisphere can then be illustrated by drawing each fiber vertically. If the bottom of each fiber corre-

sponds to the reference direction, the angle any other direction makes with the reference direction can be represented by a height along the fiber [*see illustration on next page*]. The point halfway up the fiber represents 180 degrees from the reference direction and the point at the top of the fiber represents 360 degrees from the reference direction. Since the 360-degree direction coincides with the reference direction, the top and the bottom of each fiber represent the same point of the total space.

How can one define a path-lifting rule for the bundle of directions on the sphere? A path in the base space of the bundle of directions is simply a path on the surface of the sphere. Lifting such a path into the total space requires that a direction chosen from the circle of directions be assigned to every point in the path. Think of a watch with one hand, transported so that the center of the watch moves along the path on the sphere while the hand moves freely around the dial. A path-lifting rule must determine the position of the hand at each point in the path, once the hand's initial position is given. If one considers only the topology of the sphere, so that the surface can be stretched and deformed (but not torn) as if it were made of a sheet of rubber, there is no preferred way to give such a rule. The fundamental reason is that there is no topological relation between directions at one point on the surface of the sphere and directions at another point. If one considers the geometry of the sphere, however, there is a natural principle that will determine the movement of the watch hand. The principle is called parallel transport.

Parallel transport on a sphere can best be understood by imagining the sphere to be rolling on a flat surface. Suppose there are straight lines and curved lines drawn on the flat surface in wet ink, and suppose there are arrows spaced frequently along the lines, all pointing, say, to the lower left. The term parallel transport is based on this picture. Since the arrows in the plane are all parallel, the arrows printed onto the sphere exemplify parallel transport along the printed curve on the sphere.

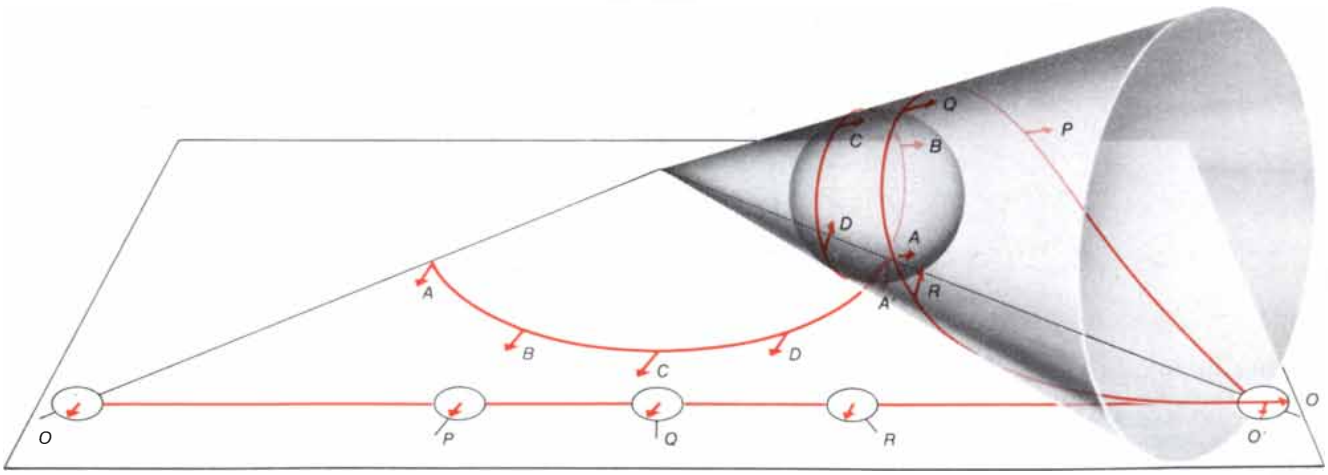
Parallel transport determines a path-lifting rule in the bundle of directions on the sphere. Given a curve on the sphere and an arrow representing a direction at the start of the curve, one can place the sphere on the plane so that the starting point is the point of tangency. Imagine there is an arrow at each point in the plane, drawn in ink and parallel to the original arrow. If the sphere is rolled along the plane so that it prints the curve on the sphere onto the plane, the arrows from the plane will also be printed onto the sphere. The latter arrows give a path in the bundle of directions that starts with the initial arrow and lies above the given curve.

When a straight line is printed onto the rolling sphere (or onto any other curved surface), the curve it forms on the surface is called a geodesic. On the sphere a geodesic is a great circle. The shortest path on the surface between two points is a geodesic. The arrows that all point in the same direction on the plane retain a vestige of this property when they are printed onto a geodesic. At each point along the geodesic curve the angle between the arrow and the line tangent to the geodesic is the same.

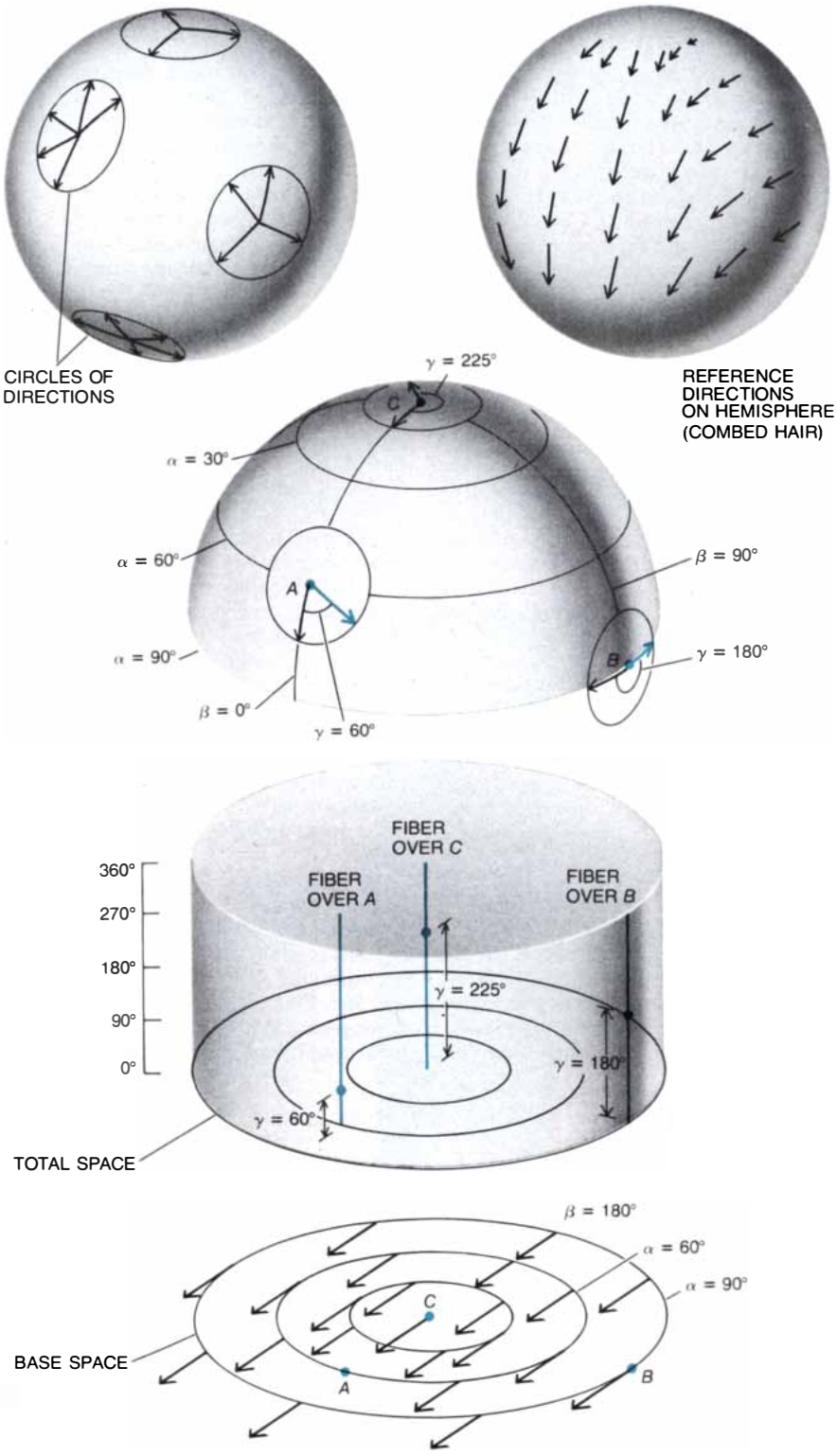
Parallel transport can be described for a curved path made up of geodesic segments without reference to rolling. It can be carried out by maintaining a constant angle between the transported arrows and the tangents along successive geodesics. From a perspective above the surface of the sphere, however, parallel transport along a geodesic may seem anything but parallel: the arrows may appear to rotate. When an arbitrary curved line is printed onto the sphere, the rotation of the arrows as viewed from above the surface may appear to be even more chaotic.

Parallel transport provides a way of making quantitative and explicit the intuitive difference between a curved surface and a flat one. When an arrow is carried by parallel transport around a closed path in the plane, the directions of the arrow at the start and at the finish coincide. Parallel transport around a closed path on a curved surface, however, may not lead to such coincidence. If there is a change in the direction of the arrow when it completes a single circuit of a closed path, the angle between the final direction and the initial one is called the angular excess of the path. It follows from the way parallel transport is defined that the angular excess does not depend on the initial direction of the arrow.

Mathematicians commonly express angular measure in radians instead of degrees. (Conversion from degrees to radians is made by multiplying the degree measure of an angle by the constant $2\pi/360$. One radian is therefore about 57 degrees.) When the angular excess is measured in radians, the result is a num-



PARALLEL TRANSPORT carries a direction along any curve in a plane so that an arrow points in the same direction everywhere on the curve. To extend the idea to parallel transport along a curve on a surface that is not planar, one can imagine that parallel directions in the plane are printed onto the surface as the surface rolls on the plane, without slipping or twisting about the vertical, so that the point of tangency always remains on the curve. To roll a sphere along one of its circles of latitude it is convenient to draw a cone tangent to the sphere along the circle. As the cone rolls on the plane, the sphere rolls along the circle of latitude. When a surface is rolled along a straight line in the plane, the curve printed onto the surface is a geodesic. If a geodesic is printed onto a cone of sufficiently small vertex angle, it can circumscribe the cone and intersect itself at an angle called the angular excess, a measure of the curvature enclosed by the path.



BUNDLE OF DIRECTIONS on the surface of a sphere is an important example of a fiber bundle. At each point on the sphere there is a circle of directions along which one can look on the surface. To label these directions with angles one must assign a reference direction to each point. If the reference directions could be assigned everywhere in a continuous manner, one could "comb the hair" on the sphere, but that is not possible; there must always be a "cowlick." Hair can be combed, however, over any region smaller than the entire surface. For example, on a flat map of the northern hemisphere the description "downward and to the left" specifies a direction at each point and so defines a continuous set of reference directions on the hemisphere. A picture of the bundle of directions on the hemisphere can be made by adopting the flat map as a base space. Every direction at a point on the hemisphere appears on the vertical coordinate line above the corresponding point on the map, and at a height that corresponds to the angle the direction makes with the reference direction. Heights 0 and 360 correspond to the same direction. The total space of the bundle is a cylinder where points at the top and the bottom of each vertical fiber are identical. The arrows that represent the reference directions on the map are parallel, but their counterparts on the sphere do not represent parallel transport.

ber called the total curvature of the region enclosed by the path. One can then define the average curvature of a region as its total curvature divided by its area. By convention the sign of the average curvature is given correctly when the arrow is transported along the path counterclockwise, so that the region is on the left of the path. The curvature of a surface at a point can be defined as the limiting value of the average curvature of progressively smaller regions containing the point.

Parallel transport makes it possible to define a path-lifting rule from the surface of the sphere to the total space of all directions. In the total space the angular excess is represented as the angular distance along the fiber that corresponds to the point on the closed path where the circuit of parallel transport begins and ends [*see illustration on opposite page*]. Hence the total curvature of the region enclosed by a path in the base is represented by a distance along one of the fibers in the lifted path. It turns out that replacing parallel transport by an arbitrary path-lifting rule generalizes the notion of curvature to other bundles on which the operation of parallel transport does not make sense. Such path-lifting rules have to be formulated without any reference to geodesics or angles.

Instead of focusing on the geometry of the base space, as one does in parallel transport, it is possible to lift a path by imposing structure on the total space. One way of doing this is to associate a set of parallel, sloping planes with each fiber. The slopes of the planes determine how fast a lifted path rises or falls as it moves from fiber to fiber in the total space. The planes must never be parallel to the fibers. Their slopes must vary continuously from point to point, and they must have the same slope at every point along a given fiber. The latter condition is an analogue of the guarantee provided by parallel transport that the angular excess (and thus the curvature) is independent of the initial direction of the arrow being transported. Such a set of continuous planes in the total space is called a connection in the fiber bundle. At the point on a fiber where a lifted path crosses the fiber the path must be tangent to the sloping plane associated with the point. This is how the plane defines the slope of the lifted path at the point. The illustration on the opposite page shows the connection that lifts paths on the sphere in the same way that parallel transport does.

The curvature of a connection can be defined by a procedure similar to the one employed for measuring the curvature of a surface. The aim of the procedure is to assign to each point in the base a number that represents the curvature at that point. (For higher-dimensional spaces the curvature is specified not by a single number but by a collection of

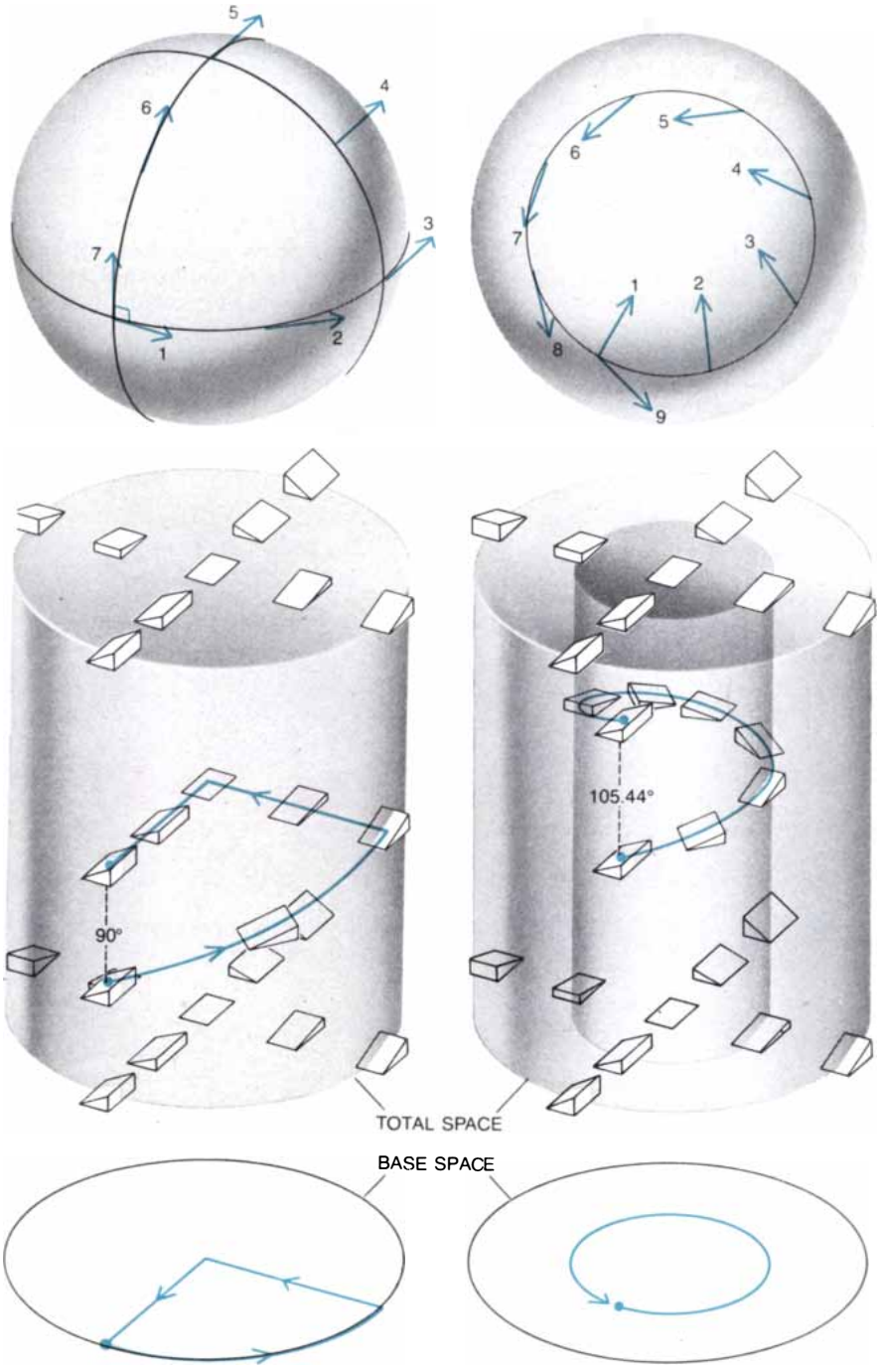
numbers that are the components of a mathematical object called a tensor.) The number that measures the curvature of the connection is obtained by finding the analogue to angular excess for progressively smaller paths enclosing the point. A connection has zero curvature over a region in the base if every sufficiently small closed curve in the region lifts to a closed curve in the total space. When this happens the connection is called flat, by analogy with parallel transport in plane geometry. If the connection has nonzero curvature, small closed paths in the base lift to yield curves in the total space that do not close.

Motion along a lifted path in a region where the connection is flat is analogous to motion on a hillside. Following a closed path may take one to various elevations above sea level. No matter how circuitous the route, however, when one returns to the surface coordinates (longitude and latitude) of the starting point, one has also returned to the same elevation. Motion in a curved region of the connection is analogous to motion in a cave. Some routes through the tunnels of the cave may lead back to the same longitude and latitude coordinates, but one's elevation may be quite different from what it was at the start of the path. Mathematically a connection above a region in the base is flat only if the planes of directions defining the connection are tangent to a family of surfaces. Each surface corresponds to the hillside in the analogy. The surfaces must conform to one another and pack together like spoons to fill up the part of the total space that is above the region of the base. Each surface must have the same number of dimensions as the base has.

It is now possible to give some account of how fiber bundles can represent the distinctive features of a gauge field. A more descriptive name for a gauge field would be a phase-shift field. In current gauge-field theories of nuclear forces the phase shifts act on quantum-mechanical waves to change the identity of the particle the wave describes. For example, to turn the probability amplitudes for a proton field into those for a neutron field and back again, which is in effect to alter continuously the probability that a particle is a neutron or a proton, it is sufficient to shift generalized phases.

The quantum theory of magnetism provides a much simpler example of the phase-shifting properties of a gauge field. The gauge field in question is called the magnetic vector potential, and it determines how electrons interact with a magnetic field. The clearest way to demonstrate the effect of the magnetic vector potential experimentally exploits the interference patterns of electron waves.

An electron in a beam can be repre-



LIFTING A PATH in a fiber bundle is a means of finding a path in the total space starting at a given point and lying directly above the path in the base. For the bundle of directions on the surface of the sphere, parallel transport of directions on the surface gives a unique lifting for every path. The bundle of directions over the northern hemisphere can be represented as a cylindrical total space. Each direction is at a height corresponding to the angle it makes with the reference direction, here chosen to be pointing to the lower left in the flat map of the northern hemisphere. For the path around the spherical geodesic triangle the reference directions point south along the meridian at the start of the path. The angle between the transported direction (*colored arrows*) and the tangent to a geodesic is constant. On the first leg of the triangle the geodesic curves with respect to the reference direction, so that the angle between the transported direction and the reference direction increases at a constant rate. Along the second and third legs of the triangle the transported direction maintains a 180-degree angle with the reference direction. When the arrow returns to its starting point, its direction has changed by 90 degrees, the angular excess of the closed triangular path. The changes in transported directions are plotted as a lifted path in the bundle of directions on the hemisphere. For the path around the 45-degree latitude the transported direction begins 180 degrees away from the reference direction and increases its angle with the reference direction at a constant rate. A connection can define a path-lifting rule without reference to parallel transport by assigning planes to every point in the total space. The lifted path must be tangent to the planes. The slopes of the planes are the same for all points along a single fiber, but they vary continuously from fiber to fiber; the planes are never vertical. Such a collection of planes is called a connection in the fiber bundle.

sented by waves whose length is inversely proportional to the electron's momentum and whose frequency is proportional to its energy. At any point in space such a wave has a definite height at each instant, just as a water wave has an instantaneous height above or below the average surface level of the water. The sequence of heights at a point varies periodically from a maximum to a minimum and back again. A graph of the heights resembles a cosine curve. Because the cosine is a function whose argument is an angle the instantaneous height of a wave can be stated by giving the wave's maximum height and an angle that corresponds to the instantaneous height of the cosine curve. The curve attains its maximum height, for example, at 0 degrees and its minimum, or greatest negative value, at 180 degrees; it has a height of zero at 90 degrees and 270 degrees.

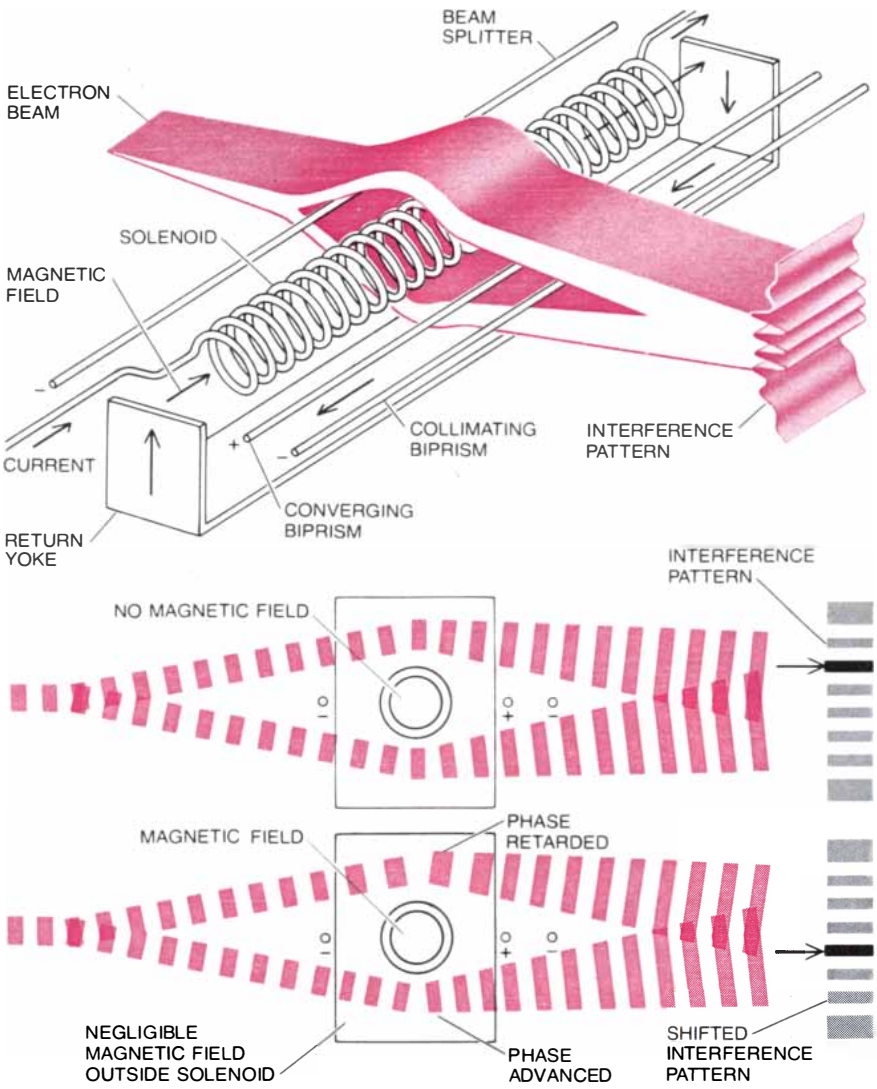
The angle corresponding to the instantaneous height of an electron wave is called the phase angle of the wave. The phase angle does not affect the probability of finding an electron at a point, because the probability amplitude is proportional only to the maximum height of the electron wave. (As usual, the probability is found by squaring the probability amplitude.) Thus if one shifts the phase of the electron wave arbitrarily at each point in space, the probability of finding an electron at any point does not change. A function that assigns such a local phase shift to each point in space is called a gauge transformation.

Although the overall phase of a single electron beam has no effect on observed quantities, the relative phase with which two partial beams arrive at the same point does have physical consequences. A phase difference between two interfering beams can alter the maximum height of the electron wave and so can change the probability amplitude.

An interference pattern shows up as a variation from one point to another in the probability of finding a particle. Hence wherever two partial beams overlap they create an interference pattern. Suppose the experiment is symmetrical and the beams are exactly in phase at the center of a detecting screen. The value of the phase at the center of the screen can be changed at will, but the same change will then be made in each of the partial beams. Therefore the interference is constructive under any gauge transformation at the point.

The interference pattern is formed because for points on the detecting screen, say to the left of the center, the waves from the left partial beam have traveled a shorter distance to reach the screen than the waves from the right partial beam. As one searches farther from the center of the pattern the phase difference or relative phase takes on greater values, creating the periodic variation in intensity that constitutes the interference pattern. The pattern is detected as a variation in the counting rate in the region of overlapping partial beams. If some physical effect introduces a phase difference between the two partial beams arriving at the center, the same phase difference will also be found at each point to the right or left of the center. The "fringes" in the interference pattern will be uniformly shifted.

For electrically charged particles the magnetic vector potential is a field that acts by shifting phases at each point in space. The vector potential determines the magnetic field; indeed, magnetic effects on charged particles can be completely explained in terms of the phase shifts given by the vector potential field. The underlying reason the vector potential field is a gauge field is that the magnetic force acts on a charged particle to change its direction without changing its energy. When an electron enters a magnetic field, the frequency of the electron waves therefore remains constant but the spatial pattern of the waves is changed. It is as if the wavelength were to vary from point to point. The phase shift of an electron caused by a magnetic field therefore depends on the path of the electron.



ELECTRON PHASE SHIFT can be induced not only by a magnetic field but even by the passage of the electron through a region near a magnetic field. The effect can be detected by splitting an electron beam with a negatively charged wire, passing the two partial beams around opposite sides of a solenoid within which a magnetic field has been confined and then recombining the partial beams so that they interfere. As the current in the solenoid is increased the magnetic field increases and the interference pattern of the overlapping beams is shifted. Because the electron beams remain coherent only if they are separated by no more than 60 micrometers the experimental apparatus is exceedingly small: the diameter of the solenoid is less than a seventh the thickness of a human hair. An experiment of this kind was first proposed in 1959 by Yakir Aharonov of Yeshiva University and David Bohm of the University of London.

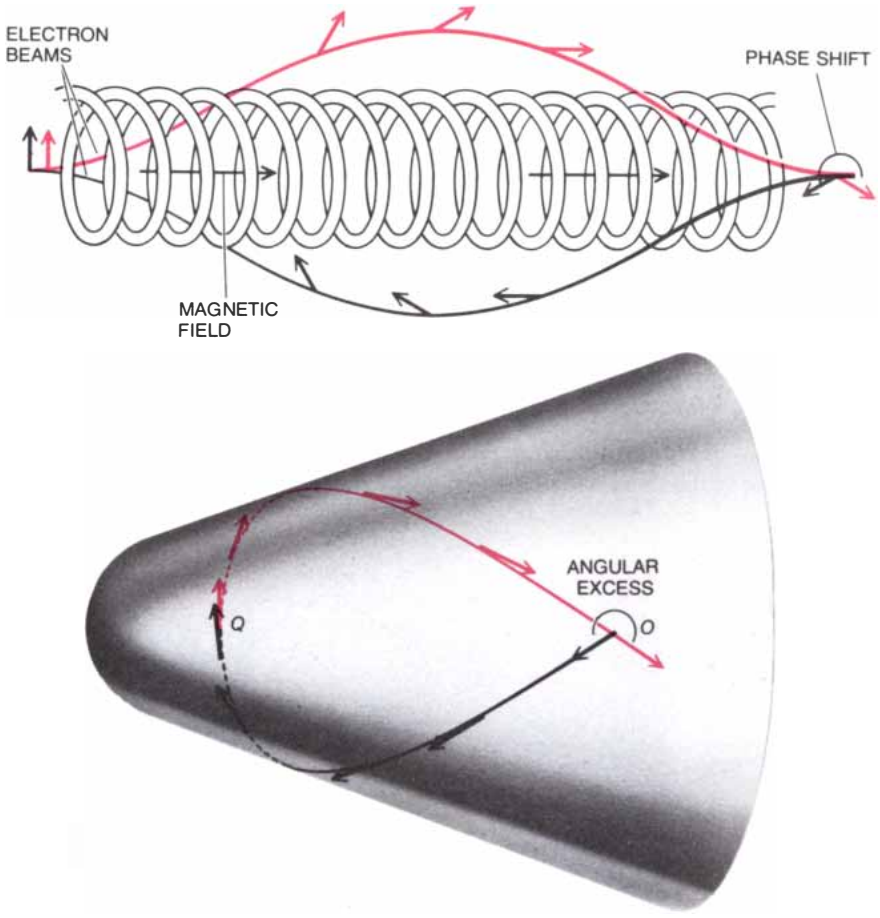
In the classical theory of magnetism the magnetic vector potential was conceived as an auxiliary device for calculating the magnetic field. The calculations showed that the magnetic vector potential could be nonzero in regions where there is no magnetic field. As a result physicists thought the magnetic vector potential would not have observable consequences of its own.

In quantum mechanics the magnetic vector potential field does have observable consequences. Its expected effect on phase was mentioned by W. Ehrenberg of the University of London and R. E. Siday of the University of Edinburgh in 1949. It was not until 1959, however, that Yakir Aharonov of Yeshiva University and David Bohm of the University of London proposed an experiment in which the effect could be observed directly.

Like the neutron-rotation experiment, the Aharonov-Bohm experiment calls for splitting a beam of subatomic particles, recombining the partial beams and observing the resulting interference. The particles employed are electrons rather than neutrons. This presents a difficult technical problem, because the maximum separation over which the two parts of a split electron beam remain coherent is only 60 micrometers. Even to achieve such a small separation the entire experiment must be set up inside an electron microscope. The experiment was first carried out by R. G. Chambers of the University of Bristol in 1960. The effect was confirmed in 1961 by Gottfried Möllenstedt and Werner Bayh of the University of Tübingen in a somewhat more elaborate experiment.

The incoming electron beam travels toward a wire bearing a negative electric charge, which splits the beam into two coherent partial beams. The diverging beams pass around opposite sides of a solenoidal electromagnet with an outside diameter of 14 micrometers (less than a seventh the thickness of a human hair). The electromagnet is installed in such a way as to confine the magnetic field to the inside of the solenoid. Since the electrons remain outside the solenoid, they pass through a negligible magnetic field; any phase changes they undergo must therefore be attributed to the magnetic vector potential field that surrounds the solenoid.

Beyond the solenoid in the electron microscope is a wire bearing a positive charge that brings the beams back together. Another negatively charged wire deflects the beams so that they intersect at a small angle in order to increase the width of the fringes in the interference pattern. The resulting pattern has broad dark bands from diffraction at the beam-splitting wire and fine fringes from the interference.



PHASE SHIFT of the electron beams in the Aharonov-Bohm experiment can be modeled by the parallel transport of a direction on the surface of a truncated cone capped with a spherical dome. The shifting of phase along each beam is represented by the rotation of an arrow. The two partial beams begin in phase. The phase is then retarded along the colored path and advanced along the black path. The phases shift even though the magnetic field along both paths is zero; the shifts are directly proportional to the magnetic field between the paths. On the domed cone the geometry of the conical region is like the geometry of a plane: the region can be slit longitudinally and unrolled onto a plane without stretching or compressing. Parallel transport along two paths around the cone, however, will not lead to the same direction when the paths meet again, even though the curvature along both paths is zero. The paths shown form a curve that is a self-intersecting geodesic; the arrows represent the direction tangent to the curve. Because the tangent direction undergoes parallel transport along a geodesic the angular excess of the curve is identical with the angle at which the two paths meet. The angular excess measured in radians is equal to the total curvature of the region between the paths. The curvature is concentrated on the dome, just as the magnetic field is confined to the solenoid.

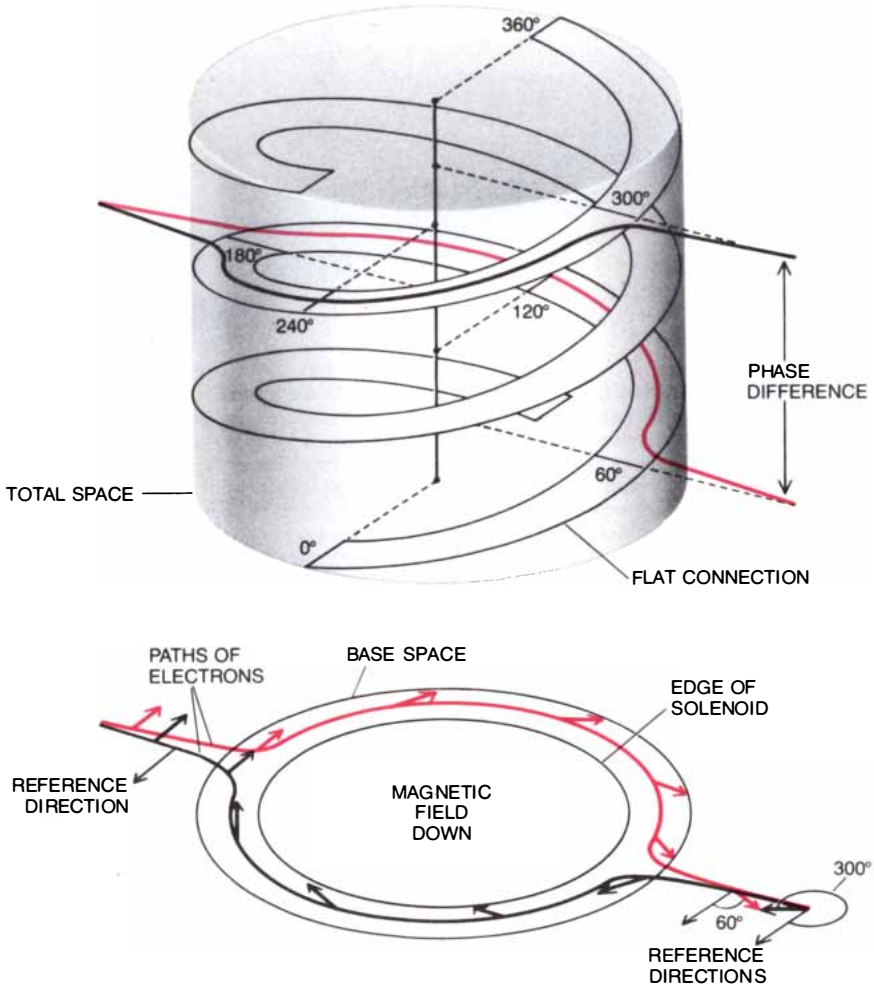
As the current in the solenoid increases, the magnetic field becomes stronger and the fine lines of the interference pattern shift with respect to the broad diffraction bands. The magnetic flux can be calculated from the dimensions of the coil and the current passing through it. Hence the experiment gives an exact numerical check on the relation between the magnetic flux and the phase shift; the results confirm the predictions of quantum theory.

The fiber-bundle model of this experiment illuminates the general correspondence between gauge fields and connections in fiber bundles. The model describes parallel transport on the surface formed by truncating a cone and capping it with a dome. Topologically the surface formed in this way is equivalent to the surface of a hemisphere, but it has a quite different geometry. Note that if the conical region is slit longitudinally, the cone can be unrolled onto a planar surface without stretching or compressing. Lengths and angles measured on the unrolled cone must be the same as they are when they are measured on the cone itself. Thus a two-dimensional observer measuring curvature on the surface of the cone by carrying out parallel transport over small closed curves would always find the angular excess to be zero. The surface would appear to be flat. This part of the surface is analogous to the region outside the solenoid in the Aharonov-Bohm

experiment, where the magnetic field is negligible.

If sufficiently large paths on the cone are considered, however, the difference between the cone and the plane will become apparent. A straight line on the unrolled cone that intersects both sides of the slit at the same distance from the apex will roll up into a closed loop that is a geodesic on the cone. Suppose a pair of two-dimensional observers stand, both facing forward, at the point on the loop directly opposite the point where the loop crosses the slit. If the observers move along the loop, one walking forward and the other backward, they must think they are moving away from each other along a straight line. When they meet again at the point where the loop crosses the slit, however, they no longer face in the same direction. The angle θ between them is equal to 360 degrees minus the angle at the top of the unrolled cone. The observers must conclude that the surface they live in is not flat after all, because this angle is the angular excess of the closed curve formed by their paths. The region to their left must have a total curvature equal to the measure in radians of the angular excess.

What happens to the pair of two-dimensional observers as they pass around opposite sides of the cone is similar to what happens to the pair of partial beams as they pass around opposite sides of the solenoid in the Aharonov-Bohm experiment. The two observers initially face in the same direction. Analogously, the two partial beams of electrons are in phase immediately after the beam is split. If the observers measure the curvature of the surface along either one of their paths, they will find it is zero because they are traveling on the flat part of the capped cone. Similarly, the magnetic vector potential, which causes the beams to be out of phase when they are recombined, cannot be detected by measurements made along either one of the separate paths.



FIBER BUNDLE of phase angles over a cross section of the Aharonov-Bohm experiment is almost exactly like the fiber bundle of directions on the surface of a spherically domed cone. Over each point in the base the fiber in the total space represents the possible phase angles (from 0 to 360 degrees) of the electron at that point. For an appropriate choice of magnetic-field intensity the connection defined in the bundle of phases by the magnetic vector potential field is identical with the connection given by parallel transport on the domed cone. The curvature of the connection corresponds to the magnetic field in the solenoid. The region inside the solenoid corresponds to the spherical dome. There the rule for lifting a path is the same as the parallel-transport rule on the surface of the sphere, and it can be described by the same system of sloping planes (not shown). The region outside the solenoid corresponds to the truncated cone. There the connection is flat. The sloping planes are tangent to a family of spiral ramps that fill up the total space. The phase shift in the base is represented by the rotating arrows.

The key to the similarity of the two effects is the interpretation of the magnetic vector potential as a connection in a fiber bundle. The bundle is called the bundle of phases. Its base space is the three-dimensional space in which the experiment takes place. The fiber over any point is the set of all possible phases of electrons at that point, and so the total space is made up of all possible phases of electrons at all points of three-dimensional space. Since a phase can be described by an angle measured in degrees or radians, the fiber over each point is a circle, just as it is in the bundle of directions on a surface. With a circular fiber for every point in three-dimensional space, however, the total space is four-dimensional.

The three-dimensional base space and four-dimensional total space make it difficult to visualize the bundle of phases. A picture can be drawn, however, of the bundle of phases over a plane that passes through the experiment perpendicular to the solenoid. This bundle, which has a three-dimensional total space, is completely adequate for describing the outcome of the Aharonov-Bohm experiment.

The bundle is now a cylindrical bundle over a surface; it can be drawn with vertical fibers like the bundle of directions on the hemisphere or on the capped cone. The connection in the bundle is a field of planes transverse to the fibers. The slope of one of the planes in a given direction (say up or down the ramp) is proportional to the component of the magnetic vector potential field in that direction at the corresponding base point. The curvature of the connection is proportional to the intensity of the magnetic field. Hence the curvature is different from zero only at points inside the solenoid. For an appropriate choice of the field strength inside the solenoid, the picture of the magnetic-vector-potential connection over the plane perpendicular to the coil is identical with the picture of the parallel-transport connection over the capped cone [*see illustration at left*]. The region of the base inside the solenoid corresponds to the spherical cap on the cone. There the connection is identical with the connection that is generated by parallel trans-

port on a sphere. In the region of the base outside the solenoid the curvature of the connection is zero, since the magnetic field is zero there. The connection planes are tangent to a densely packed family of ramps that spiral around the center. The phase difference between the two partial electron beams is exactly the height gained by following one of the ramps through one full rotation.

If the curvature is calculated for the entire bundle of phases, it is found to be a tensor, since the base is three-dimensional. The tensor has three components, each proportional to the corresponding components of the magnetic field. The construction can be extended to the bundle of phases over space-time, where the base is four-dimensional. The electric potential and the magnetic potential taken together give a connection that can explain all electromagnetic interactions of charged particles. The curvature of the connection is a tensor with six components, corresponding to the three components of the electric field and the three components of the magnetic field.

Each of the quantum gauge fields can be understood as a connection in a fiber bundle where the base is space-time. The fiber of the bundle is the set of internal symmetry transformations of particles that interact by means of the gauge field. Fiber bundles have now been constructed for quantum electrodynamics, for the weak and the strong nuclear interactions and for versions of theories that attempt to present a unified account of these forces. Although a complete scheme for the unification of all the fundamental forces has not come forth, it is clear the mathematics of connections in fiber bundles is destined to play an important role in theoretical physics.

We believe the current usefulness and physical significance of such mathematical concepts is no accident. Neither mathematicians nor physicists are insulated from their cultural, political and physical milieu, and the ideas and perceptions of workers in each discipline are influenced by the other discipline. Moreover, mathematicians and physicists unavoidably share unspoken assumptions about the everyday world and the logic by means of which the world is projected onto abstract science. Indeed, they share a passionate commitment to such rational work. What seems most marvelous is not what has been called the "unreasonable effectiveness" of mathematical concepts in physics, or the fecundity of physical intuition as a source of new mathematics. Rather one must admire the success of the common intellectual approach of mathematicians and physicists in creating a rich, coherent and powerful image of the physical universe.

MATING FIREFLIES hang in the shelter of a leaf. The male, the lower of the two, is grasping the female's wing covers with both forelegs and has raised both middle legs. These fireflies are the eastern U.S. species *Photuris hebes*. The female lures and eats alien males.

PREDATORY FEMALE of the eastern U.S. species *Photuris versicolor* dismembers an alien male firefly. When alien males advertise themselves during their search for mates, *P. versicolor* females attract them by mimicking responses of a female of the same alien species.

Mimicry in the Sexual Signals of Fireflies

Male fireflies flash as they fly seeking mates at night, and females of the same species flash answers from the ground. Females of other species also answer and devour any males they can lure within reach

by James E. Lloyd

The myriad flashes and flickers in a meadow full of fireflies fills the spectator with a bucolic content that is a world removed from the reality of competition, deception and predation lying behind that twinkling facade. Limited resources, among them the supply of unmated females and food, make competition an inevitable part of life. Such competition is more easily perceived among fireflies than it is among other insects because much of it is mediated by bioluminescence.

Most of the dancing lights in a firefly meadow are the self-advertising signals of male fireflies seeking a mate. With a little effort an observer can learn to recognize several firefly species from their flashes alone. As a general rule, in any one area each species has its own unique male signal and female response. The identifying features of the signal are the duration of the flash, the number of flashes in a signal pattern, the rate of the flashes in the pattern and the rate of repetition of the pattern. The male signals of some 130 firefly species around the world are known, at least to some degree, and I have analyzed the patterns of more than 80 species on the basis of recordings of free-flying males in the field.

About 50 of the 130 signal patterns consist of single flashes in a continuous sequence, emitted at various characteristic rates. Another 30 firefly species have signal patterns consisting of groups of flashes that range from two to 11 flashes per group. Another 30 species have a flickering signal pattern, usually with five to 11 peaks of brightness, at rates of up to 20 peaks per second. Finally, some 20 species emit flashes with what is called cryptic modulation because their frequency (more than 30 flashes per second) is too high for the human eye to detect.

What makes the pattern of a particular species distinctive is not only the rate of flashes or flickering but also how the intensity of the flash increases and decreases and how the intensity is modulated within the flash. The shortest flash I have observed is the 40-millisecond spike of *Photuris potomaca,* a firefly that lives along the Potomac River and its larger tributaries. Although some firefly species emit glows that last for minutes, the longest true flash I have seen is the five-second crescendo of a Missouri race of *Photuris lucicrescens.* The most rapidly modulated flicker is that of *Photuris* "D," one of several Florida species not yet named; on warm evenings its flicker reaches more than 40 peaks per second.

With all fireflies the flashing rate changes predictably with the ambient temperature. Because of such changes and the subtlety of the species-specific elements of many display patterns positive identification of male fireflies in the field usually calls for several additional kinds of information, both behavioral and ecological. They include flight speed, altitude during light emission, maneuvers during light emission, the habitat, the activity space within the habitat, the time of the activity and the season of the year.

The flash patterns of the male firefly are only half of the species-specific code. The other half is the female response in the courtship dialogue. In most instances the female's responses to the male signal are short, unmodulated pulses. Among the exceptions to this generalization are the repeated flashes of the female *Photinus obscurellus,* a marsh-dwelling species of the northeastern U.S., and the seconds-long glow that *Photinus tanytoxus,* a species of the Florida grasslands, sometimes emits after flashing. These elaborations offer an approaching male a conspicuous target.

The characteristic of the female flash that often shows significant variation among species, and the most critical characteristic of the female response, is its timing with respect to the male flash. For example, at 76 degrees Fahrenheit the female of the eastern U.S. firefly species *Photinus ignitus* waits for three seconds before answering the male signal. (At 55 degrees the delay time is about nine seconds.) The female delay times of other firefly species are shorter. It is the combination of the male signal pattern and the delayed female answer that constitutes the species-specific code. That code has long been recognized as playing an important part in maintaining reproductive isolation when individuals of different species come within signaling range of one another.

On a given evening in the firefly meadow mate-seeking males greatly outnumber the females; the ratio may be 50 to one. As a consequence male competition is keen and a male searches a long time before finding a potential mate. When a female shows a light, however, she is quickly discovered. In order to quantify male mating efforts and successes I have followed males of the species *Photinus collustrans* in their twilight hunt.

This firefly is sexually active for 15 to 20 minutes each evening, beginning about 20 minutes after sunset. Its mate-seeking period is squeezed between the time when it is still twilight and the onset of sexual activity in the closely related firefly species *Photinus tanytoxus,* which begins its mate seeking about five minutes after the early-flying species finishes. The short duration of the *P. collustrans* males' search for mates and the fact that their activity space in Florida pastures is virtually two-dimensional make them ideal subjects for a quantified study of mate-seeking behavior.

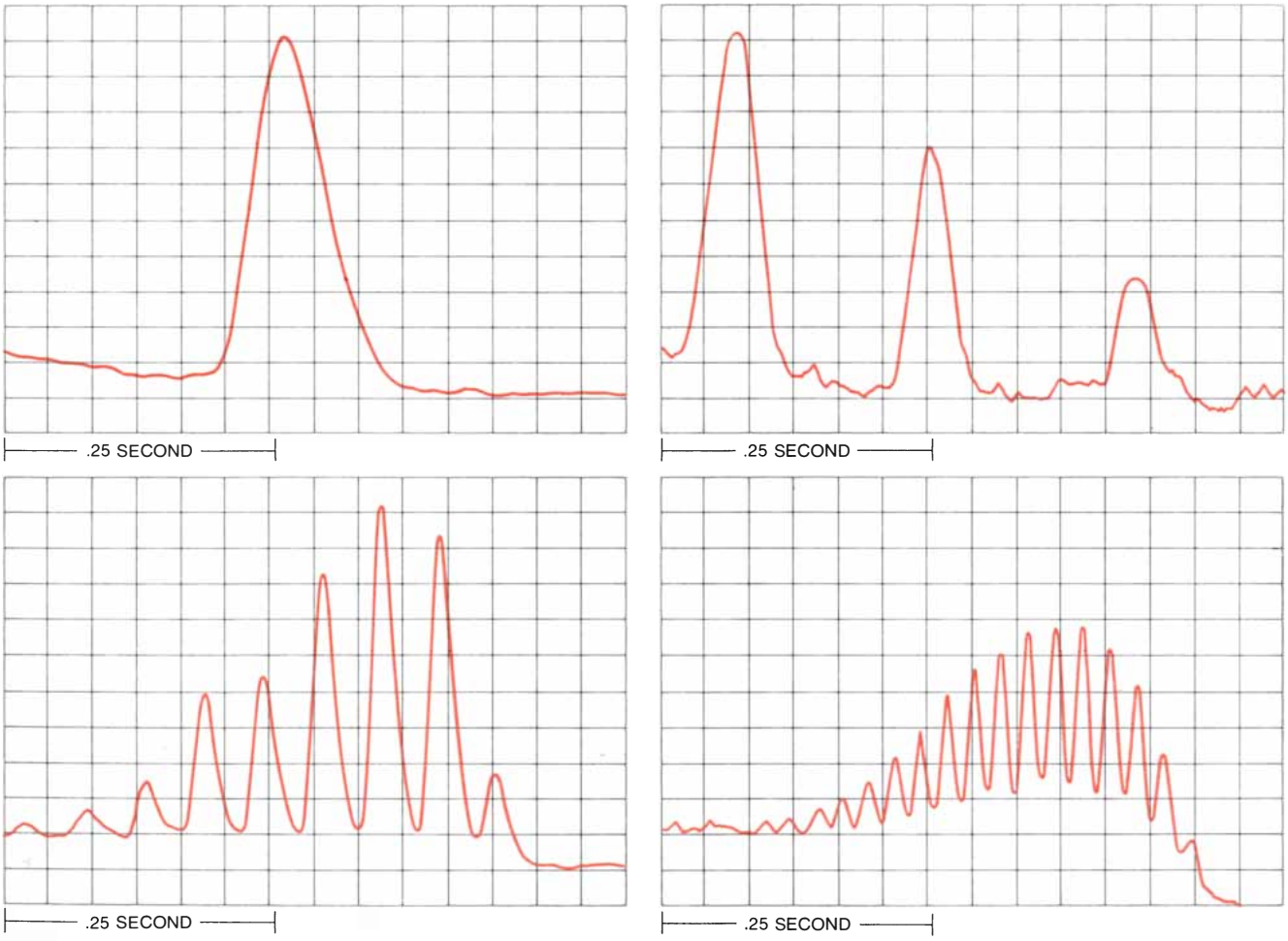
When I was following the males, I used a rolling wheel with a counter to measure distance. I also counted the males' flashes and crashes and recorded the duration of each flight. Over several seasons I have tracked 199 *P. collustrans* males for a total distance of 10.9 miles. The males flashed 7,988 times. (Their flash pattern is a single .3-second signal that is repeated every two seconds.) There were only 11 crashes, each result-

ing in the loss of up to 15 seconds of flight time. Of the 199 mate-seeking males only two found females of their own species and mated. The males' flashes were answered, however, by 11 females of another firefly species that mimics the response signal of *P. collustrans* females, and one of the males I was tracking wasted four minutes of his limited time investigating the response of one of these mimickers. It thus appears that during the average mate-seeking *P. collustrans* male's evening flight the firefly travels a distance of .62 mile and emits 455 flashes. That average searcher has a .63 chance of a crash and a .14 chance of finding a mate. This means that an average male must undertake 7.1 evening searches if he is to find a mate.

Female success is much quicker. After the *P. collustrans* female emerges from her burrow at twilight she is able to catch a flying male's attention, attract him to the ground, copulate (for 90 seconds) and return to her burrow—all in about six minutes. When the female disappears, the male immediately returns to the air and to renewed competition. Both the male and the female run a risk of predator attack if they remain at the mating site.

Most adult fireflies are not carnivorous, but among the 60-odd species of the genus *Photuris* the females of most species are an exception to this rule. They attract and devour the males of other firefly species, and their mimicry of the females whose males they attract is remarkable and versatile. The hunting female flies to an area where a prey species is active and takes a station on or near the ground. When a male prey flashes nearby, the *Photuris* female answers with the female response characteristic of the prey species. The male flies closer and flashes again. The male must always hurry. If he is slow about it, rivals close at hand will see the courtship and intrude on it. A mistake in identification, however, could be fatal. The attracted male may respond by flashing several times, moving in close and backing away. Some males then leave, some dally and some land a few inches from the answering light. From my observations of four prey species I estimate that 16 percent of the attracted males are duped and are eaten when they land near the predaceous female (which may be several times their size).

The females of two Florida *Photuris* species have an extensive repertory of

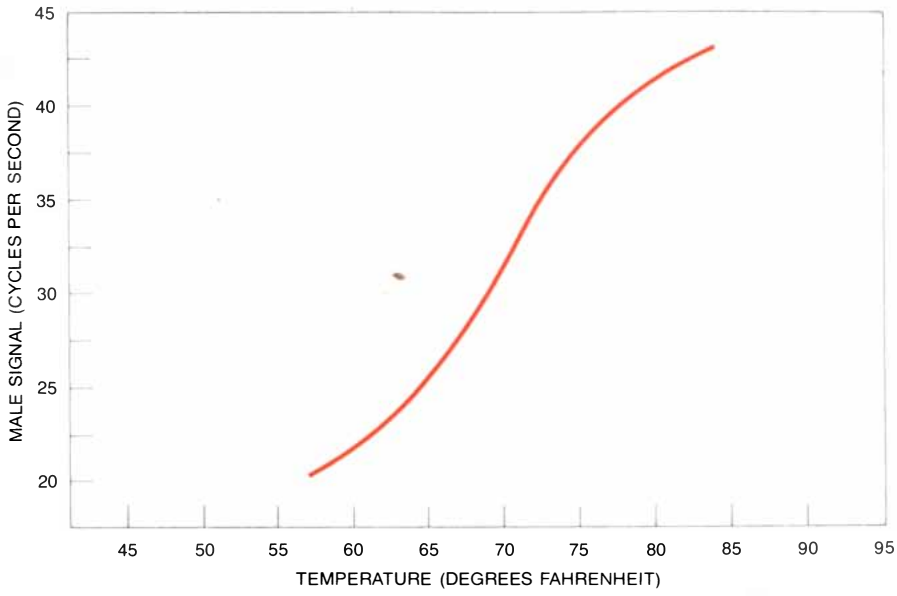


SEXUAL SIGNALS of male fireflies fall into four categories, illustrated here by the traces of field recordings that convert photomultiplier readings into sound on tape. The time bars on all four examples (*color*) indicate .25-second intervals. Roughness in the traces is due to electronic "noise" (*valleys*) or recorder overload (*peaks*). At the top left is a single flash that, emitted at various characteristic rates, is typical of about 50 of the 130 species of fireflies whose male signals are known. It is a recording of the species *Luciola peculiaris* made in New Guinea at a temperature of 70 degrees Fahrenheit. At the top right is a group of flashes; this kind of signal, which may consist of between two and 11 flashes per group, is typical of 30 species of fireflies. The three-flash group seen here is the signal of *Photuris versicolor,* recorded in Florida at a temperature of 71 degrees F. At the bottom left is a flicker; this signal, with as many as 11 peaks of brightness at rates of up to 20 peaks per second, is typical of 30 species of fireflies. The multipeaked flicker shown here is the signal of *Photinus evanescens,* recorded in Jamaica at a temperature of 71 degrees F. At the bottom right are the peaks of what is called a cryptic signal: rapid peaks of brightness that the human eye cannot resolve. Typical of 20 species of fireflies, these cryptic signals can reach a rate of 45 peaks per second. This is the signal of an unnamed species of *Photuris,* "D," recorded in Florida at a temperature of 76 degrees F.

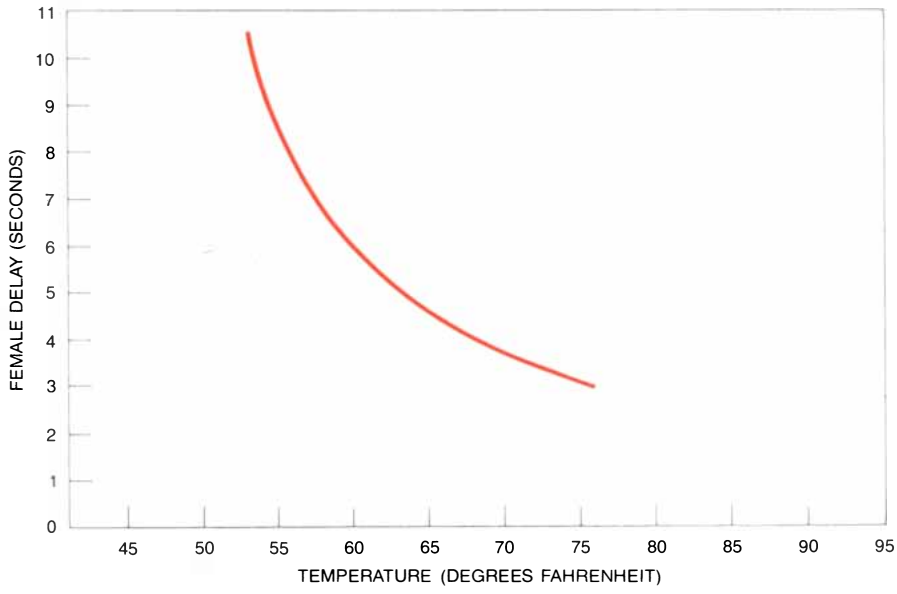
mimicry and prey on males of more than one other species. The *Photuris versicolor* female can attract males of at least five other firefly species, including members of all three of the major firefly genera of North America: *Photinus, Pyractomena* and her own genus, *Photuris.* In modifying their false signals according to the species they seek, the females of these *Photuris* species adjust not only the characteristics of the flash form but also its timing. Depending on the male flash pattern they detect, the females switch their reply appropriately.

For example, a *Photuris versicolor* female responds to the half-second flash of a *Photinus tanytoxus* male by giving the reply of the female of that species: a long flash followed by a diminishing glow. Presented with two short flashes two seconds apart (the pattern of the male *Photinus macdermotti*), the *Photuris versicolor* female flashes after the second flash just as a *Photinus macdermotti* female does. I have experimented with these females in the field, using a penlight to simulate the flashes of various prey species, and when a signaling firefly male of a species other than the one I was simulating passed by, my experimental subjects would answer the intruder appropriately.

The *Photuris* female may have a reserve tactic to improve her percentage of captures if false signals alone do not suffice. On several occasions when I was sending prey flash patterns to the predaceous females, they have landed on my hand or penlight or have flown into my face. Do they perhaps launch aerial attacks on hesitating and wary males?



AMBIENT TEMPERATURE affects the timing of the male firefly's sexual signal. The curve traces the increased frequency of the flicker of a male of the same *Photuris* species, "D," from 26 cycles per second at an air temperature of 58 degrees F. to 42.5 c.p.s. at 82 degrees F.



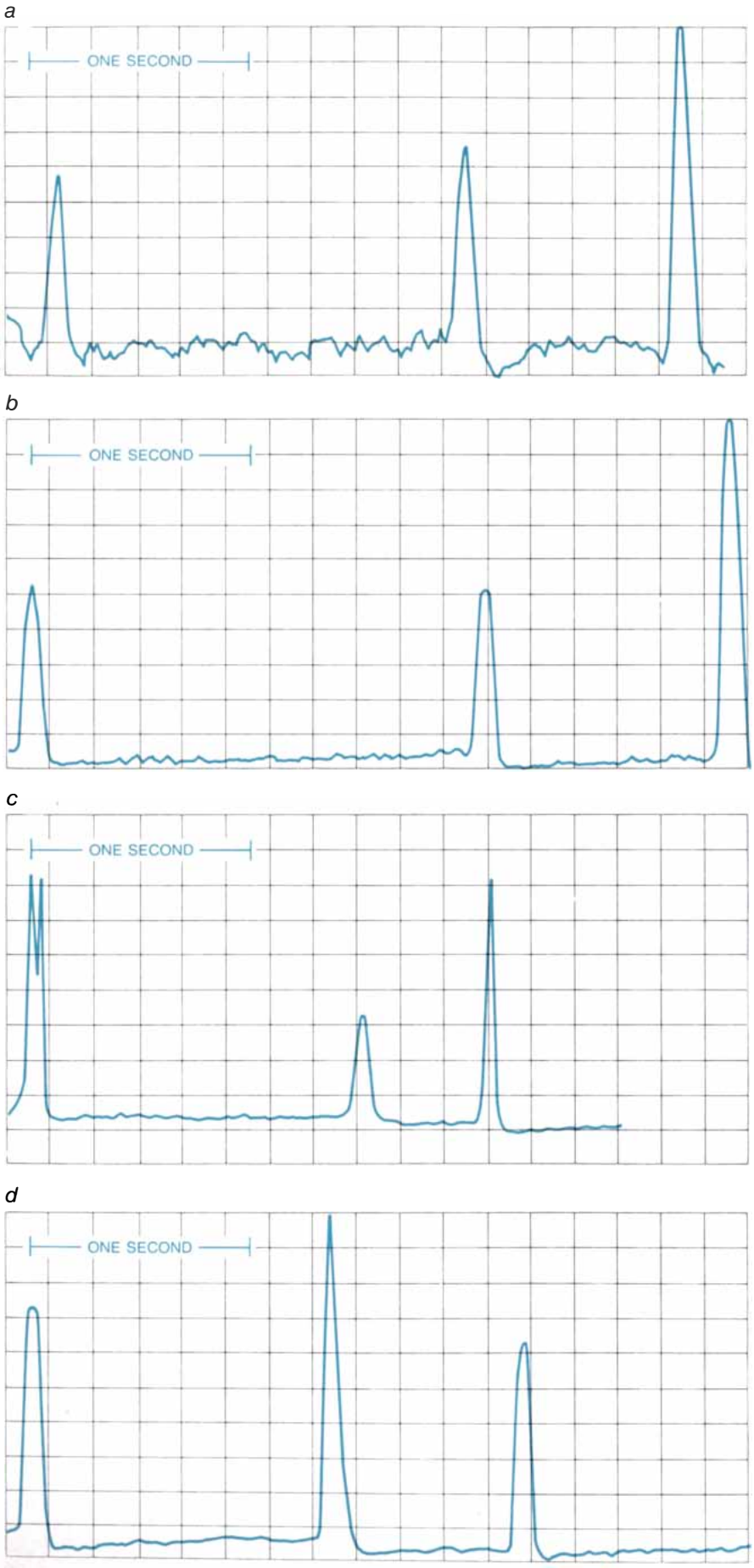
DELAYED RESPONSE of a female firefly to a male's sexual signal is also affected by the ambient temperature. The curve traces the delay of a female of the species *Photinus ignitus.* The delay is more than 10 seconds at 53 degrees F. but falls below four seconds at 76 degrees.

The mimicking behavior of *Photuris* females has had a profound effect on the mating behavior of *Photinus macdermotti* males. This species of small firefly is preyed on by the females of at least three *Photuris* species. The male *Photinus macdermotti* approaches an answering female cautiously. Rather than quickly wheeling in to land beside her, like the male of *Photinus collustrans,* the attracted male may hover and flash several times and then land some feet away from the female. Such a slow pace leaves time for other *Photinus macdermotti* males either to discover the female themselves or to observe the courtship interaction and intrude on it. Near single *P. macdermotti* females I have found congregations of up to five males.

Such competitors have recourse to a number of tactics. The commonest one is the "injection" of a flash between the first and the second of the two short flashes, two seconds apart, that constitute the pattern of the rival. The flash of the standard pattern has a single peak of intensity; the injected flash can also have a single peak, but often it has two peaks and occasionally it will have three. The competitor generally injects these flashes during the latter half of the interval between the flashes of the rival's flash pattern, and at least three different tactics may be involved.

When a competitor injects a single flash in the last quarter of the interval between the flashes of the rival's flash pattern, the effect as far as the female spectator is concerned is to shorten the interval between flashes in the male flash pattern. The female will often respond to the shorter-interval pattern and is more likely to do so as the pattern approaches the "correct" two-second timing. This gives an advantage to the flash-injecting male because the female normally aims her response flash toward a courting male's second flash. Therefore the flash-injecting male may see a bright beacon from the female and the standard-flashing male may not see the female's response at all. In this way the male emitting an early second flash can sometimes "steal" the female's answer from a rival.

The competitive advantage of such behavior is easy to appreciate, assuming that the flash-injecting male has found a mate and not an executioner. The fact



is, however, that most of the injected flashes are early rather than late. They are emitted in the third quarter of the standard male flash pattern, earlier than the injected flashes most likely to elicit the female's response. Could these early injected flashes mimic the shorter signal patterns of males of a related firefly species that is often present with *Photinus macdermotti?* In other words, does the competitor make it appear to the female that an alien male is only a flash away? Natural selection in the context of reproductive isolation has certainly resulted in strong female discrimination against the patterns of alien males. I have never been able to induce *P. macdermotti* females to answer my simulations of the patterns of related species.

In order to determine the effect of injected male flashes on female responses I presented simulated flash patterns to caged *P. macdermotti* females and put caged males of the same species where they could see the dialogue. The males sometimes injected flashes into my simulations. At other times I put a screen between the cages so that the females could not see any of the injected flashes. When the females were screened, they answered 98 percent of my simulations. When they were not screened, they replied to only 18 percent of the simulations that included the caged males' injected flashes. Next I used two artificial lights simulating competition between a pair of males. The screened females answered 94 percent of the uninjected patterns and 18 percent of the patterns that were injected when they were two-thirds complete. When I injected flashes halfway through the pattern, however, the females answered more frequently: 39 percent of the time. The significance of the fact that the females are more inhibited in responding to the first kind of injection than they are in responding to the second, and how this difference might be related to past and present selective

EXAMPLES OF MIMICRY of female responses and male rivalry by predatory females of the species *Photuris versicolor* appear in these traces of photomultiplier field recordings. Trace *a* shows the twin flashes that constitute the sexual signal of the male fireflies of the species *Photinus macdermotti*, followed by a single flash that is the response of the female *P. macdermotti*. In trace *b* the author has imitated the male flashes with a penlight; the response (*right*) is that of a *Photuris versicolor* female mimicking the female *Photinus macdermotti* response. Trace *c* shows another male *P. macdermotti*'s twin-flash signal, interrupted by the injected flash of a rival male of the same species. In trace *d* the author's imitation of a *P. macdermotti* signal is interrupted by a *Photuris versicolor* female's injected flash that mimics the male rival's.

pressures within the context of reproductive isolation, remain a matter for conjecture.

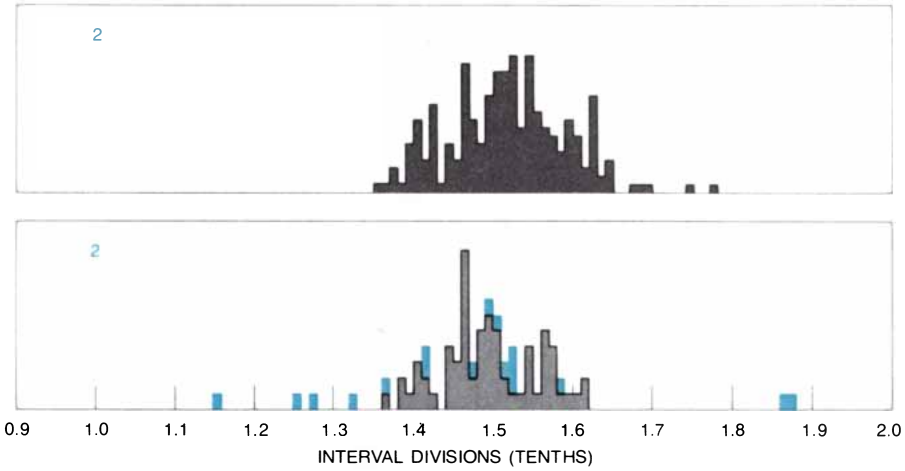
One aspect of flash injection relates to the mimicking signals of female predators. For example, when female *Photuris* predators mimic *Photinus macdermotti* females, they usually answer after the male's second flash. Sometimes, however, they also answer after the male's first flash. This early answer is not a mistake; its timing is like that of a flash injected by a *P. macdermotti* male. In effect the predator is simultaneously mimicking not only a much-sought-after female but also a rival male who has almost reached her.

Viewed in this way the significance of the flashes injected by the male *P. macdermotti* appears clear. The flash-injecting males are actually mimicking the female predators in such a way as to further retard a rival's cautious approach to a contested female. Such mutual (that is, reciprocal) mimicry is maintained in nature because by using their respective ruses individual *Photuris* predators enhance their success in the acquisition of prey and individual *Photinus macdermotti* males enhance their success in mating; each supports the other in continuing codependent mimicries. But what is the explanation for the *P. macdermotti* females' being variously inhibited in responding to the injected flashes? It could lie in the fact that the females too are at risk of predation by *Photuris* females. If they respond infrequently both to flashes injected by males of their own species and to the mimicking flashes injected by female predators, that risk is diminished.

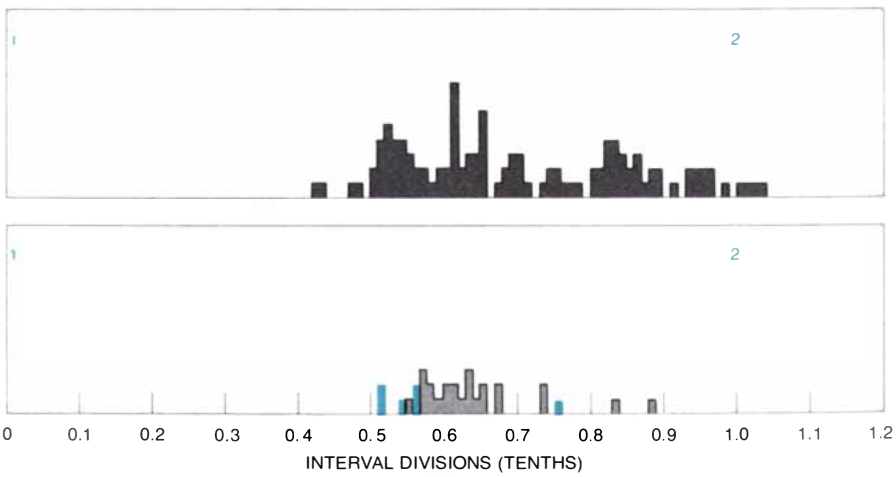
Experiments with simulated flashes have yielded evidence of other forms of sexual rivalry and deceit in *Photinus macdermotti*. For example, male competitors not only inject flashes but also flash synchronously with the second flash of their opponent's pattern. Such synchrony is observed in many firefly species, probably always in the context of competition for a mate. *P. macdermotti* males also mimic female flashes. They emit extra flashes after the female response flash, double and triple injected flashes and various combinations of all these emissions. The single injected flash, however, is by far the commonest of the male tactics.

Photuris females have mimicked the signals of their prey for millions of generations. Their multiple-signal tuning is remarkable, but it represents only one aspect of these strained and dynamic relations. The mating signals of prey species must have twisted and sidestepped repeatedly because of the pressures exerted by the predaceous mimics. Indeed, the most important influences on the signals of New World fireflies have probably been the prey-luring actions of the *Photuris* females. It could reasonably be argued that if this genus existed in Asia, it would soon wipe out the sedentary mating swarms of Asian fireflies, such as certain fireflies in Thailand that flash in unison. Some New World firefly species may have had similar synchronous mating protocols before the advent of *Photuris*. After that they would have become, so to speak, sitting ducks, and they would have soon disappeared.

Prey males are evidently receptive to a number of clues that would enable them to detect predators. For example, my firefly-meadow observations show that *Photinus collustrans* males were answered by more than five times as many female predators as females of their own species, and yet not one male was caught. On the average a *P. collustrans* male took only 42 seconds to reach and couple with a female of its own species from the time of first sighting. None of the 11 males signaled by *Photuris* females, however, ventured close enough to be seized. What kind of clue gave the



FEMALE MIMICRY by two *Photuris* species is shown in these histograms; time is measured along the abscissa as divisions of the interval (0 to 1.0) between the twin flashes that are the sexual signal of *Photinus macdermotti* males. This allows comparison of recordings made at various temperatures. The top histogram shows the time spread of 234 female *P. macdermotti* replies to a male signal. The earliest reply came a little more than three-tenths of an interval after the second male flash (*colored numeral* 2); the latest came a little less than eight-tenths of an interval after. Most replies fell on either side of half an interval. The bottom histogram shows the time spread of mimic replies of 85 *Photuris* females, *P. versicolor* (*gray*) and an unnamed species "B" (*color*). Most false answers also fell on either side of half an interval.



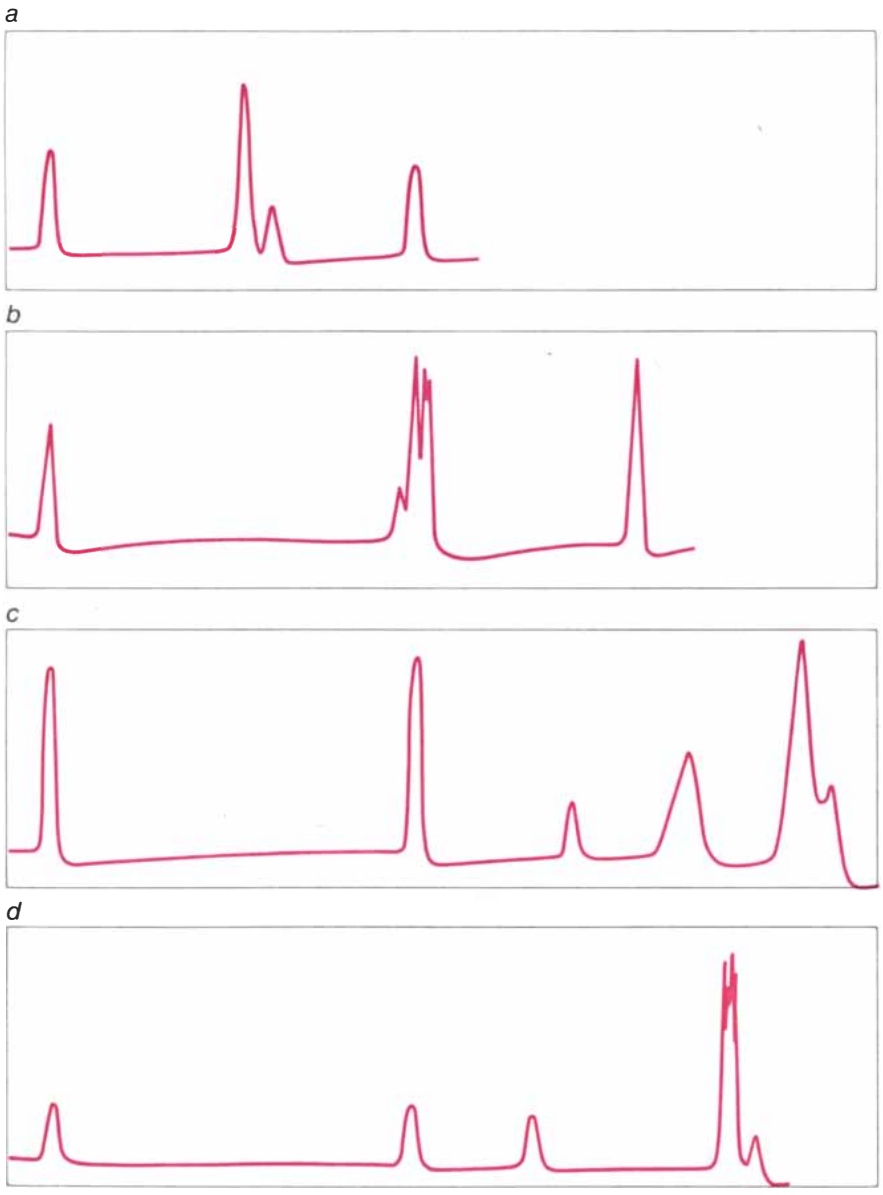
MALE MIMICRY by the same two *Photuris* species is shown in these histograms. The divisions on the abscissa (from 0 to 1.0) measure the interval between the first and second flashes (*colored numerals* 1 *and* 2) of a male *Photinus macdermotti* signal. The top histogram shows the time spread of 114 flashes "injected" by rival *P. macdermotti* males. Few of the injected flashes came earlier than half an interval after the first male flash or later than one full interval. The bottom histogram shows the time spread of 49 flashes injected by *Photuris* females, *P. versicolor* (*gray*) and unnamed species "B" (*color*) in mimicry of flashes injected by *Photinus macdermotti* males. None of the false flashes fell outside the time spread of the genuine ones.

predators away? Might the predators have been tricked by the males into betraying themselves?

Males probably respond to the brightness of answering flashes, and they certainly respond to the form and timing of female flashes. If the level of ambient light allows, they may also see the answering insect itself. A hunting *Photuris* female, if she is visible, could readily be distinguished from a prospective *Photinus* mate by her size, shape and stance. Could this be the reason most *Photinus* males fly at twilight? (There is no evidence that the colors of flashes are discriminated, although the luminescence of *Photuris* species is usually green and that of most *Photinus* and *Pyractomena* species is yellow.)

The mating signals of some fireflies *Photuris* females prey on may have become devious, incorporating tricks that cause the predators to botch their false answers. One simple scenario for the evolution of such countermeasures would make it possible to escape from *Photuris* signal tracking because the predator is simultaneously exploiting other species with a similar code. A *Photuris* female might not be able to coevolve with a prey species that gradually changed its code because of the "drag" exerted by the predator's continued tuning to other species and its nutritional dependence on them. Or if the timing of the predator's false answers began with the onset of its victim's flash, a prey species that shifted its response timing from the onset of the flash to its termination could cause deceivers to delay improperly. Hence the flash delays of an answering mimic could be shifted in predictable ways, say by the prey species' varying the length of its flashes, whereas those of females of the same species would remain constant.

Another conceivable predator-detecting code incorporates two types of pulse in a single-flash pattern. For example, in Jamaica, where *Photuris jamaicensis* females are versatile predators that feed on several other firefly species, one prey species of the genus *Photinus* flashes two to six times and then flickers. The preliminary flashes, because of their resemblance to the patterns of other prey species, may stimulate precipitate responses from *Photuris jamaicensis* females. The male-female flash alternations that occur during *Photinus macdermotti* mate competition may also be of sign-countersign significance.



VARIETY OF COMPETITION among rival *Photinus macdermotti* males is shown in these traces of photomultiplier readings. Using a penlight, the author imitated the twin-flash male sexual signal of this species in the presence of both males and females. In trace *a* the injected flash of a male competitor is seen in the two-peaked (overload) pulse midway between the twin flashes; this is a typical injected flash. In trace *b* the competitor synchronized its injected flash with the author's second flash, producing a ragged (overload) pulse; the single flash at the right is the female's response. In trace *c* the competitor did not inject the two long bright flashes at the right until after the author's signal had elicited a female response. In trace *d* the competitor similarly delayed its single flash, a two-peaked (overload) pulse similar to trace *a*.

The genus *Photuris* has had a reputation for presenting taxonomic difficulties and even downright impossibilities. The reason is that its confusing morphological variations make sorting out the different members of the genus into satisfying "cabinet species" a hopeless task. In 1951, however, on the basis of work done nearly 30 years earlier, descriptions of the species-specific flash patterns of the genus compiled by H. S. Barber of the U.S. Department of Agriculture were published posthumously. Barber proposed a number of new *Photuris* species. Taking a leaf from Barber's book, I have similarly found new species in the comparatively simple, well-known genus *Photinus* by analyzing their mating signals. As a result of Barber's studies and others, investigators of fireflies have been led to a general rule: one habitat, one time, one signal, one species. To borrow from the language of physics, two species of fireflies orbiting in the same habitat at the same time must have different signal "spins."

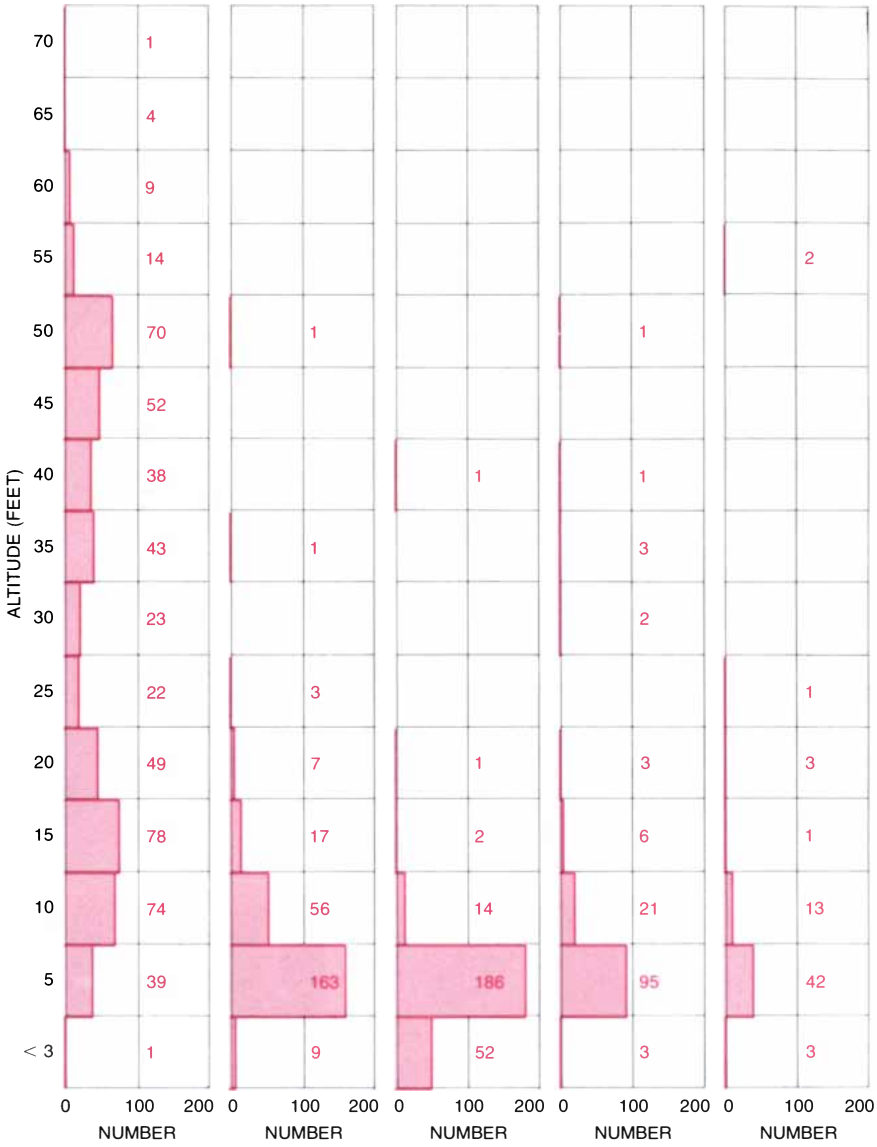
The fireflies of the genus *Photuris* dramatically break this rule. Not only do the predaceous females emit patterns like those of other firefly species that share habitats with them but also the males are mimics. For example, in northwestern peninsular Florida three species of uncommon fireflies share the cool evenings of very early spring, the pinelands and the prairies and their sexual signals. *Photuris* "VR" (Variable Red) males fly at high altitudes, particu-

larly around the tops of pines, emitting their slow, pulsing pattern. *Pyractomena barberi* males fly near the ground, emitting their ragged flicker. *Pyractomena angustata* males usually fly with *P. barberi* or a little higher, emitting their own long-sustained glow. When *Photuris* "VR" males descend to the airspace of *Pyractomena angustata* and *P. barberi,* however, they change their signals and begin to luminesce like the alien species.

The fact that the *P. angustata* signal is a glow rather than a flash is remarkable in itself; glowers almost invariably belong to behaviorally primitive firefly genera, and all *P. angustata*'s congeners are behaviorally advanced flashers or flickerers. Along with this unlikely circumstance anyone's curiosity must be aroused by the fact that *Photuris* "VR," the scion of another flashing lineage, can glow like *Pyractomena angustata* and flicker like *P. barberi* and engages in both mimicries not only in the *Pyractomena* arena of activity but also in season with these other species. It is certainly more than coincidence.

Males of other *Photuris* species engage in similar mimicry. *Photuris* "AR" (Apalachicola Red) normally signals with single flashes in the tops of pines, but it glows like *Pyractomena angustata* when that species is present. Once *P. angustata*'s season has passed or when *Photuris* "AR" flies outside *Pyractomena angustata*'s geographic range, the *Photuris* species no longer glows.

Perhaps most notable for its versatility is *Photuris cinctipennis.* Although the studies are not complete, it appears that the male of this species, in addition to emitting its own spectacular and unique signal pattern (which takes the form of chains of swooping, rapid crescendos), can emit four other signal patterns.



VERSATILE MIMIC, the unnamed *Photuris* species "VR" (Variable Red) male, usually flies between 10 and 50 feet above the ground as it emits its pulsing sexual signal. The column at the left shows the altitude of 517 "VR" fireflies emitting pulsing signals. The males' flight pattern changes greatly, however, and their signal becomes a flicker when *Pyractomena barberi,* a low-flying firefly with a flickering signal is present. The second column shows the altitude of 257 flickering "VR" fireflies; the third column shows the altitude of 256 *P. barberi* males. When males of the low-flying species *Pyractomena angustata,* which emit a glow, are present, the "VR" fireflies also emit a glow and fly low. The fourth column shows the altitude of 135 glowing "VR" males; the last column shows the altitude of 65 *P. angustata* fireflies.

What could possibly give rise to such evolutionary convergence? As Barber wrote when he began to suspect (correctly) that *Photuris lucicrescens* males may emit either long crescendos or short flashes: "Various answers will satisfy various persons, but no one knows." Students of warning coloration might invoke a similar explanation for the convergence. Those concerned with signal propagation might argue that particular forms of signals are more easily propagated in certain habitats. I have weighed the 15 or more explanations for convergence that have entered my mind and have tentatively excluded most of them as being either unlikely or overly complicated.

Perhaps the answer lies in the fact that at any one time *Photuris* males greatly outnumber the sexually active females of their own species and compete vigorously to obtain mates. I suspect that in most instances the signal-mimicking males are hunting their own females while at the same time the females are busy preying on the fireflies the *Photuris* males are mimicking. Simply put, the predators have become the prey, although the males' motive is sex rather than food.

The *Photuris* female, however, is bigger and more powerful than the male. What can the male do to mate with such a formidable predator? My best guess is that in some species, such as *Photuris* "VR," when a male has located a hunting female, he is sometimes able to shift her interest from food to mating. In effect the males "seduce" the females by accelerating a change in behavior that would have come eventually. On several occasions I have noted that when a flickering or glowing "VR" male lands near a flashing light on the ground, he next emits his own species-specific treetop signal pattern. This fits a process of seduction rather than one of rape. Whether this explanation or some other one eventually proves to be correct, one fact is clear. The interactions of American fireflies are far more complex than anyone could have predicted. Are fireflies different or are the interactions of other insects also this complex?

AS LUXURY SEDANS BECOME MORE EXPENSIVE, THEIR INHERENT WORTH BECOMES MORE CRITICAL.

Due to a variety of economic ills, the day has finally arrived when the price of a luxury automobile can rival that of a small house in the country.

A disturbing trend that has forced many otherwise casual automobile buyers to scrutinize their potential purchases as never before.

At BMW we are not alarmed by this newly enlightened consumer.

For when one examines the features that truly matter in a luxury sedan—performance, craftsmanship and resale value—no other expensive automobile justifies its price quite so thoroughly as the BMW 733i.

Accounting, no doubt, for the fact that last year the demand for the BMW 733i outpaced even our most optimistic predictions.

EVOLUTION IS PREFERABLE TO REVOLUTION.

With a certain predictability, automakers reacted to the realities of the 1980's by bringing forth a plethora of "revolutionary" re-engineered cars.

The 733i, however, is hardly the result of a single year's rush to build an enlightened automobile.

Indeed, decades ago, when luxury car makers were building ever larger cars, reckless in their consumption of fuel, BMW was championing such avant-garde concepts as six-cylinder engines, sensible piston displacements and impressive power-to-weight ratios.

The BMW 733i's 3.2-liter, fuel-injected power plant, for example, is the product of millions of miles of testing and refinement—on and off the great racecourses of the world.

Its double-pivot suspension system has been described by Car and Driver magazine as "...the single most significant breakthrough in front suspension design in this decade."

A five-speed standard transmission (automatic is available) affords one the unique opportunity to pick and choose one's own gears.

ELECTRONIC INNOVATION THAT GOES BEYOND A DIGITAL SPEEDOMETER.

On the BMW 733i, our electronic computerized monitoring system (first installed by BMW in 1975) plays a far more important role than mere decoration: it actually improves the car's efficiency.

With every engine revolution a computer receives and assesses signals from sensors deep within the 6-cylinder engine. Then instantly determines the precise air/fuel mixture to be injected into the cylinder ports.

Yet the incongruous note here is that BMW efficiency is not achieved at the expense of the sort of exhilarating performance one expects in a BMW—or the power reserves necessary to maneuver safely even under the most demanding conditions.*

LUXURY. NOT SUPERFICIALITY.

It would be difficult to imagine an appointment or an accessory that has been omitted in the BMW 733i.

Yet, all facets have been biomechanically engineered to achieve the perfect integration of man and machine.

Its seats are anatomically correct buckets and covered in wide rolls of supple leather.

All vital controls are within easy reach of the driver. Instruments are large, well-marked and totally visible.

So impressive is the total result that one automotive journalist was moved to write of the 733i, "...to drive it is to know all the wonderful things machines can do for man."

To arrange a thorough test drive, simply phone your nearest BMW dealer.

THE ULTIMATE DRIVING MACHINE.

BMW. MUNICH, GERMANY

*The 733i (with standard transmission) delivers 16 EPA estimated mpg, 25 estimated highway mileage and, based on these figures, an estimated mpg range of 360 miles and a highway range of 562 miles. (Naturally, our fuel efficiency figures are for comparison purposes only. Your actual mileage and range may vary, depending on speed, weather and trip length. Your actual highway mileage and highway range will most likely be lower.) © 1981 BMW of North America, Inc. The BMW trademark and logo are registered trademarks of Bayerische Motoren Werke, A.G.

Binocular Depth Inversion

Sometimes a solid object seen with both eyes can seem to reverse perspective. A study of this geometrically irrational experience suggests that ordinary depth perception is somewhat precarious

by John I. Yellott, Jr.

A visitor to the "Haunted Mansion" at Disneyland in California sees among other things a pair of human faces that appear to rotate in a mysterious and sinister way as he walks by. They are in fact inside-out relief masks, but because of the lighting and the way they are mounted the visitor unwittingly reverses their depth, perceiving them incorrectly as normal faces. This unconscious reversal of perspective gives rise to the apparent rotation.

Besides mystifying visitors to the "Haunted Mansion" this illusion presents a problem for theories on the perception of visual form. The problem is not the apparent rotation of the faces; psychologists have known for some time that whenever a three-dimensional object is perceived in reverse perspective, it will seem to rotate as the observer's head moves. What is puzzling is the perspective reversal itself. Ordinarily people see the three-dimensional forms of things correctly, and reversals of perspective occur only in special circumstances that deny the brain its normal visual cues to depth. One such circumstance is viewing the three-dimensional form with one eye closed, so that the depth cues provided by binocular vision are eliminated. The masks at Disneyland, however, show that sometimes objects are routinely perceived inside out in spite of the availability of all the normal depth cues, including those due to binocular vision. How can such a major perceptual mistake occur? And given that it does happen sometimes, why is it so rare?

In this article I draw a distinction between the kind of perspective reversals produced by ambiguous pictures such as the famous Necker cube and those experienced in viewing actual three-dimensional objects, such as the masks in the "Haunted Mansion." I shall refer to the latter type of illusion as "depth inversion." Reversible-perspective pictures and their perceptual consequences are quite well known; drawings such as the Necker cube have illustrated countless psychology textbooks since the 19th century, and artists have explored the same theme for much longer.

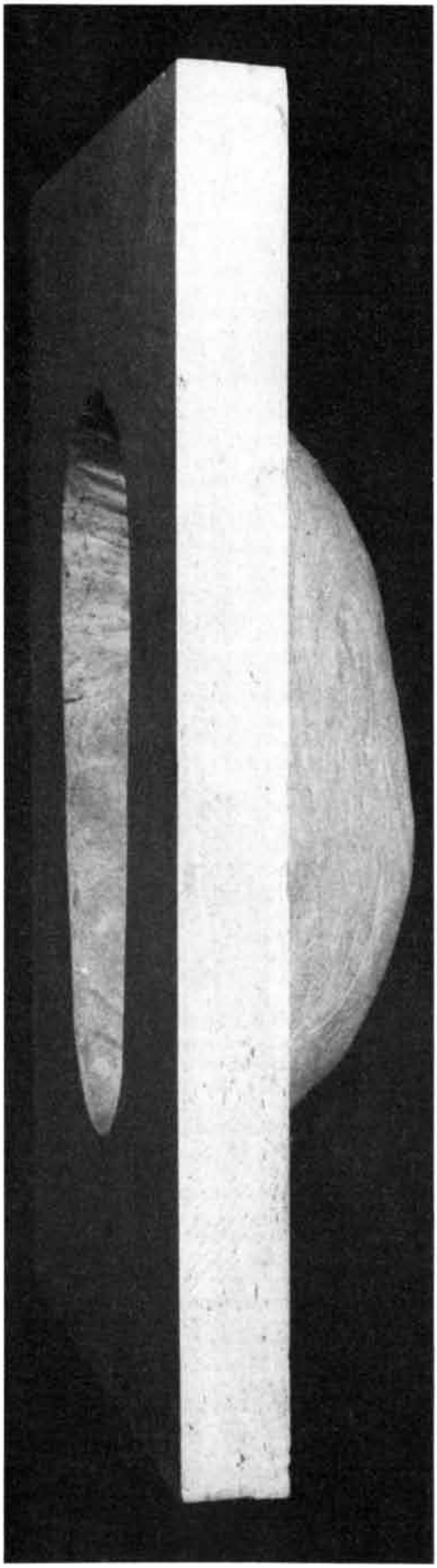
Depth inversion of solid objects also has a long scientific history. References to the phenomenon date from the 18th century, and in the 19th century it was studied by such notable figures as Hermann von Helmholtz and Ernst Mach. In this century, however, it seems to have been neglected until 1970, when the British psychologist Richard L. Gregory drew attention to it again in his book *The Intelligent Eye*. Gregory's discussion stimulated my interest and led me to devise several experiments to determine whether objects can be seen in reverse perspective when the brain is truly in full possession of all the depth information available in normal vision, including in particular the information provided by binocular vision, which according to classical accounts should make the illusion impossible. The results of these experiments indicate that under appropriate conditions the brain is prepared to override all its sensory cues to depth and create an inside-out visual world that defies geometrical analysis but nonetheless seems just as realistic as normal visual experience.

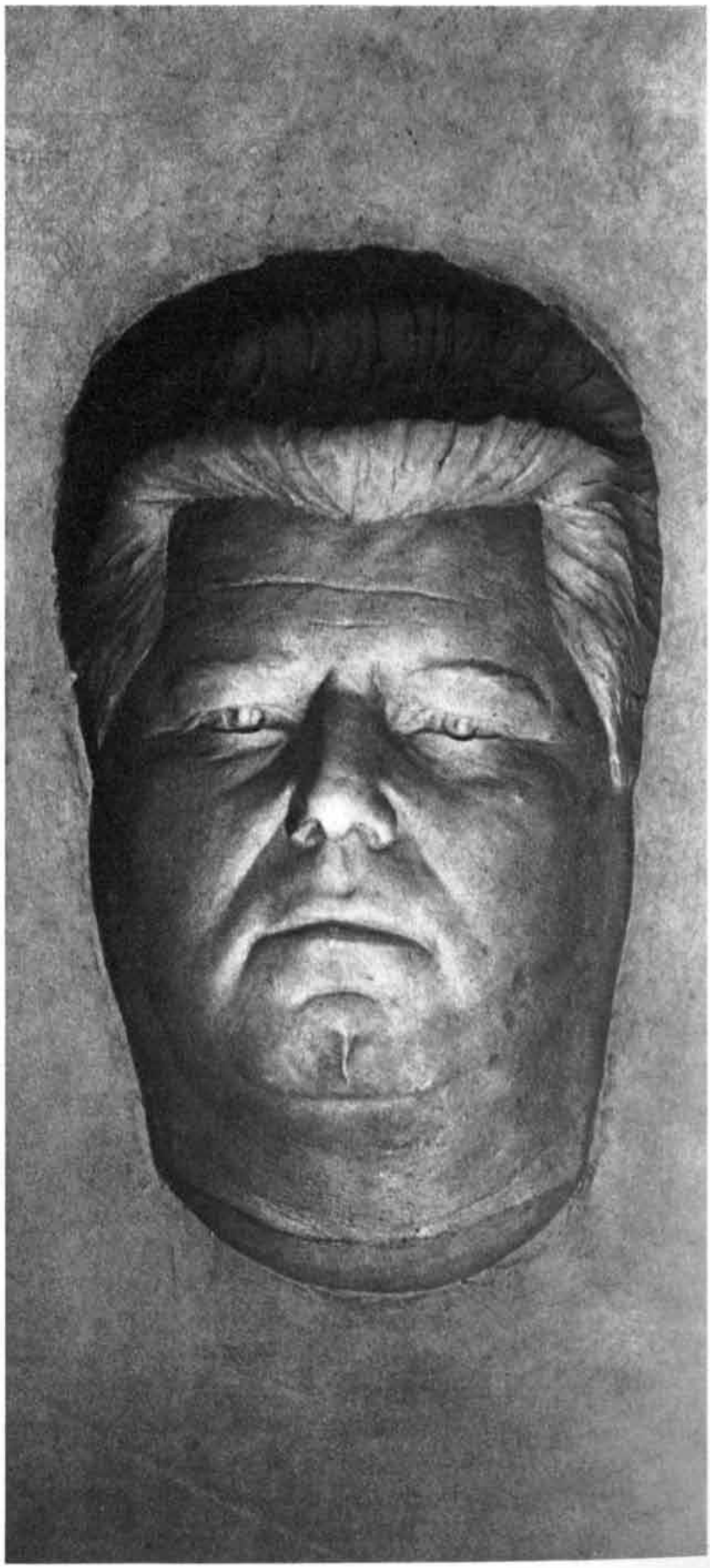
These experiments on binocular depth inversion are the subject of this article, but to put them in context it will be helpful to first consider monocular inversion, which is much easier to explain. Depth inversion of a three-dimensional object viewed with one eye can be understood if one thinks of visual experience as the outcome of a process in which the brain tests hypotheses about the three-dimensional shapes of objects against the evidence provided by their retinal images. With one eye alone the only potential source of unequivocal information about depth is accommodation, or change of focus, and the brain normally gives this cue little or no weight in its judgment of distance. Accommodation therefore presents no barrier to the acceptance of an inside-out shape as being real. All the other monocular cues to depth are intrinsically ambiguous. The evidence they provide cannot exclude inside-out hypotheses, although they can render such hypotheses statistically unlikely in the sense that, say, a tree rotating in synchrony with movements of the head is an improbable object.

Apparently this is normally enough to enable the brain to guess correctly about the shapes of things seen monocularly. If sensory evidence becomes sufficiently impoverished, however, the brain may accept an inside-out hypothesis that is compatible with the retinal image. In such a case visual experience is totally transformed to agree with the hypothesis, intellectual knowledge of the correct form notwithstanding. Yet the inverted object now seen still incorporates all the information available on the retina, just as in normal vision. The only difference is that now every depth cue is visually reinterpreted in order to agree with a false premise.

Now consider the situation in binocular vision. The key to my explanation of monocular inversion (an explanation borrowed from Helmholtz, Mach and Gregory) is the fact that all the monocular cues to depth can be consistently reconciled with an inverted-object hypothesis. When both eyes view an object, however, no such reconciliation is possible. Binocular vision provides depth information that is geometrically incompatible with depth inversion, in other words information that should enable

INSIDE-OUT FACE, made as the mold of a bust, is shown in side and front views on the opposite page. Looked at from the front it is more easily seen as a normal face because the brain overrides the depth cues that suggest an object as improbable as an inside-out face. (The reversal is made easier when, as in the front view here, the lighting eliminates shadows that might aid the brain in making the correct interpretation.) A three-dimensional inside-out face seen in reversed perspective seems to rotate and to follow an observer who is moving laterally past it.





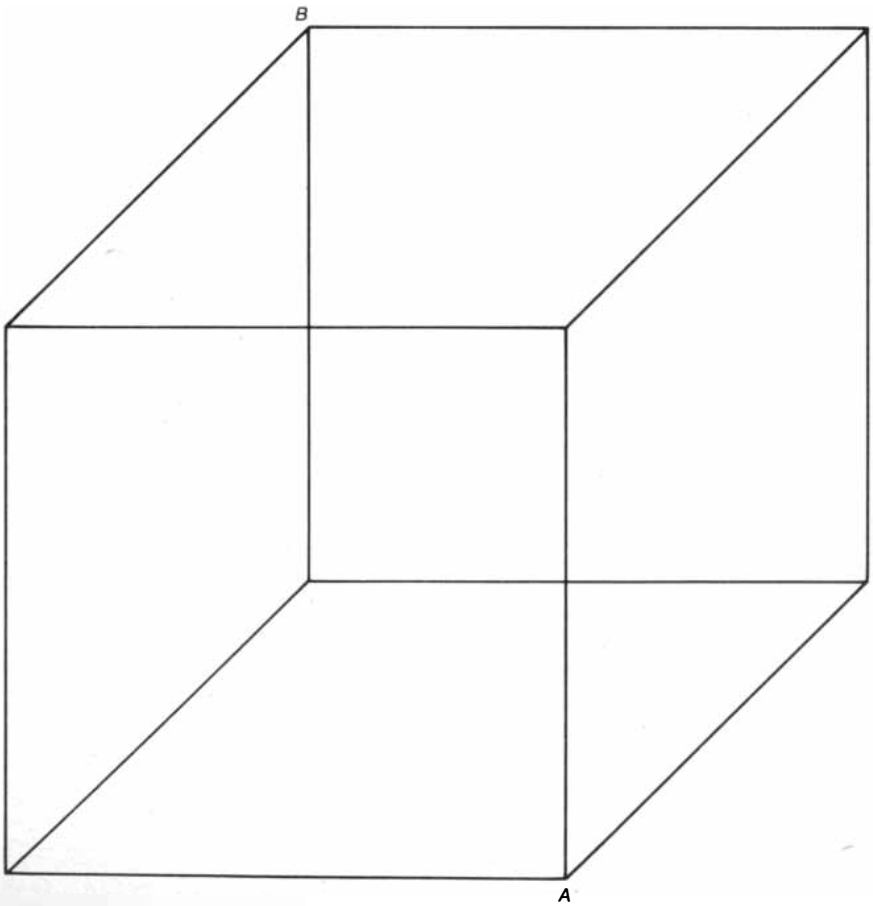
the brain to categorically reject inverted-object hypotheses.

That information stems from the difference between simultaneous retinal images of an object in the left and right eyes, a depth cue termed binocular disparity. Its effect is that none of the inverted-depth hypotheses consistent with the left eye's view can simultaneously be consistent with the right eye's view. If the key to depth inversion is geometrical compatibility between retinal evidence and inverted-object hypotheses, binocular depth inversion should be impossible, or at least a most unnatural visual experience, quite unlike monocular inversion.

Binocular vision actually provides not one new cue to depth but two cues. One is the muscular cue produced by the act of convergence, that is, the action of the eye muscles in aiming both eyes at a common fixation point. This action gives the brain information on the convergence angle of the line from each eye to the fixation point, and the angle gives a cue to the distance of that point. At a given instant, however, this muscular cue does not provide any information on the depth of other points that are not being fixated.

That information is supplied by the second binocular cue to depth: binocular disparity. When the eyes converge on a point, the images of the point fall on corresponding places on the two retinas, namely the center of each fovea (the small area that affords the sharpest vision). Points nearer or farther than the fixation point necessarily fall on noncorresponding places on the two retinas. The magnitude of this positional disparity is conventionally measured in angular units. The binocular disparity of any nontargeted point X is the difference between the convergence angle and the angle formed by the lines of sight to X. This angular measure is proportional to the absolute distances between the retinal locations of the two images of the point; one minute of binocular disparity corresponds to a six-micrometer difference in the retinal positions.

Although convergence is a better cue to depth than accommodation, it still provides rather uncertain distance information. Binocular disparity, however, is an extremely powerful cue to depth. Under experimental conditions normal observers can detect depth differences that give rise to disparities of about 10 seconds of arc, equivalent to a one-micrometer difference in retinal position. In other words, the brain can reliably detect disparities that are substantially smaller than the diameter of the smallest photoreceptors (about two micrometers). Cues furnished by binocular disparity therefore seem sufficient in principle to rule out hypotheses of depth-inverted objects.



NECKER CUBE is a reversible-perspective drawing named for the Swiss naturalist Louis Necker, who in 1832 described the perceptual consequences of ambiguous perspective in pictures. The perspective of the cube tends to reverse back and forth as one stares at the picture.

The foregoing arguments make it plausible a priori that binocular depth inversion should not occur because the brain cannot construct an inverted visual model consistent with all its retinal evidence. Both Helmholtz and Mach apparently believed binocular depth inversion is impossible. Gregory, studying the binocular inversions of a three-dimensional wire cube, noted that they are rare and brief, and that when an inversion does occur, the cube looks unnatural. Historically it seems to have been generally accepted that binocular depth inversion simply does not happen, at least in any stable way compared with monocular inversion.

It is easy to show that things are not so simple. Under appropriate conditions binocular depth inversion can occur quite easily, yielding a stable perception much like the one resulting from monocular inversion in spite of binocular-disparity cues that would be detected readily in normal vision. The trick is to use an object with an overwhelmingly improbable real form, so that it looks normal only when it is seen inverted in depth. The best example is an inside-out human face like the two in Disneyland. Such a face is the mold of a normal relief. The inside of an ordinary Halloween mask will also do.

With a little practice one can easily achieve stable binocular depth inversions of such a face at a viewing distance of about an arm's length. An excellent stimulus is a plastic mask mounted inside out on a sheet of cardboard and illuminated from behind. This arrangement eliminates informative shadows that can slow down inversion.

With such a setup monocular inversion is easy. At first opening the other eye tends to disrupt a monocularly stable inversion, just as movements of the head initially disrupt monocular inversions achieved with the head stationary. With practice, however, one learns to tolerate the new depth information provided by binocular vision, and the perceptual result is a depth-inverted face that appears to be natural and stable.

Trying the illusion for the first time, observers often find that the surface of the inside-out face does not invert all at once. Instead inversion begins in one region, typically the nose, and then other regions gradually become incorporated into the inverted percept. Thus at early stages one may find that during a movement of one's own head the inverted nose will seem to wobble on an otherwise immobile face.

The fact that binocular inversion

can occur with an inside-out face is not entirely surprising. One can simply say that the brain is prepared to override even unequivocal sensory evidence when the evidence supports a highly improbable object hypothesis. Conversely, when past experience is compatible with both an inverted and a noninverted version of an object, binocular-disparity cues tip the scales in favor of the correct hypothesis. This explains why objects such as three-dimensional wire cubes are easy to invert monocularly but difficult to invert binocularly.

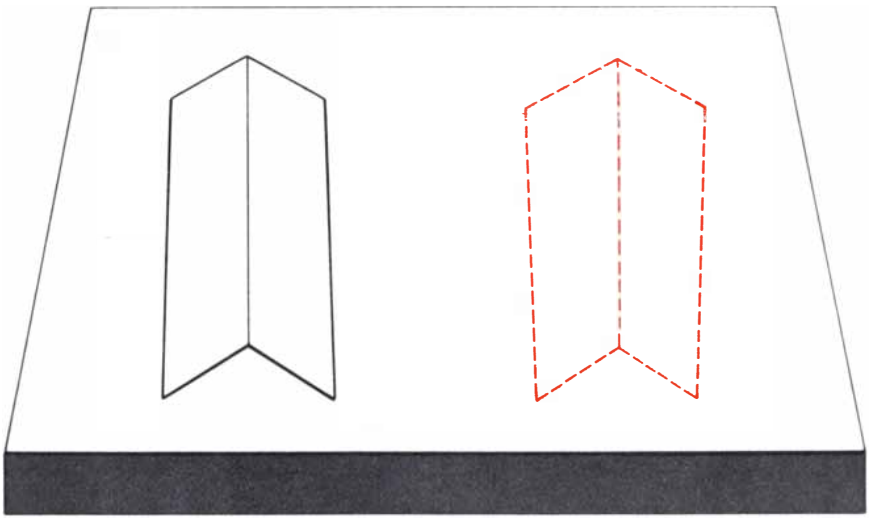
When binocular inversion does occur, however, there still remains the critical question of what happens to the binocular-disparity information that should, if it is properly incorporated into perception, prevent inversion altogether. To put the issue another way, how can the retinal images in the two eyes be combined into a single three-dimensional visual experience when that experience cannot be geometrically reconciled with both images simultaneously?

Two answers suggest themselves immediately. The first and simplest one is that a binocular inversion is not truly binocular. Even though both eyes view the object, perhaps the information from only one eye is incorporated into visual experience. This would mean that information from the other eye is suppressed at some preconscious level.

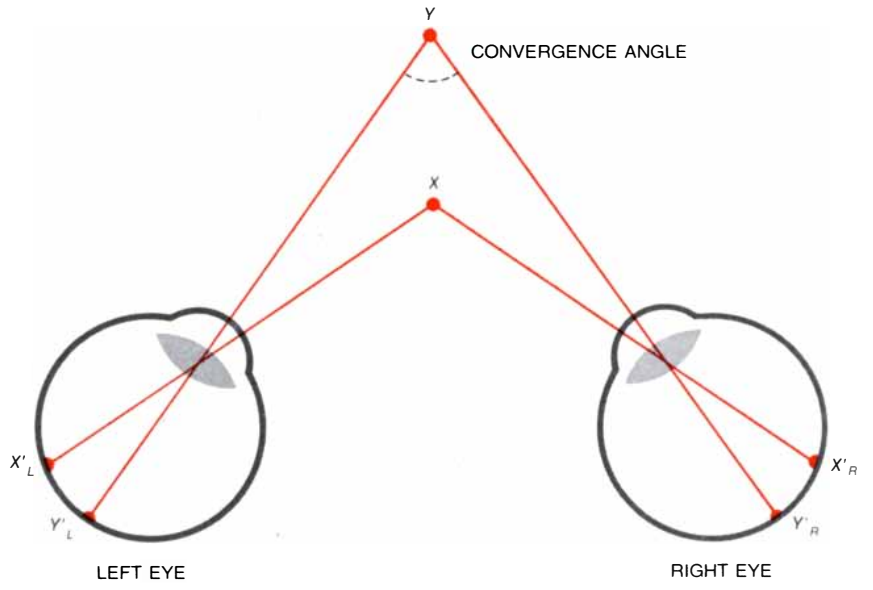
Such suppression occurs regularly in normal vision when the two eyes are exposed to quite different stimuli. The phenomenon is known as binocular rivalry. It is easily demonstrated. With both eyes open hold your right hand about six inches in front of your right eye and look across the room at, say, a lighted lamp, making sure the lamp is visible to the left eye but not to the right eye. Closing your left eye, you see your hand, and closing your right eye, you see the lamp: two irreconcilable views of the same region of visual space. Yet when both eyes are open, you see only the lamp. Your hand is suppressed, at least in the region of the visual field where the two stimuli are in conflict. Indeed, you can see the lamp "through" your hand.

Since this kind of suppression of one eye's view in favor of the other's is routine in normal vision, one might suppose it could account for binocular depth inversion. If the brain discards the information from one eye, it is then free to construct an inverted object that is entirely consistent with information from the other eye. On this hypothesis binocular inversion would be only monocular inversion coupled with suppression of the information from one eye.

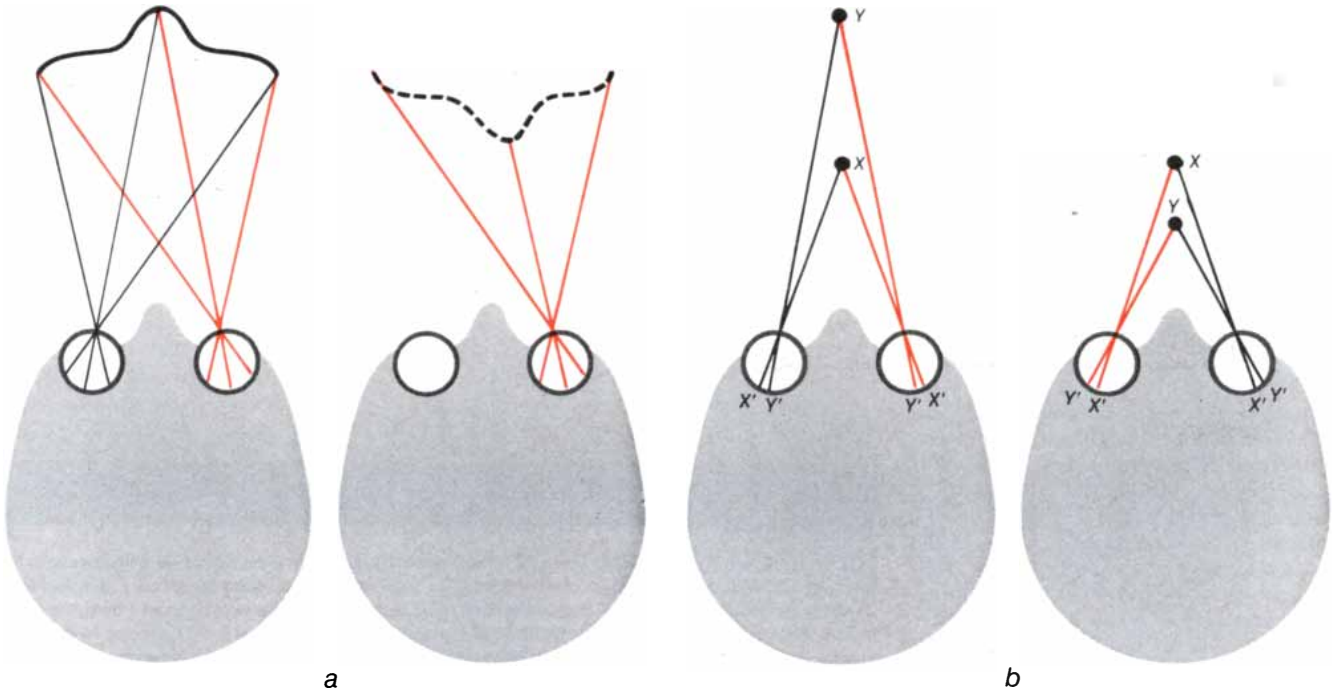
The other answer I thought of originally was that binocular inversion might be truly binocular in the sense that the information from both eyes is incorporated into visual experience but with the signs of all the binocular-disparity cues reversed, as though the brain had lost track of which eye is which and had interpreted the retinal image in the left eye as coming from the right (and vice versa). In experiments that present flashes of light randomly to one eye or the other, normal observers often have great difficulty telling which eye has been stimulated. Moreover, in normal vision involving binocular rivalry one is not consciously aware of which eye sees what. (For example, hold a finger a few inches above this page so that some letters are invisible to one eye and some to the other. With both eyes open you can read every letter, but unless you alternately close one eye and then the other you will not be able to tell whether a given letter is seen by the left eye or the right eye.) Hence it seemed possible that the brain might exploit this condition in order to reconcile an overwhelmingly plausible object hypothesis with all the sensory evidence. Such an explanation would at least make binocular inversion a more or less direct extension of monocular inversion. Visual experience would still incorporate all the depth



MONOCULAR INVERSION is readily achieved with an experiment devised by Ernst Mach. A blank card such as an index card is folded in half lengthwise and placed on a flat surface with the bend upward (*left*). If one looks at it long enough with one eye, trying to imagine that it is in the configuration of an upright book open for reading (*right*), it will appear to reverse and be turned inside out, so that its bend points down and the card seems to perch on one end.



GEOMETRY OF BINOCULAR VISION suggests that depth inversion should be difficult or impossible. Here both eyes view points *X* and *Y*, with *X* the nearer one; the respective retinal positions are designated X'_L and so on. If the eyes converge on *Y*, the images of *Y* fall on the center of the fovea of each retina. The lines of sight projected outward from the images X'_L and X'_R meet at a unique point in space: the real location of *X*. Hence the combined retinal images in the eyes cannot be reconciled with hypothesis that *X* is more distant than *Y*.



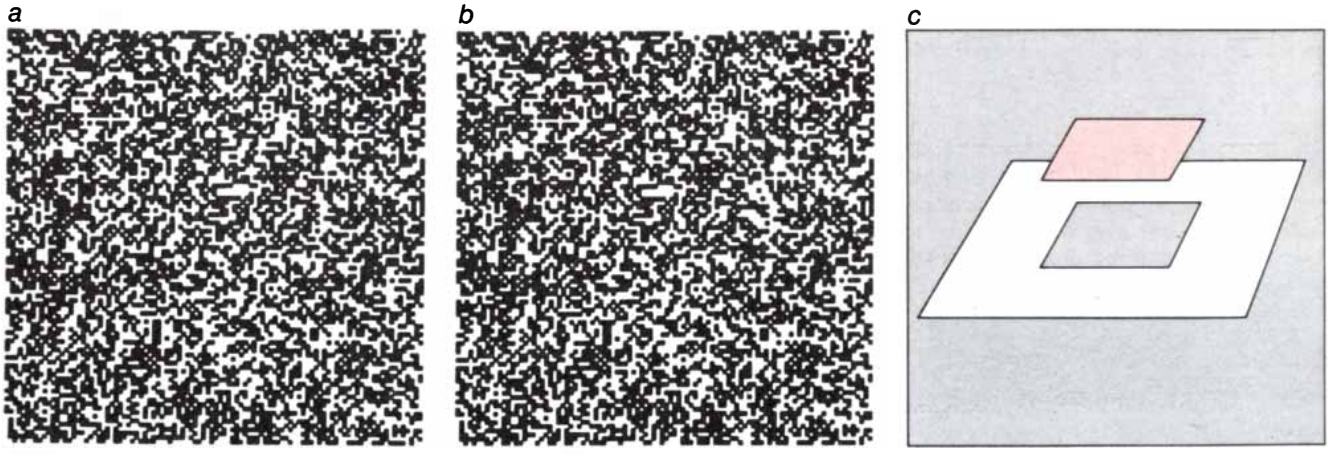
TWO HYPOTHESES on binocular depth inversion are monocular suppression (*a*) and disparity reversal (*b*). In each case what is actually seen is portrayed at the left and the inversion the hypothesis would explain is shown at the right. In monocular suppression the left eye's view is depicted as being suppressed, so that the inside-out face is seen depth-inverted as it would appear in a monocular inversion when it is viewed by the right eye alone. In disparity reversal point *X* is really nearer the observer than point *Y*, but the brain treats the retinal image in the right eye as though it came from the left eye and vice versa, with the result that point *Y* appears to be closer than point *X*.

cues available to the brain, but one cue, namely binocular disparity, would appear to be transformed in an unjustifiable way.

These two potential explanations of binocular depth inversion can be termed "monocular suppression" and "disparity reversal." I shall describe two easily reproducible experiments, each of which tests both hypotheses simultaneously. Both experiments lead to the same conclusion: neither monocular suppression nor disparity reversal can account for what one sees during binocular inversion.

The first experiment makes use of a fascinating class of stimuli known as random-dot stereograms, which were invented in 1959 by Bela Julesz of Bell Laboratories. (I did the experiment in my laboratory at the University of California at Irvine in collaboration with Jerry Kaiwi, who was then a graduate student.) A stereogram is a pair of pictures designed to create a sensation of depth when one picture is viewed by the left eye and the other is viewed simultaneously by the right eye. The sense of depth is elicited by discrepancies between the left and right pictures that simulate the binocular disparities a solid object would generate. The process of perceiving depth on the basis of binocular-disparity cues is known as stereopsis.

Random-dot stereograms provide a definitive test of whether the viewer is achieving stereopsis. Not everyone can; about 2 percent of the population is "stereo blind." To make such a stereogram one constructs a pair of identical pictures consisting of randomly scattered dots. Then all the dots in a given region are shifted slightly to the left in one picture and slightly to the right in the other to create a binocular disparity



RANDOM-DOT STEREOGRAM provides a basis for testing the monocular-suppression and disparity-reversal hypotheses. The left and right halves of the stereogram (*a*, *b*) are identical except that the dots in a square region in the center are shifted horizontally (to the right in the left-hand picture and to the left in the right-hand picture) to create a binocular disparity. An observer with normal stereoscopic vision, viewing the pictures in a stereoscope or by some other means that presents the pictures separately to the left and right eyes, perceives the central square as floating above the background (*c*). Random-dot stereograms were made by Bela Julesz of Bell Laboratories.

Artwork created by Frank van Herck on a Scitex Response-300.
Use of original document by courtesy of Pre-Press Group DeSchutter BV, Antwerp, Belgium.

This man has forgotten his computer system.

Deep at work as he is, the man behind this hand is oblivious to everything but his interactive artwork. His medium is Scitex's Response system, the first to prepare images by computer for color printing plates. The operator thinks of it as a movable window which lets his stylus through to paint and reshape, alter contrasts and colors, and generate color pages with text. For the printing industry worldwide, Response systems give better color and detail, in less time, than conceivable before.

This advertisement developed, like others in this magazine, as 20 megabytes in the memory of a Response system.

In America, Europe, and Japan, the talent to see through equipment into processes is common to operators of the Response. At Scitex in Israel, where Response systems are developed, the talent to see through processes into systems is the common factor. Here north of Tel Aviv, scientists whose backgrounds range from electro-optics to real-time software, and span many countries, have united to win leadership in color image processing with a product so responsive it can seem transparent.

For more information:
Scitex Corporation — POB 330,
Herzlia B 46 103, Israel

scitex

Physician, did you miss any of these significant developments in medical science?

• ***Campylobacter fetus* subsp. *jejuni* is associated with a colitis that can clinically and sigmoidoscopically resemble acute idiopathic ulcerative colitis. Stool cultures are in order for *C. fetus* before beginning nonspecific anti-inflammatory therapy.**

• **Coumarin derivatives cross the placenta. A recent study shows that the consequences for the fetus can be severe. These include embryopathy, stillbirth, and premature delivery.**

• **Nonsteroidal anti-inflammatory drugs may produce a marked reduction in glomerular filtration rate; with termination of the drug, GFR returns to normal.**

• **Pittsburgh pneumonia agent (PPA) and atypical *Legionella*-like organisms (ALLO) are particularly dangerous in immunosuppressed patients.**

IF THESE ITEMS are familiar you must be a prodigiously energetic or prodigiously lucky reader. With 2,000 or more journals published each year, information that significantly affects patient management all too easily slips by. Textbooks are out-of-date before they are published.

SCIENTIFIC AMERICAN *Medicine* is the busy clinician's answer to this problem.

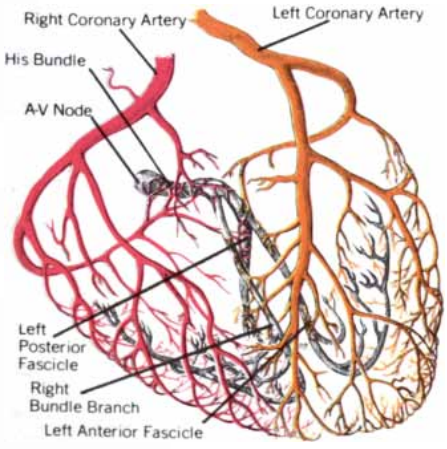
Because its authors update SCIENTIFIC AMERICAN *Medicine* every month, it is always current. Because the new information appears in a single source, it is there when you need it.

This 2,000-page, innovative union of publishing and electronic technology is the work of leading scholar-practitioners from Harvard and Stanford. The editors are Edward Rubenstein, M.D., F.A.C.P., and Daniel Federman, M.D., F.A.C.P.

Each month as authors update their contributions, revisions are entered on the magnetic tape on which the text and index are stored. The tape drives high-speed phototypesetting equipment so that subscribers receive about eight new chapters and a new index every four weeks; a bulletin highlights new developments.

New material replaces old material in the living text, so that the information is there – up-to-date, at your fingertips.

A CME program of eight patient management problems offered over a 12-month period is available at no extra cost. As an organization accredited for continuing medical education, the Stanford University School of Medicine designates this continuing medical education activity as meeting the criteria for 32 credit hours in Category 1 for Educational Materials for the Physician's Recognition Award of the American Medical Association, provided it has been completed according to instructions. This program is approved by the American Academy of Family Physicians for 32 Elective Hours of CME credit.



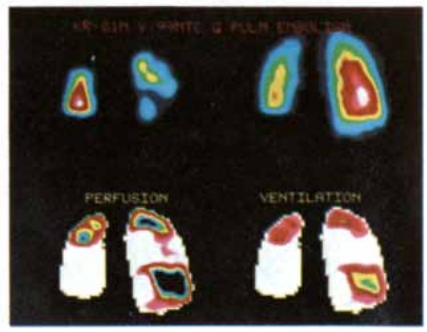
Branches of the right and left coronary arteries supply blood to the AV node and intraventricular conduction system.

SCIENTIFIC AMERICAN *Medicine* is lucidly illustrated with drawings and photographs. Some examples are seen here and on the facing page.

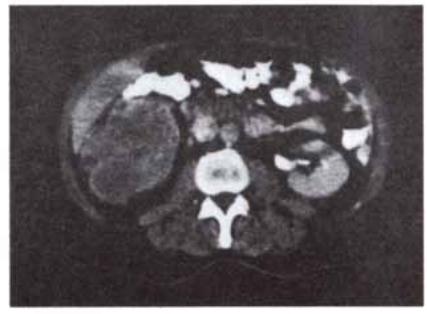
Trial Offer

We invite you to try SCIENTIFIC AMERICAN *Medicine* — for two months at no cost. Send us the coupon and you will receive the two-volume text and two monthly updates. You may also take a CME test for credit. At the end of 60 days, if you decide to continue the subscription, we will bill you for $185 for the full 12 months (renewal is $160); otherwise just return the volumes to us.

Please mail the coupon today and let us take the hassle out of keeping up.



Computerized scintigraphy reveals pulmonary thromboembolism.



Abdominal computed tomogram reveals large renal carcinoma replacing part of right kidney.

THE DISTINGUISHED AUTHORS AND THE FIFTEEN SECTIONS OF SCIENTIFIC AMERICAN ***MEDICINE***

1. Cardiovascular Medicine

Edgar Haber, M.D., Harvard Medical School and Massachusetts General Hospital

E. William Hancock, M.D., F.A.C.P., Stanford University School of Medicine

Roman W. DeSanctis, M.D., Harvard Medical School and Massachusetts General Hospital

Adolph M. Hutter, Jr., M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Eve Elizabeth Slater, M.D., Harvard Medical School and Massachusetts General Hospital

2. Dermatology

Eugene M. Farber, M.D., Stanford University School of Medicine

Elizabeth A. Abel, M.D., Stanford University School of Medicine

3. Endocrinology

Daniel D. Federman, M.D., F.A.C.P., Harvard Medical School

4. Gastroenterology

Gary M. Gray, M.D., Stanford University School of Medicine

Peter B. Gregory, M.D., Stanford University School of Medicine

John Austin Collins, M.D., Stanford University School of Medicine

5. Hematology

Stanley L. Schrier, M.D., Stanford University School of Medicine

6. Immunology

John David, M.D., Harvard Medical School and Robert B. Brigham Hospital

7. Infectious Disease

Thomas C. Merigan, M.D., Stanford University School of Medicine

Morton N. Swartz, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Cyrus C. Hopkins, M.D., Harvard Medical School and Massachusetts General Hospital

Adolf W. Karchmer, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Robert H. Rubin, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Harvey B. Simon, M.D., F.A.C.P., Harvard Medical School and Massachusetts General Hospital

Peter F. Weller, M.D., Harvard Medical School

8. Intensive and Emergency Care

Edward Rubenstein, M.D., F.A.C.P., Stanford University School of Medicine

9. Metabolism

George F. Cahill, Jr., M.D., Harvard Medical School, Howard Hughes Medical Institute, and Peter Bent Brigham Hospital

10. Nephrology

Roy H. Maffly, M.D., Stanford University School of Medicine and Palo Alto Veterans Administration Medical Center

11. Neurology

Robert W. P. Cutler, M.D., Stanford University School of Medicine

12. Oncology

Saul A. Rosenberg, M.D., F.A.C.P., Stanford University School of Medicine

13. Psychiatry

Ned H. Cassem, M.D., Harvard Medical School and Massachusetts General Hospital

14. Respiratory Medicine

Eugene D. Robin, M.D., F.A.C.P., Stanford University School of Medicine

15. Rheumatology

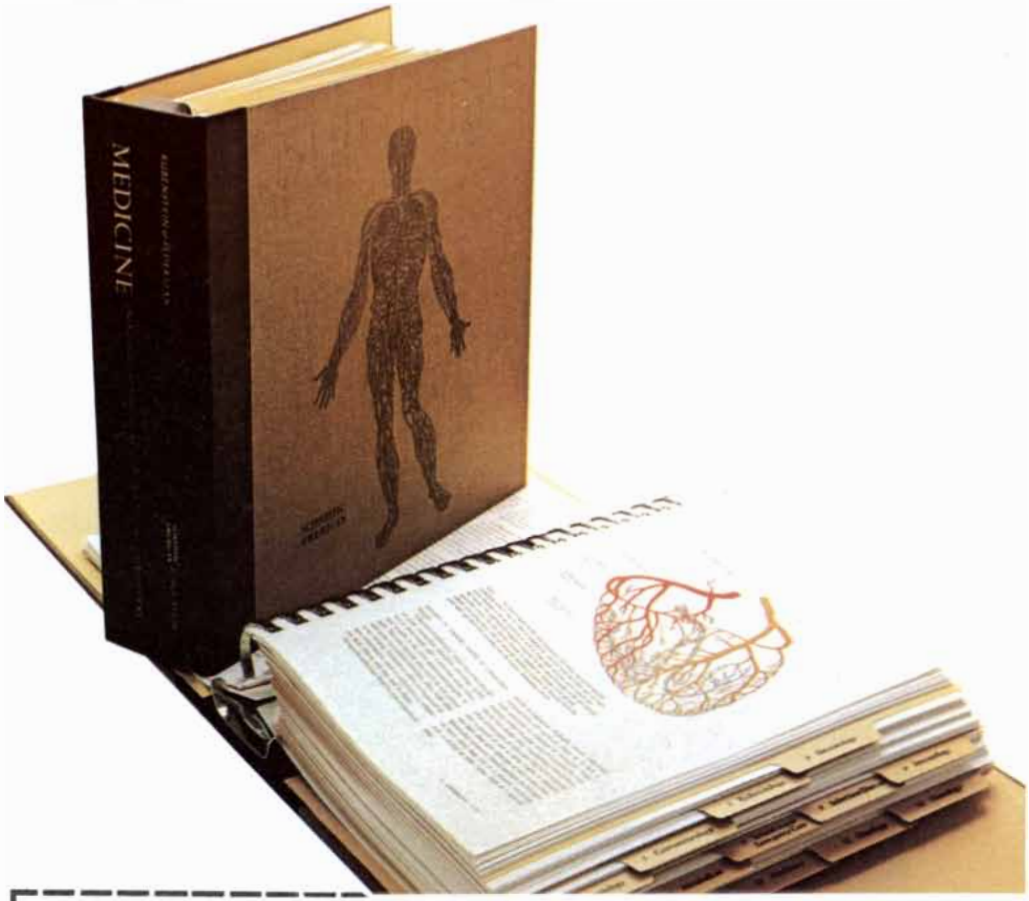
Stephen M. Krane, M.D., Harvard Medical School and Massachusetts General Hospital

Dwight R. Robinson, M.D., Harvard Medical School and Massachusetts General Hospital

Andrei Calin, M.D., M.A., M.R.C.P., Stanford University School of Medicine and Palo Alto Veterans Administration Medical Center

Order by Phone

You can order SCIENTIFIC AMERICAN *Medicine* by telephone. **Please call this toll-free number: 1-800-227-3900 (in California call 800-632-2122);** you will be billed after your subscription begins. Toll-free calls are acceptable only for orders placed in the continental United States.



SCIENTIFIC AMERICAN MEDICINE

415 Madison Avenue, New York, N.Y. 10017

Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine*. On receipt of this coupon you will send me the advanced two-volume text described in your announcement and update it regularly by sending me new monthly subsections. I understand that the price of $185 for the first year of service is tax deductible, as is the renewal price of $160. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a *complete refund.*

☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*
☐ I shall also enroll in the CME Program
☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $185*
☐ Please bill me

* Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York

Name _____

MD Specialty _____

Address _____

City _____ State _____ Zip _____

Signature _____

Subscribers outside of the U.S. and possessions will be charged extra for shipping and handling; updates will be sent by surface routes unless airmail delivery is requested. Please allow 6-8 weeks for delivery. 1Z

Important talks in the Middle East

Statesmen aren't the only people in the Middle East who have important talks.

In Saudi Arabia, where we live, Cub Scouts have important talks with Den Mothers. Car owners have important talks with mechanics. Batters have them with umpires. And schoolgirls have *lots* of them with other schoolgirls.

We're Aramco, the Arabian American Oil Company. There are 13,000 North Americans in Saudi Arabia with us. And even though you hear a lot of news about Saudi Arabia, there are things that might surprise you about our lives there.

1. We're doing something important. Aramco produces more oil than any other company. Badly needed oil. Including about 15 percent of the oil the U.S. imports.
2. Aramco is working on some *incredibly* large energy projects. And on huge communications networks, electric utilities, and more.
3. Our people are glad to be in Saudi Arabia with Aramco. They came for excellent pay and professional challenge.
4. After 46 years in Saudi Arabia, Aramco is still growing fast. So is the number of interesting and rewarding jobs we offer.
5. The model airplane took off on the first try.

ARAMCO
SERVICES COMPANY

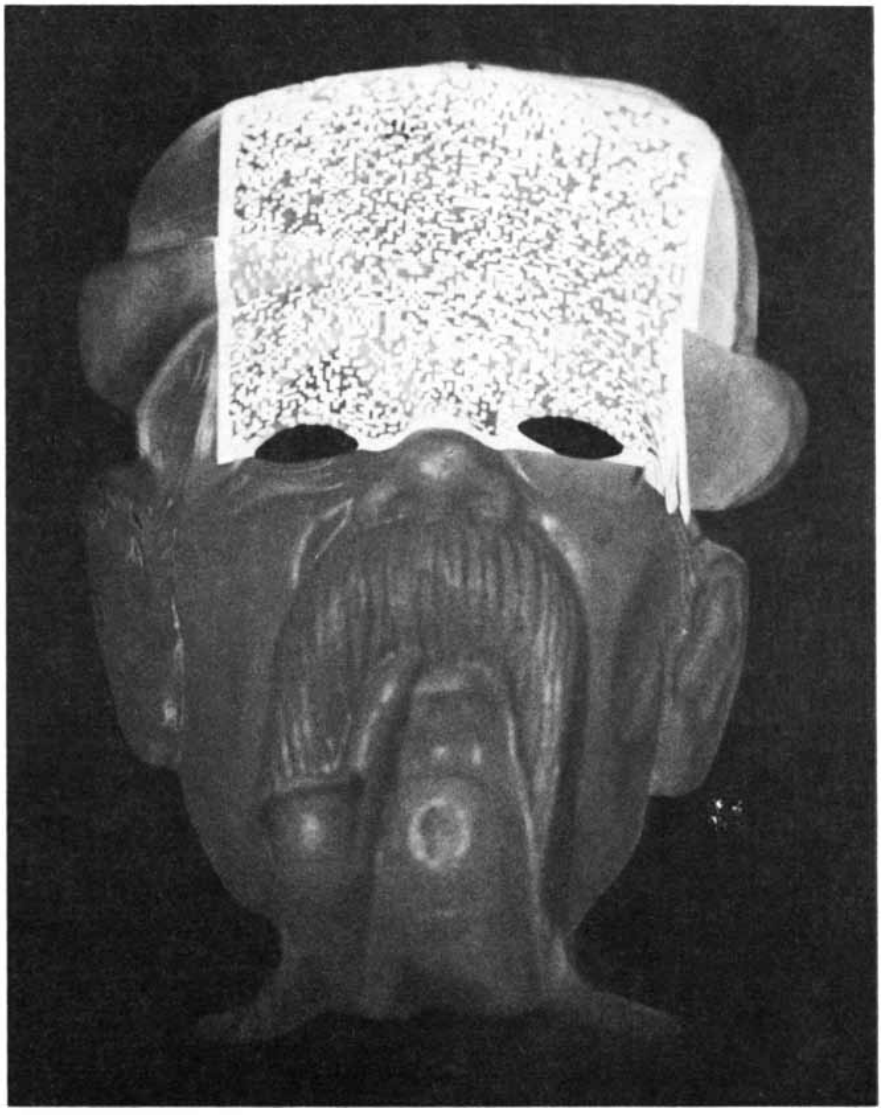
1100 Milam Building FS • CA
Houston, Texas 77002
(713) 750-6965

in that region. When the two pictures are viewed separately with both eyes, the regions of shifted and unshifted dots seem to lie in different depth planes. Through either eye alone one sees only a flat field uniformly speckled with dots.

To present the left and right halves of a random-dot stereogram (or any other stereogram) separately to the two eyes a convenient technique is to print one picture in green ink and the other, superposed, in red and to view the composite through glasses equipped with a red filter for one eye and a green filter for the other. Through the red filter green dots look black and red dots are invisible; that eye therefore sees only the green half of the stereogram. Conversely, the eye covered by the green filter sees only the red half. This kind of stereogram is called an anaglyph.

The next step is to project an anaglyph version of a random-dot stereogram onto an inside-out face mask and to view the combination through red-green glasses. If the stereogram properly perceived seems to show a central square region floating in front of a plane background, what will it look like when the face is seen in depth inversion? According to the monocular-suppression hypothesis, no depth should be seen in the stereogram (that is, the central square should be invisible) because stereopsis requires the integration of the view from both eyes. On the other hand, according to the disparity-reversal hypothesis, depth should be seen in the stereogram but with its direction reversed from the normal perception. The central square should seem to be recessed behind its random-dot surround instead of floating in front of it.

The experiment therefore provides a straightforward test of both hypotheses, and both turn out to be wrong. Every observer reports that depth can be seen in the stereogram while the face is perceived as being depth-inverted. Thus monocular suppression cannot be a factor. Every observer also reports that the direction of depth in the stereogram is the one implied by the actual disparities of the dots, not the opposite as predicted by the disparity-reversal hypothesis.



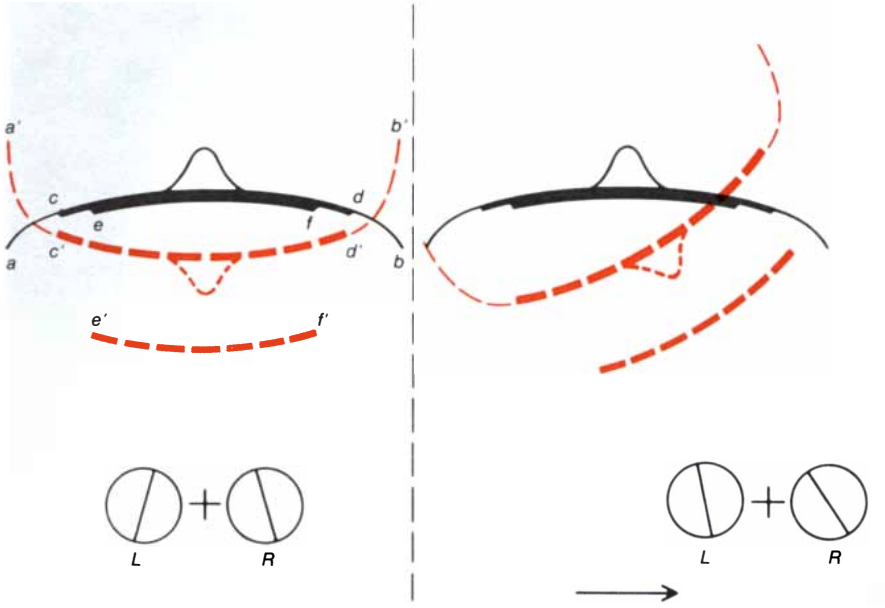
TEST OF HYPOTHESES was made by the author with a setup in which an anaglyph version of a random-dot stereogram (in which the left and right pictures are superposed, with one printed in green and one in red, and viewed through a hand-held filter that is green for one eye and red for the other) is projected onto an inside-out face mask. Because of the color each eye sees only half of the stereogram. According to the monocular-suppression hypothesis, no depth should be seen; according to the binocular-disparity hypothesis, depth should be seen but with its direction reversed. Observers do see the depth, however, and its direction is not reversed.

The second experiment exploits an illusion of depth known as the Pulfrich effect (after the German physicist Carl Pulfrich, who described it in 1922). To demonstrate the illusion one needs a pendulum that swings in a plane arc. A weighted yardstick swinging on a nail works well enough. The observer stands in front of the pendulum and views it binocularly with one eye covered by a light-attenuating filter, such as one lens from a pair of dark sunglasses.

Seen with either eye alone the end of the pendulum appears correctly to be swinging back and forth in a plane. Viewed with both eyes, however, the end of the pendulum appears distinctly to be swinging back and forth in an ellipse. If the filter covers the observer's right eye, the pendulum seems to swing outward toward him as it moves from left to right and away from him as it moves in the opposite direction. If the filter is placed over the other eye, the direction of this apparently elliptical movement is reversed. The magnitude of the illusion (the bulge of the ellipse) increases with viewing distance and also as the filter is made darker, provided that the observer can still see through it.

The accepted explanation for the Pulfrich illusion was proposed initially by Pulfrich himself, apparently following a suggestion from an associate. (Pulfrich was blind in one eye and therefore could not see his own illusion.) The explanation is that the eye covered by the filter has a slower response time than the uncovered eye and that the delay gives rise to what is in effect a binocular disparity between the left and right retinal images as registered at some higher level in the brain. As the pendulum moves across the visual field its momentary position on each retina is the same, but the signal sent to the brain from the covered eye indicating the presence of the pendulum at any given retinal location lags behind the corresponding signal from the uncovered eye. Hence at the level of the brain where simultaneous left and right retinal images are compared it seems there is a disparity between the two eyes' images of the swinging end of the pendulum, and the "disparity" is interpreted in the usual way to signify depth. Pulfrich's original explanation has subsequently been confirmed by many experiments.

My variation was to mount an inside-



TWO ASPECTS of binocular depth inversion of a random-dot stereogram on an inside-out face are depicted. In the panel at the left *ab* represents the face, *cd* the stereogram and *ef* the central square. During depth inversion of the face (*a'b'*) the stereogram appears convex (*c'd'*) and the central square floats in front of it (*e'f'*). Panel at the right shows how the inverted-face-stereogram combination rotates when the observer moves his head to the right.

out face mask on the end of a pendulum and then carry out Pulfrich's demonstration in the usual way. Here again the basic question is whether binocular depth inversion of the face can occur at the same time as stereopsis. According to the monocular-suppression hypothesis, the Pulfrich effect should be absent when the face is seen as being inverted, because that effect depends on the brain's registering binocular-disparity cues and incorporating them into visual experience. On the other hand, the disparity-reversal hypothesis implies that the Pulfrich effect should arise during a depth inversion of the face but that the apparent direction of the illusory elliptical arc should be reversed, as though the filter had been shifted to the other eye.

Neither prediction stands up. Instead one finds that the face can be seen as depth-inverted and can still appear to swing in an elliptical arc. The monocular-suppression hypothesis is therefore ruled out. And since the direction of the movement is the one normally seen, the disparity-reversal hypothesis can be ruled out too.

What do these experiments reveal about depth perception and about the perception of form in general? The central result is that inversion can occur even when the brain mechanism responsible for constructing visual experience has demonstrably registered all the depth information available in normal vision, including the geometrically unambiguous information provided by binocular disparity. On this point, then, Helmholtz and Mach were wrong; monocular vision, with its inherent three-dimensional ambiguity, is not a prerequisite for seeing things inside out. Evidently binocular vision can be equally precarious when the stimulus offers sufficient provocation.

This finding raises two questions. How is the apparent three-dimensional form of a binocularly depth-inverted object related to the two retinal images that give rise to it? What prevents inversion in ordinary vision?

The first question is perplexing because, according to the standard geometry of binocular vision, depth inversion creates an impossible object. What one sees in the mind's eye cannot be geometrically reconciled with the retinal images. The brain appears to ignore this paradox, presenting consciousness with a seemingly coherent visual object. Apparently depth-inverted percepts are constructed from sensory evidence according to definite perceptual rules, but it is not obvious what the rules are.

Initially I was inclined to look for the rules among variations on the disparity-reversal theme. That idea now seems to me to be increasingly implausible. For one thing the theme's geometrical implications for the apparent shapes of binocularly inverted objects do not seem to agree well with what one actually sees in the illusion.

Moreover, the brain appears not to reverse disparity readily, even when the provocation is strong. Mark Georgeson of the University of Bristol has tested this point by creating reliefs of human faces in which the depth is given entirely by binocular disparity; they are faces sculptured by the three-dimensional surfaces seen in random-dot stereograms. Depending on which eye sees which half of such a stereogram, the disparities may create a face that is either normal or inside out.

When the stereogram actually depicts an inside-out face, one might expect perceptual depth inversion to develop easily if the brain is geared to create inverted percepts by reversing binocular disparities. This does not happen. Georgeson finds that these purely stereoscopic faces are always seen in correct depth (that is, inside out, as is implied by their actual disparities), notwithstanding the normal human bias for seeing faces the other way. The finding suggests that disparity reversal is not a trick the brain performs easily, and so it seems to be an unlikely basis for understanding the perceptual geometry of binocular depth inversion. The architectural key to this novel visual world remains to be found.

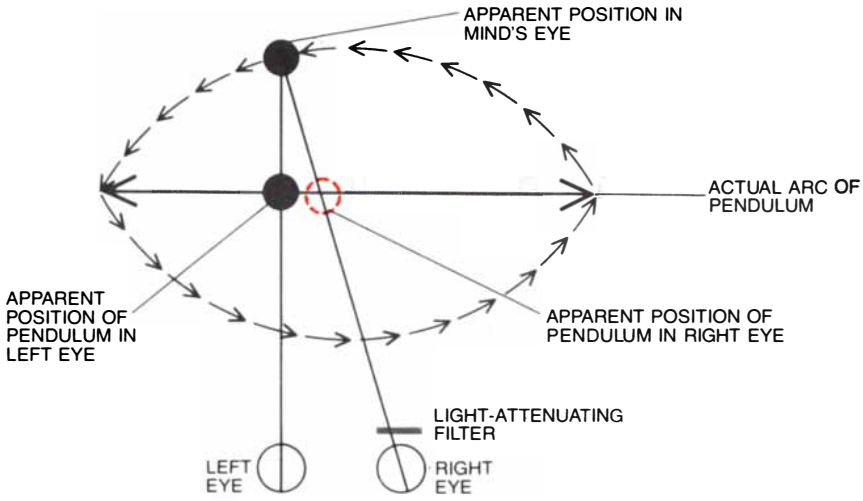
The second question raised by binocular depth inversion is more general. If one can sometimes see objects as being inverted in spite of the availability of every possible cue to depth, why is the mistake so uncommon in normal vision? One thought might be that binocular inversion is an anomaly confined to the special case of inside-out human faces. Perhaps the perception of faces invokes unique mechanisms that do not apply to the perception of other objects.

This idea is easily disposed of, because one can achieve binocular inversion with a broad range of familiar objects. The critical factor seems to be not "faceness" but rather a lifelong habit of seeing certain classes of objects in standard three-dimensional configurations. Thus to explain why inversion is rare one can only say that most of the time the object hypotheses favored by perceptual biases turn out to be correct.

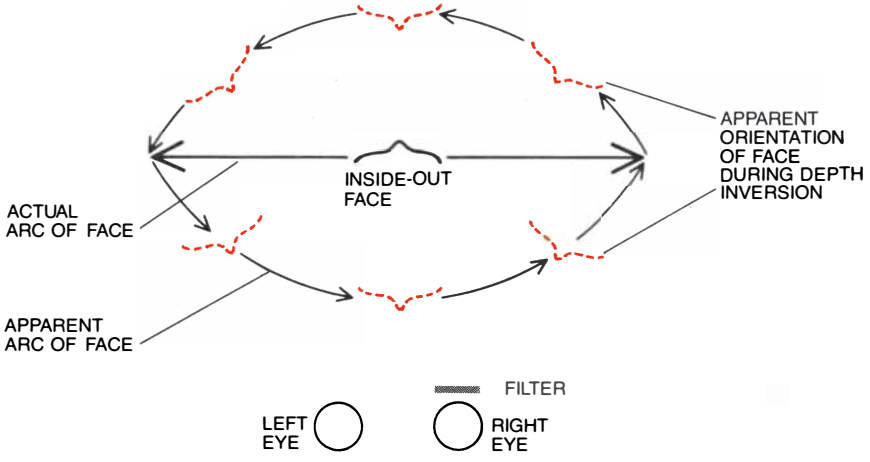
To say that is to say not enough. The basic problem is to understand precisely how mental preconceptions mesh with immediate sensory input to create visual experience. The metaphor of the brain as a tester of hypotheses does not carry one very far.

For example, it is clear that before the visual system can decide to interpret its sensory data according to some specific object hypothesis it must be guided to a roughly appropriate class of hypotheses, otherwise each new object would present an impossible problem in searching. (One could spend a lifetime testing cow-shaped hypotheses against retinal trees.) This guidance must come primarily from immediate sensory evidence, and so the key problem in perception is how the visual system manages so successfully to pull itself up by its own bootstraps.

The question of how much of what a person sees is forced on him by immediate sensory stimulation and how much is



PULFRICH ILLUSION provides the basis for another test of the two hypotheses on binocular depth inversion. In the illusion a pendulum that is in fact swinging in a flat plane appears to follow an elliptical path when one of the observer's eyes looks through a dark but not opaque filter.



SECOND TEST of hypotheses involved mounting an inside-out face mask on a pendulum. The heavy black lines denote the face and its real arc, and the curved arrows show its apparent arc when the right eye is covered by a light-attenuating filter. The colored shapes indicate the apparent orientation of the inverted face along the illusory arc. According to the monocular-suppression hypothesis, the Pulfrich effect should not occur when the face is seen inverted; according to the binocular-disparity hypothesis, the Pulfrich effect should be observed during depth inversion of the face but the apparent direction of the illusory arc should be reversed, as though the dark filter had been shifted to the other eye. Neither prediction proves to be correct.

supplied by the imagination is a long-standing issue in visual science, and this is not the place to discuss it at length. One point, however, does seem to emerge from the study of depth inversion: the critical role of unconscious perceptual learning.

After achieving depth inversion with many different objects one gains the impression that in order to achieve a stable inversion of any object the brain must construct a complete visual model of the object in inverted form: a model that assigns a three-dimensional interpretation to all the idiosyncratic features of the object. This process is evidently much easier with some objects (such as faces) than it is with others.

Nevertheless, it does not seem that the real-world plausibility of the inverted model is really a decisive factor. Plausibility does determine the amount of time and mental effort required to achieve inversion, but my impression is that there is no sharp division between objects that can be inverted in depth and those that cannot be inverted in depth. Instead it seems to be just as likely that the critical factor is the time required to construct an appropriate visual model. Some objects can be inverted in a reasonable period of time and others simply take too long to be inverted. In other words, it may be that you could learn to see everything inside out if you only had time to practice.

The Architecture of Christopher Wren

He was a prominent man of science and a renowned architect at the height of the Scientific Revolution, but he did not exploit principles of theoretical mechanics in designing his buildings

by Harold Dorn and Robert Mark

In the crypt of St. Paul's Cathedral in London is a tombstone with the inscription (translated from the Latin): "Reader, if you wish to see my monument, look around you." The epitaph marks the tomb of Sir Christopher Wren, who designed the cathedral and some 50 parish churches as well as libraries, hospitals and royal residences in England. Wren was also one of Britain's most renowned men of science in the era of Isaac Newton, Robert Hooke and Robert Boyle. Newton described him as one of "the greatest Geometers of our times." Hooke, who contended with Newton for preeminence and generally praised no one, put Wren in a class with Archimedes.

Wren's dual eminence in architecture and science stemmed from a paradoxical career. Although Wren was a professor of astronomy, Charles II appointed him surveyor general, or chief architect of England. He achieved his scientific reputation (and the presidency of the Royal Society) in spite of having published almost nothing. As an architectural theorist he repudiated exposed flying buttresses, yet in St. Paul's he employed them freely. He called for the application of statics to building, but the principles of applied mechanics he recommended were incorrect; indeed, their application would have resulted in structural failure. He often objected to useless ornamentation in architecture. Nevertheless, his most impressive technical accomplishment—the brick cone that carries the lantern, the ball and the cross of St. Paul's—is hidden by an ornamental domical shell that has no structural purpose.

What is perhaps most perplexing about Wren's career is the elusiveness of demonstrable connections between his science and his architecture. Because Wren worked at the peak of the Scientific Revolution, in the second half of the 17th century, it has sometimes been assumed that he employed principles of theoretical mechanics in designing his buildings. Indeed, the Scientific Revolution seemed to promise a merger of the science of mechanics and the art of building. In Galileo Galilei's *Dialogues Concerning Two New Sciences* the first science is not dynamics, for which the book is now better known, but rather the strength of materials, which Galileo had recognized 30 years earlier as a "science that is very necessary in making machines and buildings of all kinds." In 1624 Henry Wotton, the British ambassador to Venice, published a book on architecture in which he analyzed in a rudimentary way the structure of a stone arch. Moreover, in the 17th century it was people who would now be called scientists who were awarded the commissions to design and build monumental structures. In Turin, Guarino Guarini, a mathematician, devised the plans for such celebrated buildings as the church of St. Lorenzo, the Sindone Chapel and the Palazzo Carignano. In Paris, Claude Perrault, a physician and an anatomist, designed the facade of the Louvre and the observatory of the French Academy. In London it was Wren and Hooke who collaborated as chief architect and city surveyor after the city was devasted by the Great Fire of 1666.

We have reviewed several of Wren's structural solutions with the aid of modern engineering techniques. On the basis of our investigation and through consideration of Wren's writings and career we have undertaken to find Wren's place and the place of his architecture in the Scientific Revolution.

Wren was born on October 20, 1632, in East Knoyle in Wiltshire. At an early age he was exposed to ecclesiastical tradition, a circumstance that would later influence his architecture. His father became dean of Windsor and an uncle was bishop of Ely. At the University of Oxford, which Wren entered in 1649, his college warden was John Wilkins, who later became bishop of Chester. In his youth Wren also took an interest in mechanical contrivances and instruments. At Oxford he prepared an exhibition of some 50 devices, almost all of them of a practical nature, including "New Designs tending to Strength, Convenience and Beauty in Building." It was at Oxford that he also began to show the mathematical ability for which Newton would praise him.

In 1657, when Wren was 25, he became professor of astronomy at Gresham College in London, and four years later he was elected Savilian professor of astronomy at Oxford, a chair he held until 1673. In 1662, when the Royal Society was formally established, Wren became one of its charter members and leading lights, and indeed he wrote a draft of the preamble to the society's charter of incorporation. The utilitarian emphasis of the draft indicates that Wren's attitude toward science was already skewed toward the promotion of the practical arts. Among the functions of the society, he wrote, would be "the Advancement of Natural Experimental Philosophy, especially those parts of it which concern the Encrease of Commerce, by the Addition of useful Inventions tending to the Ease, Profit, or Health of our Subjects."

Wren's experience in architecture spanned almost 50 years. It began in 1662, when he was commissioned to design the Sheldonian Theatre at Oxford, and it ended in about 1710, with the completion of St. Paul's Cathedral. These two buildings are thought to include his most impressive solutions to engineering problems, and they exhibit two chief characteristics of his approach to architecture. The first characteristic is his reliance on classical sources, which resulted in a conservative architectural style and generally massive structures. In extreme cases, such as in parts of St. Paul's, the structure is so excessively

PORTRAIT OF CHRISTOPHER WREN shows him surrounded by scientific instruments, books and artwork. The text at his feet is open to a drawing of the Sheldonian Theatre at the University of Oxford, the first building he designed. He is holding a plan for St. Paul's Cathedral, which he rebuilt after it was demolished following the Great Fire of 1666. The London skyline in the background includes some of Wren's parish churches and a building that appears to represent one of Wren's early designs for St. Paul's Cathedral. The portrait, which is currently in the Sheldonian Theatre, was painted by Antonio Verrio, James Thornhill and Godfrey Kneller.

SHELDONIAN THEATRE, which Wren designed in 1662, departed from the Gothic tradition of the buildings at Oxford. He based the design on the theater of Marcellus in Rome.

HEAVENLY SCENE is painted on the ceiling of the Sheldonian Theatre. The flat ceiling, which spans 70 feet, impressed Wren's contemporaries because it was not supported by columns. The resemblance of the Sheldonian Theatre to the Roman theater of Marcellus is enhanced by cords attached to the painting, which shows an unfurled canvas; the Roman theater had a canvas roof that could be furled on supporting cords to open the theater to the sky. The painting is titled *Truth Descending upon the Arts and Sciences;* the figures represent the triumph of the Arts and Sciences over Envy, Rapine and Ignorance. The artist was Robert Street.

heavy that it diminishes the integrity and durability of the building fabric.

The other characteristic of Wren's design is his predilection for novel structural elements. It is these elements that have contributed to the belief that Wren's architecture is substantially in debt to his science. In the Sheldonian Theatre the novel structural elements were timber trusses supporting the ceiling above the auditorium. In St. Paul's Cathedral, which is Wren's most admired building, the innovative element was the towering dome composed of three concentric shells. Thus these two buildings provide the subject matter for a study of the relation between Wren's science and his architecture.

The Sheldonian Theatre was named for Gilbert Sheldon, who was warden of All Souls College at Oxford until he was expelled during the rule of Oliver Cromwell. Sheldon returned to Oxford with the restoration of Charles II, and although Sheldon's second tenure was a short one (because he was soon appointed bishop of London), it was long enough for him to discourage the practice of holding secular gatherings in the university's church. Sheldon provided £12,000 for a secular hall. In selecting Wren as the architect Sheldon chose a fellow of his own college who was also the nephew of a bishop.

Wren based his design for the Sheldonian Theatre on the Roman theater of Marcellus as it was depicted in the 16th century by Sebastiano Serlio, a Bolognese painter and architect. The Roman prototype had the shape of a horseshoe and was open at the top, covered only by a canvas awning. Wren's design was a polygonal approximation to the shape of the Roman theater and had a permanent roof. Wren also provided for multiple uses of the building: the basement had to house the presses of the university and the attic had to store the books printed there.

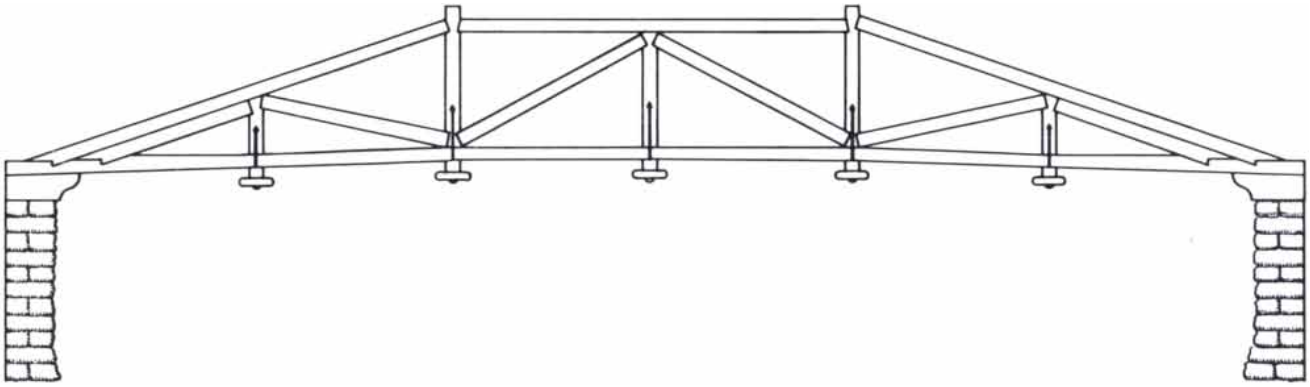
The design was notable mainly as a departure from the Gothic tradition of the buildings at Oxford. Although biographers of Wren have generally found that the design lacks the quality of his mature style, they have admired the flat auditorium ceiling constructed without supporting columns. Support for the ceiling is provided by the heavy trusses concealed in the attic. The originality of employing a truss to create a long, unobstructed ceiling can be judged from the fact that in 1663 Wren displayed a model of the theater at the Royal Society. The structure of the truss, however, is disappointing. The truss bears a greater resemblance to a medieval king-post truss than it does to the more efficient triangulated truss, which minimizes bending forces. The Venetian architect Andrea Palladio had described the role of the triangulated truss in bridge construction in *The Four Books of Architec-*

ture, published in 1570. Wren had access to Palladio's work, and so we have concluded that either he did not appreciate the greater efficiency of the triangulated form or he did not make the connection between the structure of bridges and the structure of buildings.

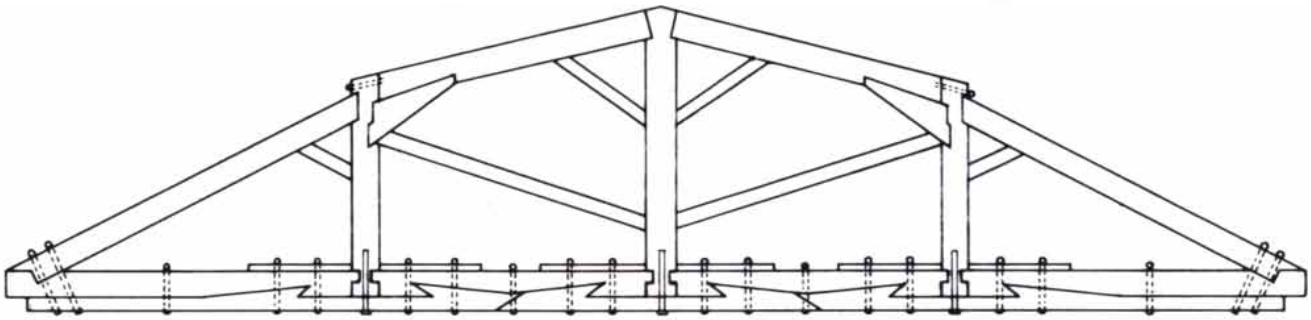
Our conclusions were confirmed by an analysis of Wren's truss by photoelastic methods, in which a model of the truss made of transparent plastic is placed between crossed polarizing filters. The interference pattern that results from passing light through this sandwich of materials can be interpreted as a contour map of stress intensity [see "The Structural Analysis of Gothic Cathedrals," by Robert Mark; SCIENTIFIC AMERICAN, November, 1972]. We found, however, that the deficiency of Wren's design would not result in severe distress under normal loading. The original trusses were replaced in 1802, so that perhaps the weight of the stored books was greater than anticipated.

Whatever the quality of Wren's ceiling design, the trusses owe nothing to theoretical analysis. The theory of the truss, even of relatively simple configurations, remained far in the future. There is no indication either in Wren's notes or in the testimony of his contemporaries that the design of the Sheldonian trusses was anything but empirical and traditional. If any of the theoretical novelties of the Scientific Revolution entered into Wren's approach to design, they cannot be detected in the Sheldonian Theatre. Therefore we looked for them in the design of the structure that was the summit of Wren's career.

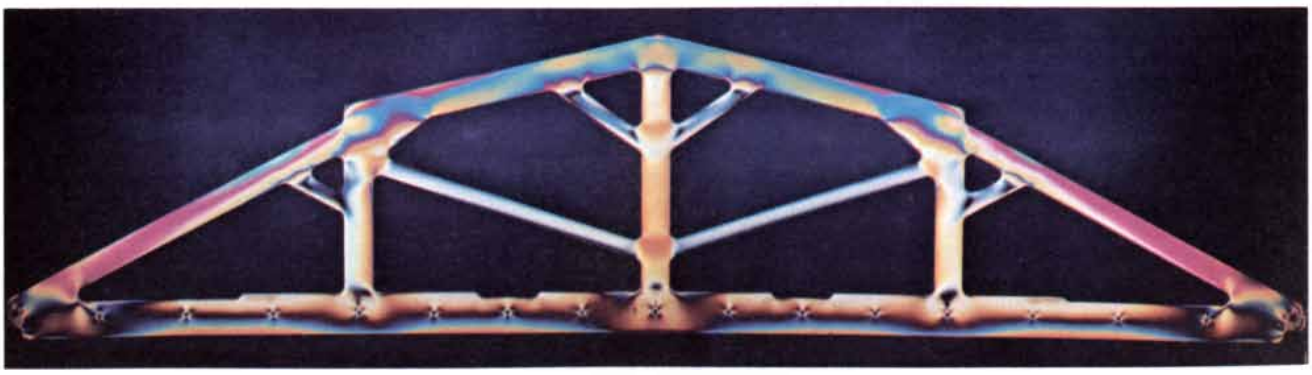
The Great Fire of 1666 had two salutary consequences. It halted the plague that had been sweeping through London and it gave Wren an opportunity to deploy his genius as an architect. One of the most pressing needs was the rebuilding of the parish churches, which had



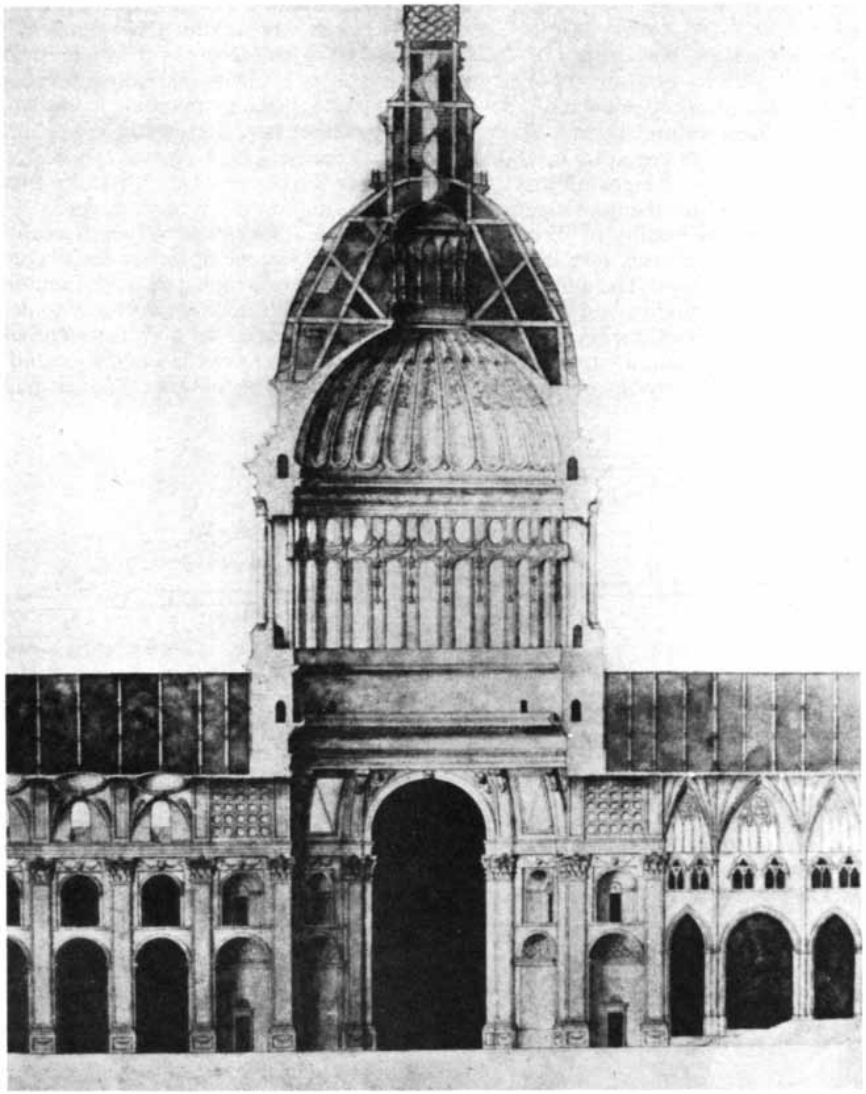
TIMBER BRIDGE TRUSS, described in 1570 in *The Four Books of Architecture* by the Venetian architect Andrea Palladio, is efficient because it is made up of triangular segments. The truss is reinforced near the supports so that it will hold up under the increased forces there. Wren had access to Palladio's work but did not adopt the triangulated truss in his design of the ceiling of the Sheldonian.



TRUSS CONCEALED IN THE ATTIC of the Sheldonian Theatre supports the ceiling. The Sheldonian truss bears more resemblance to a medieval king-post truss than it does to the triangulated truss Palladio described, which keeps the bending forces at a minimum.



PHOTOELASTIC MODEL of the Sheldonian roof truss indicates that the stress is not distributed evenly. The transparent plastic model is viewed between crossed polarizing filters. The interference pattern of the light is a map of stress intensity. Each color represents a different degree of interference, which corresponds to a degree of stress. The magenta areas of upper chords near the supports are subject to about twice as much stress as the blue areas of upper chords. There is almost no stress in the diagonals, which hardly show interference.



PLAN OF A DOME for St. Paul's Cathedral is shown in a sketch Wren made some months before the Great Fire. He proposed replacing the main tower with a double-shelled dome and a spire. The inner dome would have been made of masonry and the outer dome of timber.

formed the backbone of the social and spiritual life of the city. Within days Wren developed a plan for reconstructing London. Although his plan was not adopted, he was soon appointed the principal architect for the rebuilding of the city. In that capacity, and later as surveyor general, he attended to the reconstruction of the parish churches, which constitute a major part of his architectural legacy. At the same time Wren directed much of his interest to the prospect of rebuilding St. Paul's Cathedral, which had been heavily damaged in the fire.

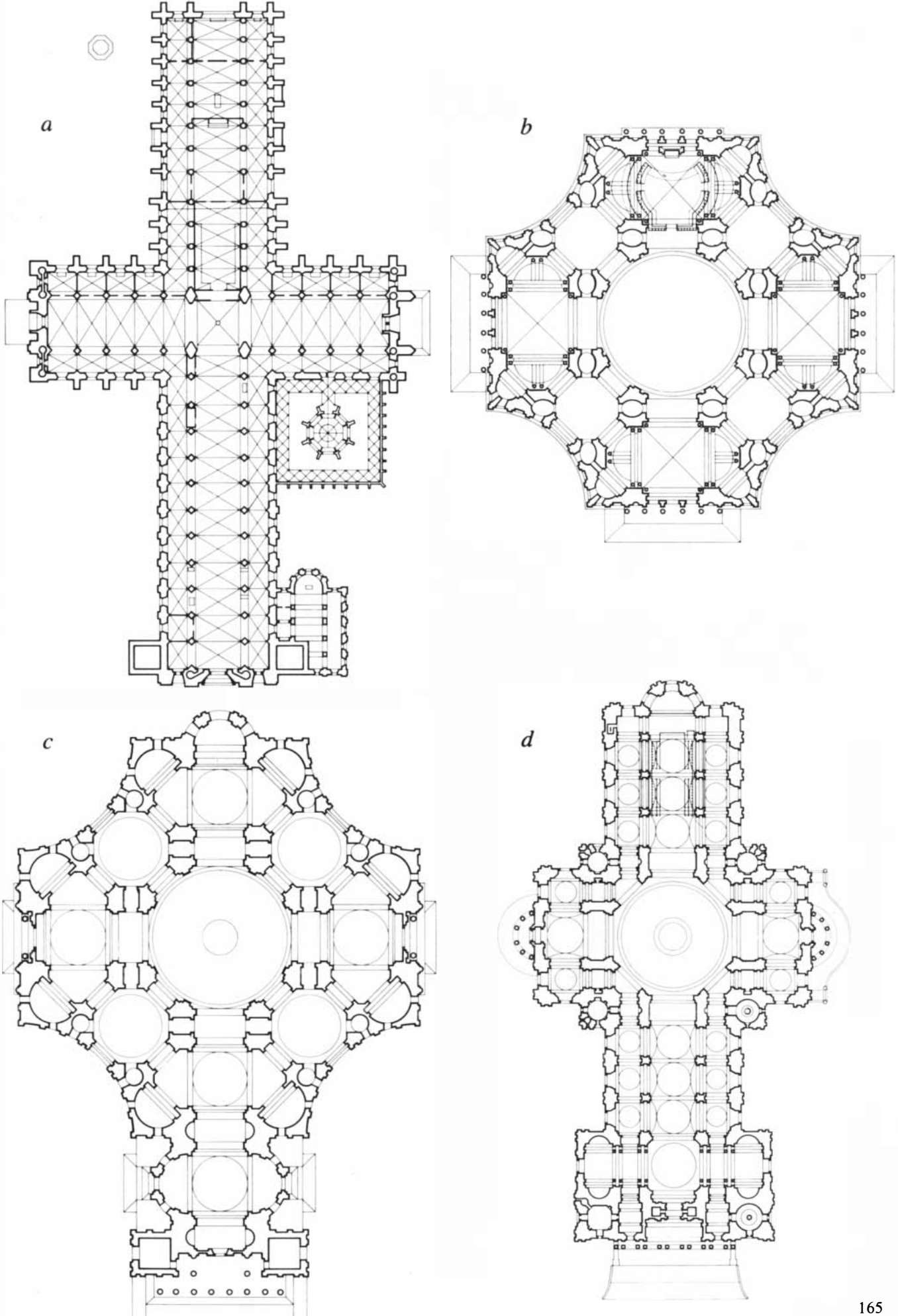
Wren had served an apprenticeship that prepared him for his role as architect of the new cathedral. Before the civil war of the 1640's the exterior of the old St. Paul's had been the object of a campaign of classical restoration under the direction of Inigo Jones. The 11th-century Norman nave and the 13th-century Gothic choir were largely covered with stone facing in the style of the Renaissance and the west facade was reconstituted into a Corinthian portico. A few months before the fire Wren had been asked to suggest further reconstruction. He called for rebuilding the interior so that it would correspond to the Renaissance exterior and for replacing the pointed medieval vaults with lighter, domical vaults. Wren found fault with what he termed the "Saracen" (or Gothic) style of the building. He noted that the piers were visibly out of plumb as a result of shoddy building, and he condemned the flying buttresses exposed to the weather. He also proposed replacing the central tower with a dome and spire.

The most interesting structural feature of this early scheme is the light, high outer dome constructed of timber and supported by an inner masonry shell. The concept of this relatively simple domical structure may have come directly from the much smaller dome of the church of the Sorbonne in Paris, built by Jacques Lemercier some 30 years earlier, which Wren probably saw in 1665. This kind of double-shell construction can be traced back through a sequence of buildings to the high domes of St. Mark's Cathedral in Venice, built in the 13th century.

After the fire, when it was eventually agreed that the old cathedral could not be salvaged, Wren evidently considered himself free to draw on Roman precedent in planning the new cathedral, as he had done in designing the Sheldonian Theatre. The ease with which he had broken with architectural tradition at Oxford, however, was not to be repeated with his ecclesiastical clients in London. Even with the backing of Charles II he was not able to prevail against the clergy, who insisted that the plan of the new building follow "cathedral fashion," namely that it have the cruciform plan of a medieval church. In order to resolve the conflict Wren submitted a series of designs and ultimately made concessions to the ecclesiastical authorities (and to popular taste) before he was allowed to proceed with the rebuilding in 1675.

What Wren initially had in mind for the cathedral can be surmised from two city plans he drew up soon after the fire. Both show a structure resembling the Pantheon in Rome and consisting chiefly of a domed rotunda joined to a rectangular hall. Some commentators have dismissed these plans as casual sketches rather than serious proposals. Nevertheless, the plans already show Wren's determination to build a domical structure and his reluctance to follow the traditional cruciform plan. Moreover, the early designs may reflect his Protestantism in that they would bring the clergy and the laity closer than they were in a medieval church, where the laity in the

SEQUENCE OF GROUND PLANS for St. Paul's illustrates the compromises Wren had to make before Charles II granted him permission to proceed with construction. At the upper left is the old St. Paul's Cathedral, which had a traditional cruciform configuration. At the upper right is Wren's Greek Cross Design of 1672. The main feature of the design is a domed, circular space at the center. The centralized structure seems to reflect Wren's Protestant beliefs in that it brought the clergy and the congregation closer than they were in a medieval church. At the lower left is the Great Model Design, named for an 18-foot model built by Wren. The plan is similar to the Greek Cross Design except for the elongated east-west axis. At the lower right is the actual cathedral, which was begun in 1675. The final plan represents a compromise between Wren's original centralized plan and the medieval basilica plan of the old cathedral.



nave are separated from the clergy and the choir in the chancel.

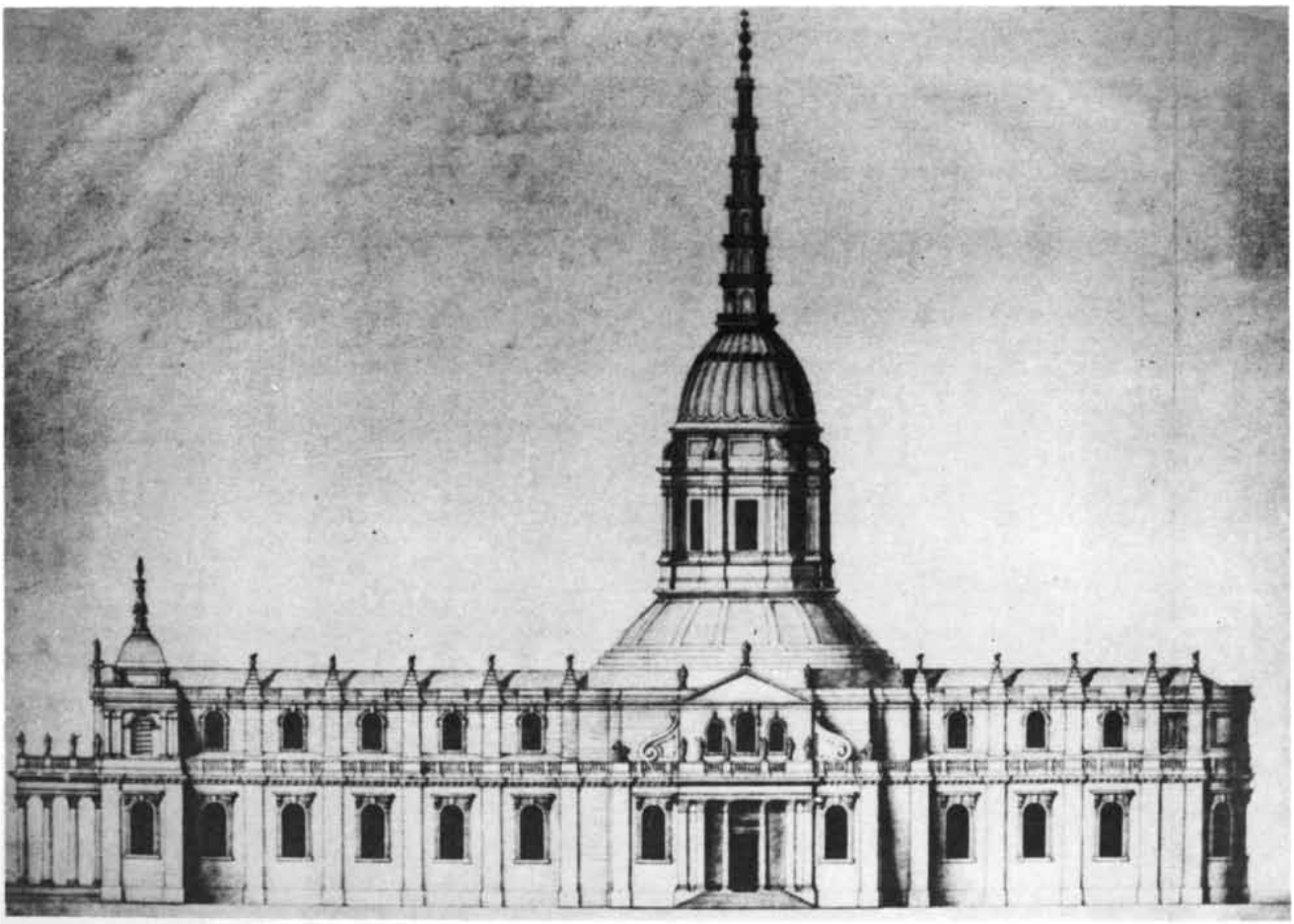
Wren had a high regard for the liturgy of the English prayer book, and he conceived of a church as an "auditory" where the congregation would hear as well as see the preacher. He estimated the maximum distance the preacher's voice might carry and designed churches accordingly. The Protestant expectation that the clergy and the laity participate jointly in the service called for a compact meeting room uncluttered by screens. The city-plan sketches, with a domed rotunda for ceremonial occasions and a simple, rectangular hall for ordinary services, were evidently designed to meet the ecclesiastical requirements of English Protestantism. Indeed, the city-plan designs may represent not casual sketches but Wren's uncompromised ideas on the design of a Protestant cathedral.

The influence of Protestantism can also be seen clearly in some of Wren's designs of parish churches. "It would be vain," he wrote, "to make a Parish church larger than that all who are present can both hear and see." In designing the parish church of St. James's, Wren tried to build the ideal auditorium, with the nave and the chancel in a single room and with balconies to increase the seating capacity without enlarging the hall. "I can hardly think it practicable," he wrote, "to make a single Room so capacious, with Pews and Galleries, as to hold 2,000 Persons, and all to hear the Service, and both to hear distinctly, and see the Preacher. I endeavoured to effect this in building the Parish Church of St. James's, Westminster, which I presume is the most capacious, with these Qualifications, that hath yet been built."

The city-plan sketches of St. Paul's Cathedral were forerunners of the next series of designs, on which Wren began to work in earnest in 1670. These designs too were based on a structure whose chief space was domed and circular. One version, the Greek Cross Design, incorporated a great raised central dome that resembles a design for St. Peter's in Rome that had been proposed by the architect Donato Bramante in the early 16th century. In the case of Bramante's plan as in that of Wren's the central-dome design was rejected by the clergy, who remained committed to the cruciform configuration.

In 1673, in an attempt to placate his ecclesiastical critics, Wren offered the Great Model Design, whose plan was similar to the Greek Cross Design except for an elongation that he grudgingly introduced along the east-west axis to form a short nave and a rudimentary choir. The name of the design comes from a large model Wren built in such a way that the interior is seen from the same perspective as the interior of the building would have been. In making the model he went to considerable trouble and expense to impress his patrons with what he thought was a satisfactory compromise. Nevertheless, the design was rejected, and he resolved "to make no more Models, or to publicly expose Drawings."

Wren's next effort was the awkward Warrant Design of 1674. It has been interpreted as Wren's exasperated and contemptuous reply to his critics. It represented a retrogression almost to the old cathedral as modified by Jones. The drawings of the Warrant Design are known to have been prepared in some haste, which suggests that Wren relied on existing designs; his biographers cite as the source a number of earlier buildings described in 16th- and 17th-century architectural literature. It is also possi-



WARRANT DESIGN done by Wren in 1674 was approved by Charles II on March 14, 1675. Wren was granted the right to make changes of certain kinds, and he exercised that right by building a cathedral quite different from the one specified in the Warrant Design.

ble the Warrant Design is one of Wren's own early plans, rehabilitated from his file. In any event the design served his immediate purpose. Charles II signed the warrant allowing construction to begin, with the understanding that Wren would be free to make certain alterations. Wren evidently interpreted his freedom quite broadly, and it seems to have been his intention from the outset to use the warrant as a cover under which he could restore many of the elements of his earlier plans.

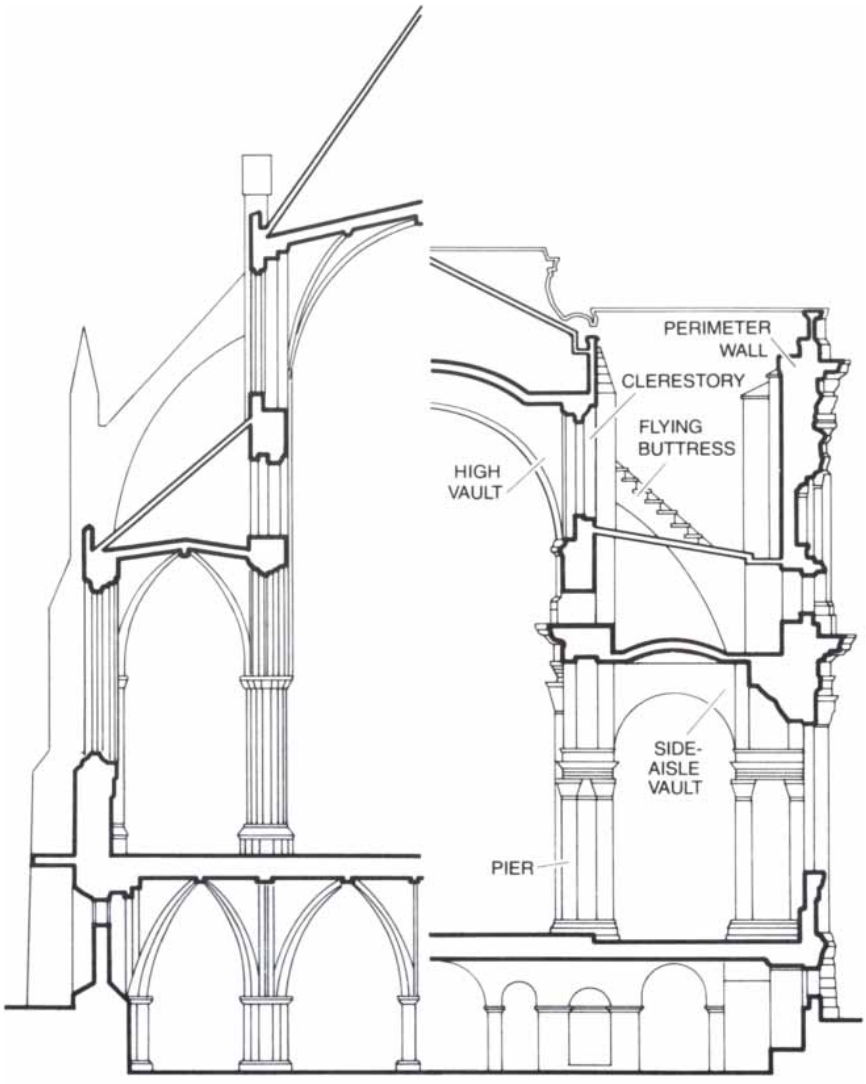
In the cathedral as it was finally constructed Wren in effect created two buildings. The interior is modeled on the plan and elevation of a medieval basilica, with a high center-aisle vault flanked by lower side-aisle vaults and flying buttresses intended to support the clerestory walls (the interior walls that extend above the adjacent side-aisle roofs). None of these structural details, however, are apparent from the outside, where the great central dome appears to rise from a massive two-story base. This effect was accomplished by raising the perimeter walls so that they conceal both the inner features of the building and the flying buttresses. The concealment has led to charges of architectural fraud. It has been said that the perimeter walls, which even have false windows, are nothing more than screens meant to hide the Gothic clerestory and the flying buttresses of which Wren disapproved in his theoretical writings.

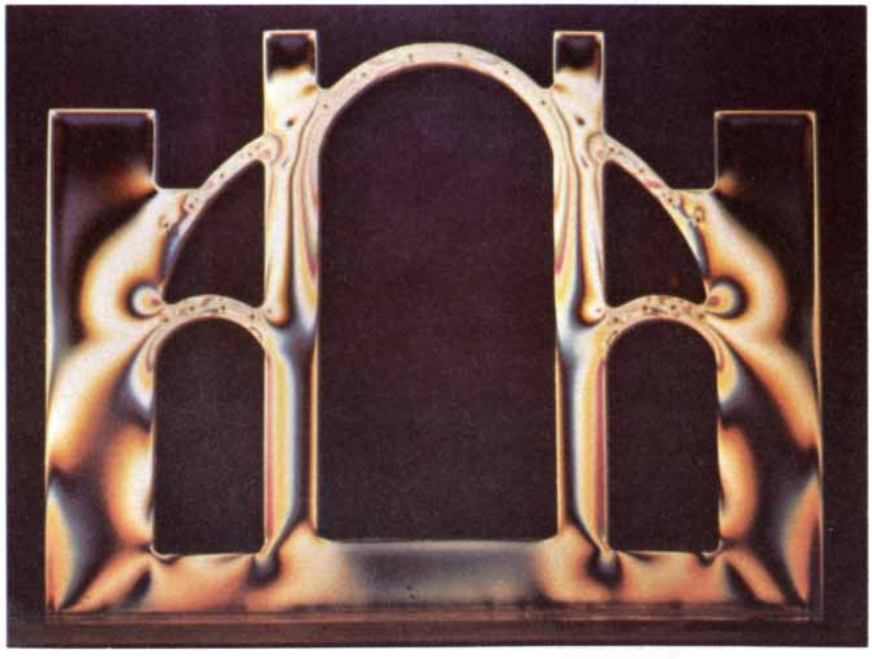
The perimeter walls are quite heavy, however, and Wren's design has been defended on the ground that such massive walls could not have been intended merely as screens. In 1923 Somers Clark, who had overseen the upkeep of the cathedral, denied that Wren had built "one half of the church to hide the other." Clark rationalized that the perimeter walls were not screens but buttresses. He argued that they not only stabilize the great bastions that help to support the dome but also serve as buttresses "to receive the thrust of the aisle vaults and of the flying buttresses which maintain the high vaults."

To confront the question of Wren's purposes, with an eye toward the possibility that he employed theoretical principles, we analyzed the structure of a typical bay of St. Paul's with the aid of photoelastic techniques. A model of the bay was tested first under loadings that simulate the actual dead weight and then under loadings that represent the effects of wind, which were determined from local meteorological data and from wind-tunnel tests. Then we removed the flying buttresses from the model and tested it again under the two kinds of loading. Stresses in the piers increased when the flying buttresses were removed, but we ascertained that well-constructed piers of solid masonry (or at least piers with a firm outer shell composed of several layers of coursed masonry) could have successfully resisted the thrusts of the vaulting and the effects of high winds.

The test results suggest that the flying buttresses were unnecessary. In view of Wren's earlier warning that flying buttresses "are the first Thing that occasion the ruin of Cathedrals," it is ironic he adopted them in St. Paul's. It is of course possible that Wren believed the perimeter walls were necessary for structural integrity and therefore he may have intended them not as screens but as buttresses. Nevertheless, even this defense is questionable, because if Wren was thinking only of structural necessity, he could have provided discrete piers to receive each flying buttress rather than continuous, massive walls along

ST. PAUL'S CATHEDRAL looks today almost as it did when Wren built it. The balustrade was added over Wren's protest in 1717, when he was 85 and no longer in charge of the construction. Isaac Newton was consulted and it is believed he visited Wren to mollify him.





the nave and the choir. Discrete piers supporting flying buttresses, however, would have given St. Paul's the decidedly Gothic look that Wren deplored (ostensibly on functional and structural grounds, although his Protestantism may also have influenced his judgment). As for Clark's suggestion that the perimeter walls buttress the side-aisle vaults, it should be noted that the thrust of these vaults reaches the walls well below the top. Even if one allows for the possibility that Wren thought the additional height (and hence the additional weight) would increase stability, the height to which he carried his massive (and expensive) walls suggests he was employing them chiefly as screens.

To deal with Clark's further assertion that the perimeter walls help to stabilize the bastions that form part of the buttressing system of the great dome, we did a separate analysis of the dome and its supporting structure. Wren hesitated for many years before committing himself to a particular design for the dome. As late as 1694 there are references in his notes to experimental models of the dome, and sketches of it made under his direction as late as 1703 are different from the final plan. Wren's hesitation may have been caused by alarming reports of structural defects in the dome of St. Peter's in Rome. The high dome of St. Peter's, constructed between 1585 and 1590, was plagued with structural problems from the beginning. The major source of the problems is revealed by comparing St. Peter's with the Pantheon in Rome, constructed in about A.D. 120. In the ancient building the outer profile is relatively flat and the outward thrust of the dome is met by a low, massive ring of masonry. In St. Peter's the central, upper portion of the dome exerts a similar outward thrust, but there is only the more vertical portion of the dome itself to provide resistance, which has proved to be inadequate. The dome has cracked along meridians as it spread outward. Over the years some 10 iron chains have been placed around the

SECTION OF THE CHOIR of the old St. Paul's (*upper left*), on which construction began in 1258, is compared with that of Wren's cathedral (*upper right*). Wren's cathedral is much heavier. The perimeter walls make the central dome appear to have a classical base. A model of the choir structure (*lower photograph*), stressed by dead-weight loading, shows that the upper story of the perimeter walls, which hides the flying buttresses, hardly contributes to structural stability. Analysis shows that the bottom section of the building, which is heavier than the corresponding section of many Gothic cathedrals, is adequately braced under both dead-weight and high-wind loadings; it would remain secure even if the flying buttresses were removed.

WHAT'S BETTER THAN SPEED READING?

SPEED LEARNING

(SPEED PLUS COMPREHENSION)

Speed Learning is replacing speed reading. It's easy to learn... lasts a lifetime... applies to everything you read... and is the only accredited course with the option of college or continuing education credits.

Do you have too much to read and too little time to read it? Do you mentally pronounce each word as you read? Do you frequently have to go back and re-read words or whole paragraphs you just finished reading? Do you have trouble concentrating? Do you quickly forget most of what you read?

If you answer "yes" to any of these questions — then here at last is the practical help you've been waiting for. Whether you read for business or pleasure, school or college, you will build exceptional skills from this major breakthrough in effective reading, created by Dr. Russell Stauffer at the University of Delaware.

Not just "speed reading" — but speed reading-thinking-understanding-remembering-and-learning

The new *Speed Learning Program* shows you step-by-proven-step how to increase your reading skill and speed, so you understand more, remember more and use more of everything you read. The typical remark made by the 75,000 slow readers who completed the *Speed Learning Program* was: "Why didn't someone teach me this a long time ago?" They were no longer held back by the lack of skills and poor reading habits. They could read almost as fast as they could think.

What makes Speed Learning so successful?

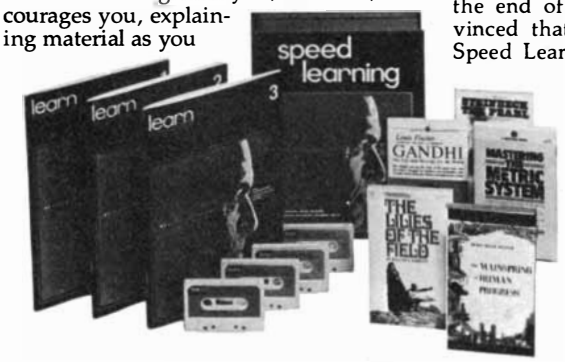
The new *Speed Learning Program* does not offer you a rehash of the usual eye-exercises, timing devices, costly gadgets you've probably heard about in connection with speed reading courses or even tried and found ineffective.

In just a few spare minutes a day of easy reading and exciting listening, you discover an entirely new way to read and think — a radical departure from anything you have ever seen or heard about. Research shows that reading is 95% *thinking* and only 5% eye movement. Yet most of today's speed reading programs spend their time teaching you rapid eye movement (5% of the problem) and ignore the most important part (95%) *thinking*. In brief, *Speed Learning* gives you what speed reading *can't*.

Imagine the new freedom you'll have when you learn how to dash through all types of reading material *at least* twice as fast as you do now, and with greater comprehension. Think of being able to get on top of the avalanche of newspapers, magazines and correspondence you have to read . . . finishing a stimulating book and retaining facts and details more clearly and with greater accuracy than ever before.

Listen-and-learn at your own pace

This is a practical, easy-to-learn program that will work for you — no matter how slow a reader you think you are now. The *Speed Learning Program* is scientifically planned to get you started quickly . . . to help you in spare minutes a day. It brings you a "teacher-on-cassettes" who guides you, instructs, encourages you, explaining material as you read. Interesting items taken from *Time* Magazine, *Business Week*, *Wall Street Journal*, *Family Circle*, *N.Y. Times* and many others, make the program stimulating, easy and fun . . . and so much more effective.



Executives, students, professional people, men and women in all walks of life from 15 to 70 have benefited from this program. Speed Learning is a fully accredited course . . . costing only 1/5 the price of less effective speed reading classroom courses. Now you can examine the same, easy, practical and proven methods at home . . . in spare time . . . without risking a penny.

Examine Speed Learning FREE for 15 days

You will be thrilled at how quickly this program will begin to develop new thinking and reading skills. After listening to just one cassette and reading the preface you will quickly see how you can achieve increases in both the speed at which you read and in the amount you understand and remember.

You must be delighted with what you see or you pay nothing. Examine this remarkable program for 15 days. If, at the end of that time you are not convinced that you would like to master Speed Learning, simply return the program and owe nothing. See the coupon for low price and convenient credit terms.

Note: Many companies and government agencies have tuition assistance plans for employees providing full or partial payment for college credit programs.

In most cases, the entire cost of your Speed Learning Program is Tax Deductible.

COLLEGE CREDITS

You may obtain *2 full semester hour credits* for course completion, wherever you reside. Credits offered through Whittier College (California). Details included in your program.

CONTINUING EDUCATION UNITS

National Management Association, the world's largest association of professional managers, awards 3.0 CEU's for course completion. CEU's can be applied toward the certificate in Management Studies.

PROFESSIONAL SOCIETIES

Speed Learning is offered internationally to members of professional associations such as: American Chemical Society, Foundation for Accounting Education, Institute of Electrical and Electronics Engineers and dozens more. Consult your Education Director for information.

BUSINESS, INDUSTRY, GOVERNMENT

Many companies and government agencies offer Speed Learning as a wholly-paid or tuition reimbursement program. Consult your Training or Personnel Director for details.

31SAH

learn INCORPORATED 113 Gaither Drive, Mt. Laurel, NJ 08054

YES! Please rush me the materials checked below:

- ☐ Please send the Speed Learning program @ $99.95 plus $4 postage and handling.
- ☐ Please send the Speed Learning Medical Edition @ $109.95 plus $4 postage and handling.
- ☐ Please send the Junior Speed Learning program (ages 11 to 16) @ $89.95 plus $4 postage and handling.

NJ residents add 5% sales tax.

Check method of payment below:

- ☐ Check or money order enclosed (payable to learn incorporated)
- ☐ Charge my credit card under the regular payment terms
  - ☐ Visa ☐ Master Card Interbank No._____ ☐ American Express
  - Card No._____ Exp. Date_____

I understand that if after 15 days I am not delighted in every way, that I may return the materials in their original condition for a full refund. No questions asked.

Name _____

Address _____

City_____ State_____ Zip_____

Signature_____

If you don't already own a cassette player, you may order this Deluxe Cassette Recorder for only $49.95. (Includes handling and delivery.)

Check here to order ☐

Outside USA add $10 per item—Airmail extra

SCIENCE/SCOPE

The first of a new breed of communications satellites is providing secure voice, video, high-speed data, and electronic mail services for U.S. business firms and industries. The powerful SBS satellite is the first domestic communications spacecraft to operate at frequencies of 12 to 14 gigahertz. This means that small antennas with diameters of 5.5 to 7.7 meters can be used in urban areas for business communications without causing interference to terrestrial microwave systems. The satellite has an outer cylindrical panel of solar cells that drops down in orbit to expose an inner panel. The telescoping feature nearly doubles the power. Hughes built SBS for Satellite Business Systems, a company owned jointly by Aetna Life & Casualty, Comsat General Corp., and IBM Corp.

More than 15 years after its launch, NASA's Pioneer 6 spacecraft still transmits data via a Hughes traveling-wave tube. The interplanetary probe measured the sun's corona, studied solar storms, and measured a comet's tail. It also made many discoveries about the sun, solar wind, solar cosmic rays, and solar magnetic field. Pioneer 6's primary TWT operated more than 122,000 hours from launch in December 1965 until February 1980. Then, due to a low voltage condition, the backup TWT was switched on. The Hughes Model 214H TWT, which operates in the S band with 8 watts of power, was developed under contract to TRW.

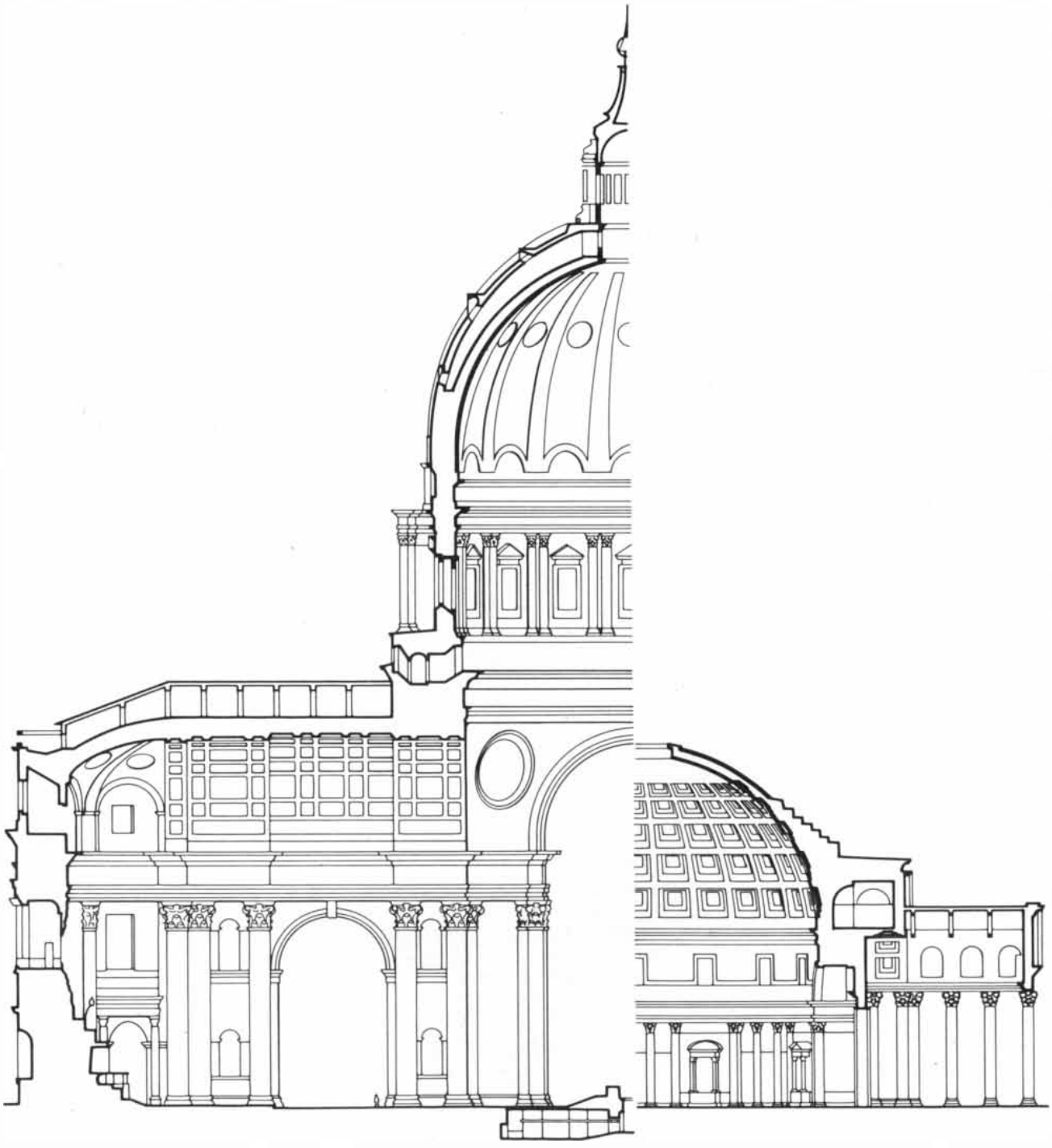
Passengers flying on a number of airlines between Miami, the Caribbean, and South America are enjoying smoother flights because of a weather satellite. Although tropical thunderstorms develop and die quickly through much of the area, airlines can avoid them and take advantage of beneficial winds by studying pictures from a GOES (Geostationary Operational Environmental Satellite) spacecraft. The airlines can even determine when turbulence will be at a minimum and then advise cabin attendants to serve in-flight meals. GOES was built by Hughes and is operated by the National Oceanic and Atmospheric Administration.

Career growth opportunities exist at all levels at Hughes Support Systems for a variety of engineers qualified by degree or extensive work experience. They include systems engineers and software and hardware design engineers for major simulation and test equipment programs. Also, field engineering posts throughout the U.S. and the world offer travel, autonomy, and responsibility for the life cycle of Hughes electronic systems. Phone collect (213) 670-1515, Ext. 5444. Or send your resume to Professional Employment, Dept. SA, Hughes Aircraft Company, P.O. Box 90515, Los Angeles, CA 90009. Equal opportunity employer.

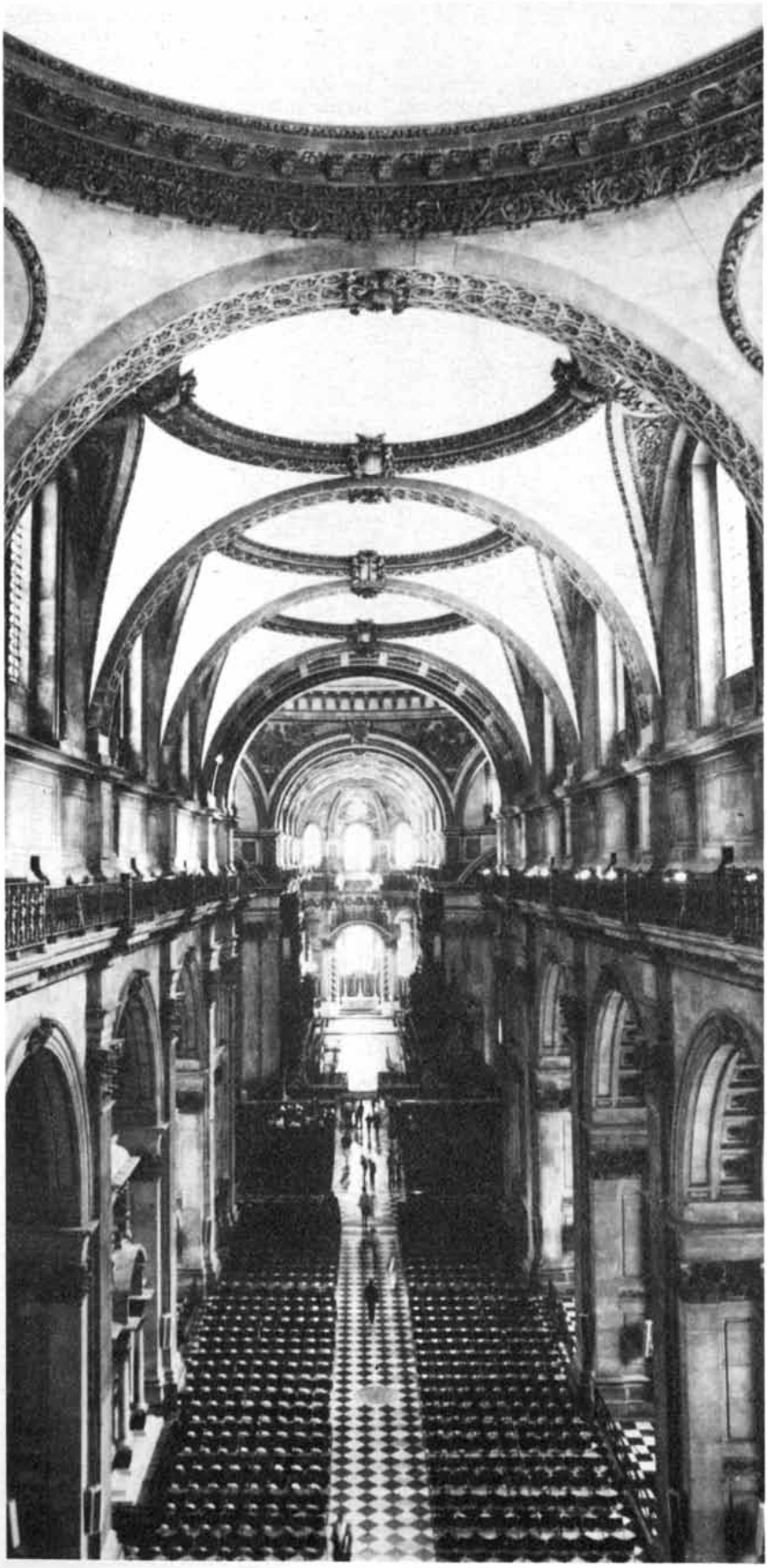
NASA's Pioneer Venus Orbiter continues to gather information after more than two years above the cloud-shrouded planet. Since arriving in December 1978, the spacecraft has relayed more than 40 million bits of data to Earth and transmitted nearly 1000 ultraviolet pictures of the planet's clouds. The radar mapper on board the Orbiter has mapped 93 percent of the surface to show a terrain of mountains, high plateaus, and great plains. The Hughes-built spacecraft is now being allowed to drift from its precise orbit to look at Venus from new vantage points. In the drift mode the Orbiter will drop into the atmosphere and burn up in 1992. Mission activities are planned to last through 1986.

dome to prevent the spreading, but the problem is aggravated by the great mass of the masonry dome itself, which generates extremely large forces.

In designing the dome of St. Paul's, Wren faced problems of another kind. He had constructed the central piers of the cathedral before he designed the dome. Building records indicate that the piers had spalled, or chipped, as early as 1691 and that in 1696 they had been noted to have settled to different depths in the ground. Wren finally produced an elegant solution: an extraordinarily light triple dome. The majestic outer dome is a light shell of lead-sheathed timber. The middle dome, which cannot be seen either from outside or from inside the cathedral, is a thin, chain-girdled brick cone that supports the outer dome and an 850-ton stone lantern. The inner dome, which is also made of brick, forms an attractive hemispherical ceiling. Unlike the outer dome of St. Peter's, the brick cone of St. Paul's, which carries the greatest part of the load, is actually stabilized by the heavy central



PROBLEM OF HIGH-RENAISSANCE DOMES is brought out in a cross-sectional comparison of the dome of St. Peter's Cathedral in Rome (*left*) with that of the Pantheon (*right*). The exterior surface of the dome of the Pantheon is comparatively flat, and the great outward thrust of the dome is resisted by massive, low circular walls. The upper part of the dome of St. Peter's, which supports a lantern, exerts a similar outward thrust. To resist the thrust, however, there are no massive walls but only the lower, steeper part of the dome itself. Some 10 iron chains have been wrapped around the dome to help prevent it from spreading, but cracking of the structure is still a problem.



INTERIOR VIEW of St. Paul's Cathedral shows the heavy structure that is characteristic of the Baroque style. The massiveness of the building is evident in the piers of the nave arcade.

lantern. The conical structure, instead of being subject to destructive bending forces, acts mainly in compression.

A single chain around the brick cone of St. Paul's has proved sufficient to maintain the integrity of the comparatively light structure against outward thrust. Studies of the triple-dome configuration we have done with the aid of a computer indicate that stresses within the supporting masonry are generally low under both gravity and wind loads and that the single chain is well placed. There have been problems with the dome's supporting structure, but they are unrelated to the structure of the dome itself. If Wren had completed the design of the dome at an earlier stage of the project, before he began construction of the central piers, he might have realized the piers could have been lightened, in which case some of the distress resulting from their differential settlement might have been avoided.

Our analysis refutes Clark's assertion that the perimeter walls are needed to buttress the dome-supporting bastions. Our verdict is that the walls serve mainly to create an appropriate stylistic effect and that the triple dome is both a visual and an engineering triumph.

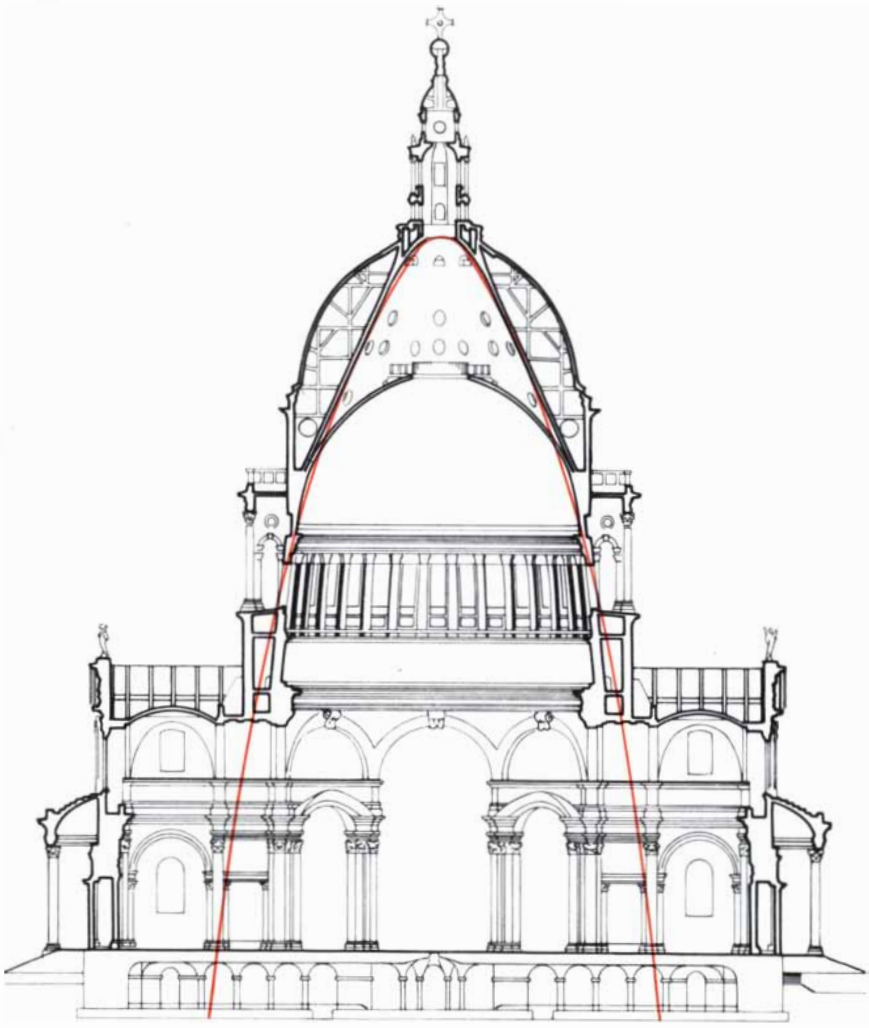
If Wren failed to achieve the elusive merger of theory and practice in his architecture, it was not through lack of purpose. In several commentaries and in the fragmentary *Tracts* his son and grandson compiled in the family *Parentalia,* Wren asserted unambiguously that the science of statics must form the basis of building. In 1713 he summed up his attitude in a report on Westminster Abbey: "It is by due Consideration of the Statick Principles, and the right Poising of the Weights of the Butments to Arches, that good Architecture depends." Indeed, he went beyond this programmatic endorsement and made the practical suggestion that where a column cannot be made stout enough to resist the horizontal thrust of an arch, its resistance can be increased by making it taller and hence heavier.

Wren's views on the place of theoretical mechanics in architecture and on the priority of the utilitarian over the ornamental are set out most clearly in an undated fragmentary work designated *Tract II.* It began with a discussion of the five classical orders of column design (Doric, Ionic, Corinthian, Tuscan and Composite). They had been reduced, he wrote, to "Rules, too strict and pedantick, and so as not to be transgressed, without the Crime of Barbarity; though in their own Nature, they are but the Modes and Fashions of those Ages wherein they were used."

Wren turned next in *Tract II* to an analysis of the arch. "It seems very unaccountable that the Generality of our late Architects dwell so much upon this

ornamental, and so slightly pass over the geometrical, which is the most essential Part of Architecture. For Instance, can an *Arch* stand without Butment sufficient? If the Butment be more than enough, 'tis an idle Expense of Materials; if too little, it will fall; and so for any Vaulting: And yet no Author hath given a true and universal Rule for this; nor hath considered all the various Forms of Arches." Wren concluded that the "true and universal Rule" can be deduced only from the science of mechanics. "What is true," he wrote, "will be shown to be only determinable by the Doctrine of finding the Centers of Gravity in the Parts of the proposed Design." Although the actual rule that Wren then formulated is incorrect because it neglects (among other things) the horizontal thrust of the arch, his analysis was a response in terms of the science of mechanics to the foremost structural problem of the time.

It is only anachronism that has led some to believe Wren's joint interest in science and architecture must have resulted in the successful application of theoretical principles to the solution of building problems. He belonged to the 17th century with its flourishing renewal of the sciences and its practical, enterprising spirit, but he came closest to applied science not in his actual building but in his tracts and fragmentary notes. The program of the Scientific Revolution can be seen in Wren's attempt to formulate a theory of the arch based on statics, in his emphasis on function and utility and in his intellectualized Protestantism that emphasized both hearing and seeing.



THREE-LAYERED DOME of St. Paul's is shown in a cutaway view. The brick cone that constitutes the middle layer supports the outer dome and the heavy lantern. Because the middle layer is conical it is stabilized by the lantern. The forces generated by the lantern on the conical structure do not thrust outward, as they do in the dome of St. Peter's Cathedral, but act mainly in compression. The colored line superposed on the interior of the cathedral has the shape of an inverted catenary (the curve formed by a chain suspended from its ends). Robert Hooke, who was Wren's architectural collaborator after the Great Fire as well as a fellow member of the Royal Society, had noted that an inverted catenary acts only in compression, as a catenary acts only in tension. Since the 19th century several investigators have tried to demonstrate that Wren was guided by Hooke's work on the catenary, but evidence is inconclusive.

Blood, sweat and software

America in crisis. A time for hue, cry and stirring rhetoric.
Yet, if America is to recapture its forward momentum—its sense of destiny—galvanizing words may be spoken in the halls of Congress and on street corners. But the true language of the second industrial revolution will be written in software.

A software manifesto

An overstatement? Perhaps. There is no question, however, that fortune has brought together at precisely this moment both the tools and craftsmen of America's renewal. Computer hardware today is both powerful and affordable. And the flood of new, young management into the decision making ranks of business is, itself, "programmed" for electronics. A stage is set.

Will our resolve be the match of our technology? Will we meet the challenge cast by an era of limited natural resources but unlimited human potential? In brief, what will be the message of the software manifesto? What will we, as businessmen and patriots, have the vision to do?

Forging the future

Manufacturing and Consulting Services, Inc. (MCS) is vitally concerned with answers to these questions. After all, our business *is* the future and we are proud of our reputation as "The Computersmiths™," helping to forge that future with Anvil CAD/CAM (computer-aided design and manufacturing) software, hardware and services.

MCS can also take pride in the fact that at least 75% of all 3-D CAD/CAM software now in use is based on concepts developed by MCS personnel—helping to make CAD/CAM the most dynamic growth area in computer technology. But more important, both for ourselves and our nation, are the philosophical concepts which have made MCS the only full-line CAD/CAM source in the country.

A chance to excel

MCS believes that the wellsprings of the American spirit are imagination and opportunity. That is why MCS is now introducing its new Anvil-4000™ software package as a significantly enhanced replacement for its AD-2000®—the most widely accepted, computer independent CAD/CAM software ever developed. The earlier system had more than proved its worth, allowing a single person at a computer terminal to produce at the rate of up to three product designers, mechanical draftsmen, or manufacturing engineers.

But the new Anvil-4000 program has loftier objectives—not just cost reduction and increased productivity, but a full release of creative innovation; not fewer designers, but an opportunity for each designer to explore more alternatives, to perform a more thorough analysis of the results, to produce a better product.

Anvil-3000 Graphic Resource Stations and the new time-sharing MCS Graphic Resource Centers expand on this theme. As a company, MCS believes that the imagination-releasing potential of CAD/CAM should not be the exclusive preserve of those few concerns which can afford an initial investment of $150,000 or more. America will succeed only if every individual and organization, large or small, has an opportunity to succeed. By making the full power of Anvil-4000 software available at per-station costs of just a few thousand dollars a month, MCS is doing its part to revitalize one of the nation's historic strengths.

These *are* stirring times, and we invite you to join us as, together, we design America's future.

Manufacturing and Consulting Services, Inc., 17942 Cowan, Irvine, CA 92714 (714) 540-3921

The Anvil Chorus

Anvil-4000L
The largest, most advanced CAD/CAM software package available to industry... an integrated design, management information and manufacturing system. It provides all the capabilities for engineers to fully automate the design and manufacturing of a product.

Unlike pre-packaged turnkey systems, Anvil-4000L can run on most of the world's wide-word (24-bit to 64-bit) computers and use virtually *any* of the new graphics display tools. It will easily adapt to new technology as it comes along. Your current hardware may be just fine. At the least you're assured of maximum flexibility... no lock into hardware that may soon become obsolete.

Anvil-4000S
A modularized version of the Anvil-4000L designed for 16-bit minicomputers.

Anvil-3000 Systems
Complete mid-range turnkey CAD/CAM systems which deliver most of the power of the Anvil-4000 software at a hardware/software price of less than $100,000. The system includes a powerful Hewlett-Packard 1000 computer and all of the basic interactive and data-storage peripherals required for CAD/CAM operation. The computer itself has excess capacity for other data processing tasks.

Anvil-3000 Graphic Resource Stations
Semi-autonomous units operating under Anvil software. They have enough local CAD/CAM capability to reduce the host computer I/O traffic by as much as 80%. They can be used to expand existing CAD/CAM systems and serve as the basis for distributed CAD/CAM processing networks.

Anvil Graphic Resource Centers
Time-sharing facilities which bring the full capabilities of the most advanced CAD/CAM system to the small-to-medium user. A time-sharing installation can also serve as a test bed for companies who are evaluating the work of CAD/CAM in their own operations. The investment is limited to the leasing of a Graphics Resource Station and the computer time used.

We need swords and plowshares

CAD/CAM software is creativity-squared: creative computer programs which foster the creative process. The writing of CAD/CAM software is also a "future's business"—directed toward tomorrow's products, processes, productive facilities, and technology.

Add those two facts together and you can see why the current debate over "guns or butter" is of more than passing interest to us. We feel that the choice is dangerously misleading, that we need both swords and plowshares if America is to fulfill its potential as a creative force in the world.

Creativity can flourish only when there is hope in the future and anticipation that individual accomplishments will be rewarded. There can be no hope when the only alternatives seem to be a weakened America, unable to protect its vital interests. We need to stake out a middle ground of self confidence and strength.

But not at the cost of a reduction in the quality of American life. We need to reward individuals and segments of our society for exceptional performance, for above-and-beyond achievements. The profit motive is just as powerful a force at the personal level as it is for corporations—and nations.

We need both, and we can have both. We simply need to be even more imaginative, more creative, more productive. Which is what CAD/CAM is all about.

Dr. Patrick J. Hanratty
President, MCS, Inc.

It's time to put America on the Anvil.

THE AMATEUR SCIENTIST

Anamorphic pictures: distorted views from which distortion can be removed

by Jearl Walker

In an anamorphic picture the image of what is represented is distorted so that it can be perceived only by viewing it in an unusual way. Some paintings and photographs of this kind must be viewed almost from the edge or by means of a reflecting cylinder or cone. Such pictures were an amusement in the Renaissance and remain largely a curiosity today. The methods of creating and viewing them are interesting as part of the larger subjects of the nature of perspective and of visual perception.

Recently I received several examples of anamorphic photography made by David G. Stork, a graduate student of visual perception at the University of Maryland at College Park. He employs a black-and-white negative or a color slide to make a print that is anamorphic. Normally a print is made by placing the negative in an enlarger. Light directed through the negative passes through a lens and falls on the print paper, which is developed for the final photograph. Stork mounts a shiny cone in the path of the light so that it reflects the rays to parts of the print paper where they would not normally go. At the center of such a print, of course, is an empty area corresponding to the circular base of the cone.

Usually the resulting print is unrecognizable, but the original perspective can be regained with the aid of the same cone. It is placed at the center of the print and the observer looks down from an appropriate height, usually with one eye. The sides of the cone show a reflection of the print with the perspective of the original scene restored.

The image from the negative is distorted in the print in two ways. The parts of the image near the center of the negative appear near the outside of the print. Conversely, the parts near the outside of the negative appear near the base of the cone in the print. This radial inversion of the center and the perimeter gives rise to the curious pattern in the print.

The other type of distortion of the image affects its angular width around the center of the print. The parts of the image cast near the outside boundary of the print are spread in a circle around the center of the print more than the parts cast near the base of the cone. As a result most of the normal clues to perspective are lost.

The equipment needed for Stork's anamorphic photography can be quite simple. Stork works with a heavy metal cone machined to size. You could substitute a sheet of reflecting Mylar formed into a cone. (These sheets are available at hobby shops and kite stores.) In addition to the standard *f*-stop diaphragm of the enlarger you need a means of controlling the width of the light beam above the cone. Another *f*-stop diaphragm mounted on a ring stand will do. A substitute can be a series of cardboard sheets each with a hole of different diameter in it. By changing sheets appropriately the photographer can control the exposure almost as well as he could with an *f*-stop diaphragm.

The cone must be positioned properly over the print paper while the photographer is working in the dark. Stork made a cardboard mask to fit into the holder for the paper. At the center of the mask is a hole of the same diameter as the base of the cone. By feel Stork can position the cone in the hole and then lift the mask off without disturbing the cone.

The procedure for exposing the paper is as follows. Put the negative into the enlarger, turn off the lights, remove a piece of print paper from its box and put it in the holder of the enlarger. Lay the mask for the cone over the paper. After you have positioned the cone in the hole lift the mask off the paper without moving the cone. Turn on the light in the enlarger to expose the paper.

Several exposures are needed for one print because the light falling on the paper near the base of the cone is more concentrated than the light that strikes the outer boundary of the paper. The time of exposure for the paper near the base is normal; for the paper near the outer boundary the time should be between 15 and 20 times longer. Stork manages to achieve approximately the right exposures by making the first one with a relatively large aperture that he has mounted on the ring stand. The aperture is either the variable *f*-stop diaphragm or a piece of cardboard with a large hole in it. The last exposure is made with a comparatively small aperture that confines the light to the apex of the cone. Since that region sends light toward the outer boundary of the print paper, only the perimeter of the paper is exposed.

Starting with the large aperture, Stork steadily reduces the aperture size so that less of the cone is illuminated. (Either the variable *f*-stop diaphragm is narrowed or the cardboard is replaced with another piece having a smaller hole.) He continues until the exposed region is confined to the perimeter of the paper. It is important that in the first exposure the circle of light not extend beyond the base of the cone. If the circle is larger, light can reach the area on the paper around the cone without having been reflected from the cone. The print would be ruined, since it would not be a true image resulting solely from reflection.

A fair amount of practice is needed to master varying the size of the aperture and making the exposures. Beginners should work with black-and-white print paper so that experimenting with the exposures does not become too expensive. Stork recommends that when you are ready to make color prints from color slides, you work with Ilford Cibachrome color print paper.

According to Stork, almost any scene will do for a color slide. A good goal is to achieve an anamorphic print that resembles something totally different from the original scene. Better prints result if the slide has a relatively dark central area. Light from that area of the slide falls mainly on the apex of the cone. No cone, not even a machined one, will have a perfect point. Hence light reflected from the apex will be distorted in an uncontrolled way.

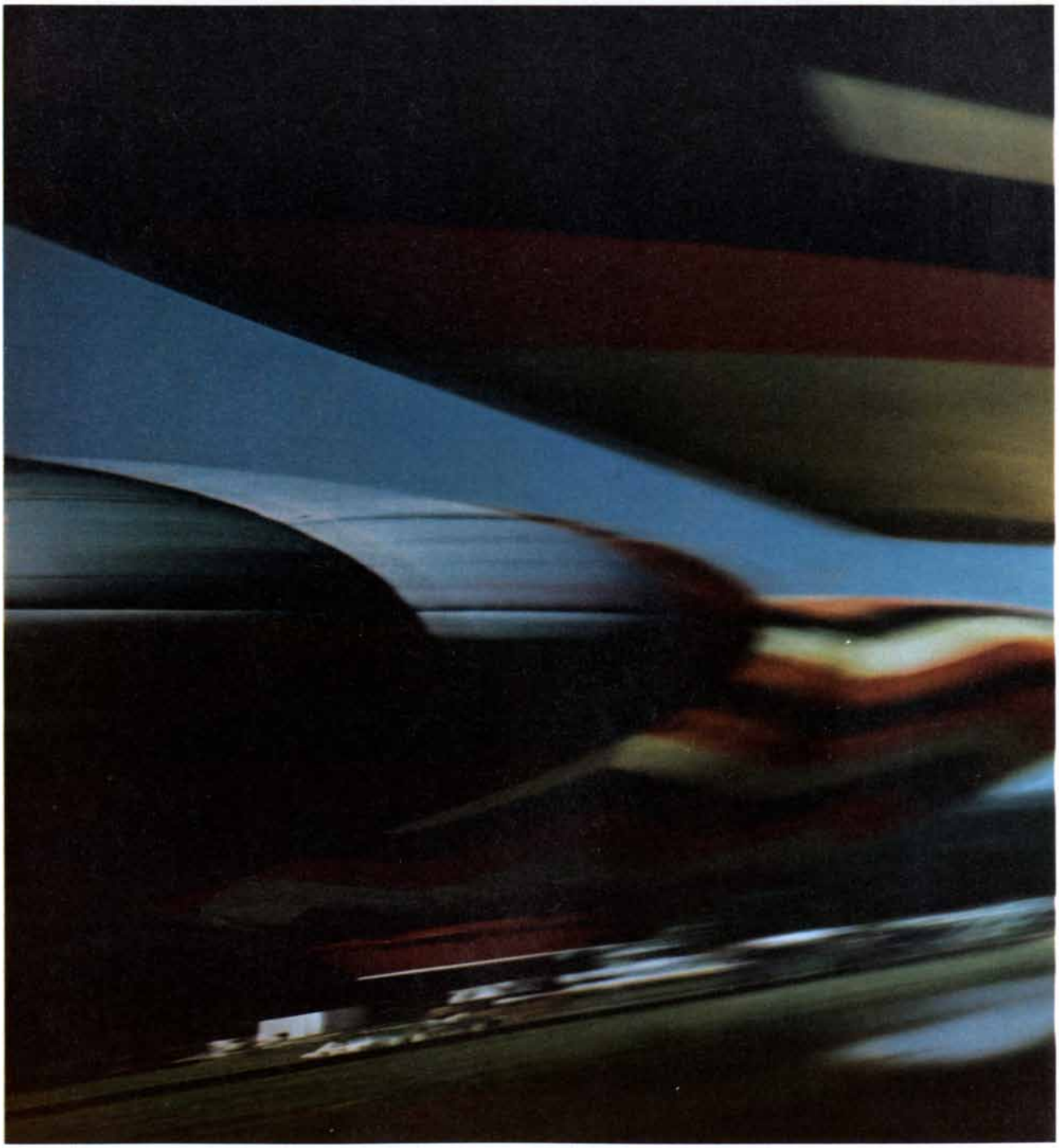
If you could place the cone on the finished photograph exactly as it stood on the print paper during the exposure, you could undo this distortion. Such precise repositioning is arduous and not worth the trouble. Another difficulty with the apex is that the light reflected from it to the print paper is comparatively faint. You can avoid both problems by having a slide with a relatively dark central area so that what is reflected from the apex makes less difference in the print.

Stork suggests that if you are unable to machine a heavy metal cone or get one machined, you should mount Mylar on a solid wood cone. You can add weight to the cone by drilling a hole into the base and inserting a short length of metal rod into it. It is obviously important that the cone not move in the course of the repeated exposures.

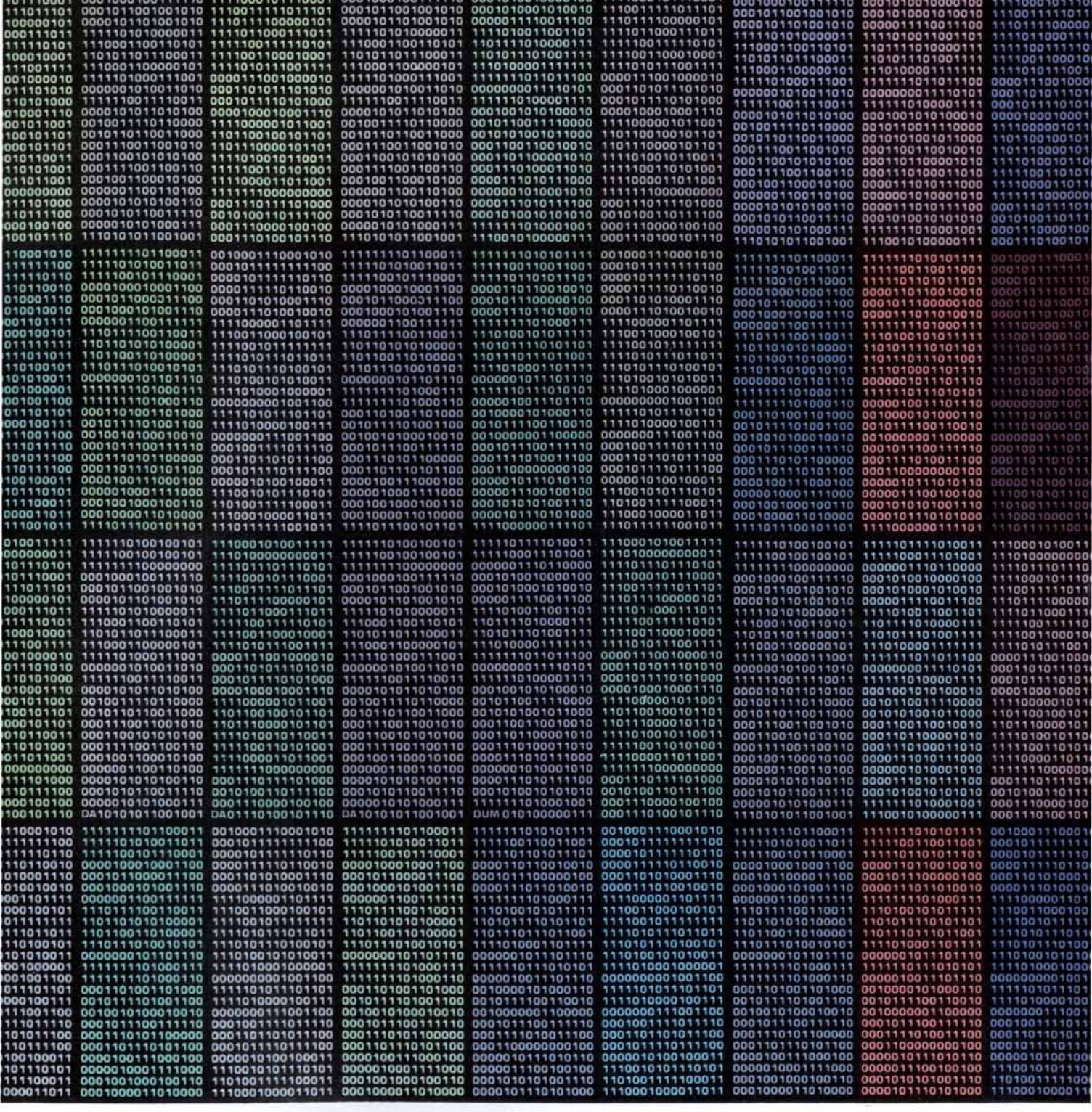
A recent article in this magazine described similar distortions produced by reflections from plane and curved mirrors [see "Mirror Images," by David Emil Thomas; SCIENTIFIC AMERICAN, December, 1980]. Thomas presented a classification of the distortions that depends on the type of mirror employed. He defined two perpendicular axes on the surface of a mirror. The type of reflection from the surface can be classified by the type of curvature the axes have. For example, the side of a reflecting cylinder has one axis along the straight section of the surface (the axis runs along the side from one base to the other) and another axis (perpendicular to the first one) that travels around the circumference of the cylinder. The surface is a combination of a plane mirror (the straight axis) and a convex mirror (the curved axis). The reflections along the straight axis behave as they would with a plane mirror, and the ones along the curved axis are the same as those from a convex mirror.

In Stork's setup the light is reflected from the side of a shiny cone, which can also be considered a combination of a plane mirror and a convex one. The axis for the plane mirror lies along the side, running from the base to the apex. The axis for the convex mirror runs around the cone, perpendicular to the other axis and parallel to the base of the cone.

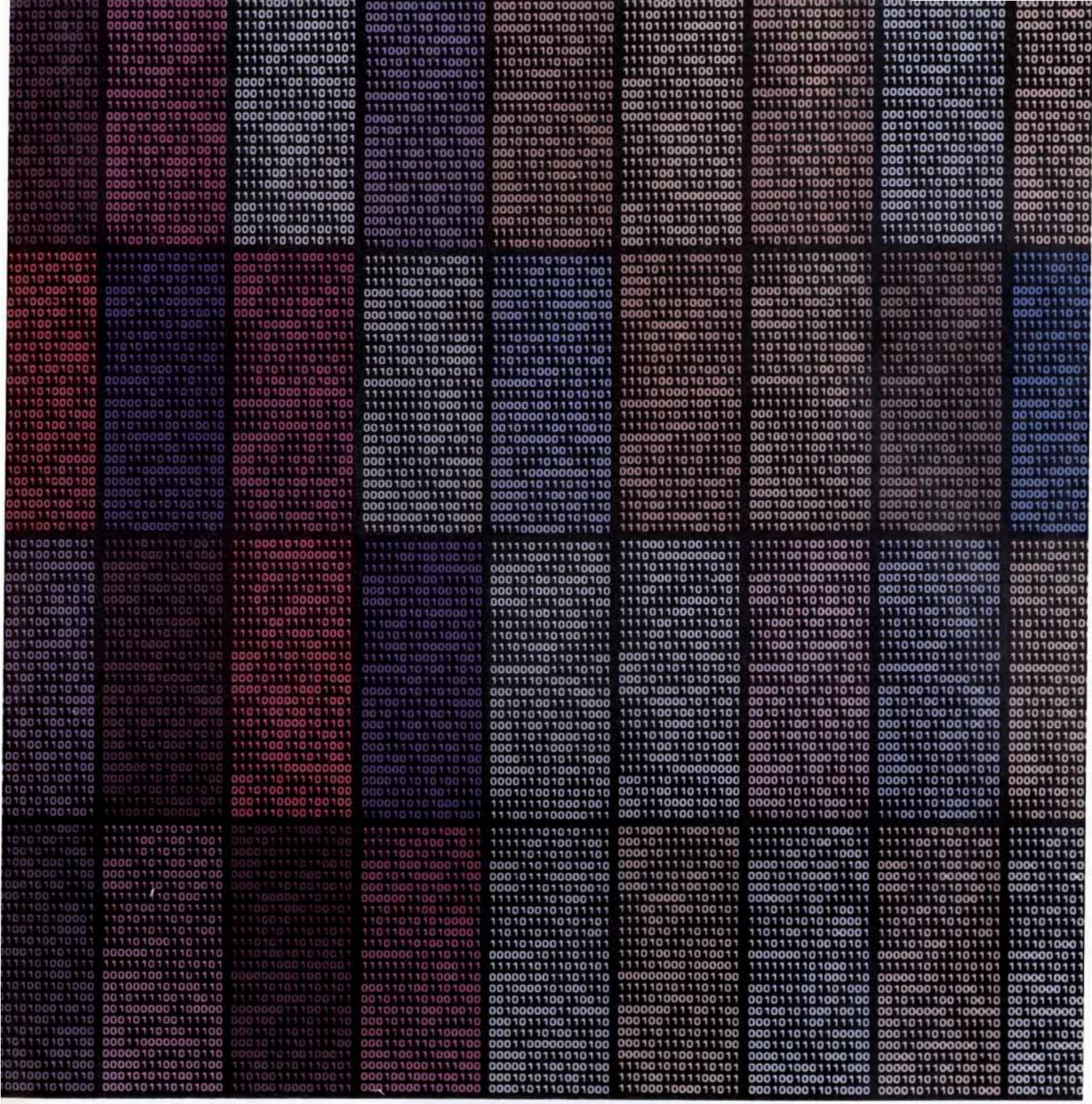
The cone introduces a complication,



David G. Stork's anamorphic photograph of hot-air balloons



These are the first four notes of Beethoven's Fifth Symphony.



You've never seen musical notes like these before, because they've never been recorded like this before. The 3M Digital Mastering System actually uses a computer to record every nuance of the music. So when reproduced, every note can be heard with uncanny crispness and clarity.

Hearing is something we pay a great deal of attention to at 3M. It's not just the technology of hearing that concerns us, but the business of hearing.

By listening to people's needs, we've responded with new ideas and innovative products like our Digital Mastering System. It's a product that not only advances the recording industry but is a step forward in computer technology.

In fact, 3M has advanced over 400 products in the communication arts field alone.

If you think you might have an application for our technologies and products, write us today for a free 3M Communication Arts Brochure: **Department 080707/3M, P.O. Box 4039, St. Paul, MN 55104.**

Or better yet, let us hear from you right now. **Call toll-free: 1-800-323-1718,** Operator 368. (Illinois residents call 1-800-942-8881)

3M hears you...

3M

however, which is that the curvature of the convex axis changes along the side. Toward the apex the axis is highly curved but toward the base it is less curved. Still, the reflections leaving any small area on the side of the cone can be thought of as coming from a combination of a plane mirror and a convex one. The axis for the plane mirror is responsible for the exchange of the center and the perimeter of the slide when the print is made. The axis for the convex mirror is responsible for the spread of the image into a circle around the center.

Thomas demonstrated how a mirror can invert and reverse an image with respect to the object casting the image. Inversion refers to an exchange of placement between the object and its image. The exchange can be of far and near, left and right or top and bottom. Plane and convex mirrors invert only far and near.

Suppose you hold a meter stick perpendicular to a plane mirror in such a way that the end with the zero mark touches the mirror. When you stand at the other end of the stick, the end closest to you in the mirror image is the one with the zero mark. The reflection from the plane mirror inverts far and near. A convex mirror does the same.

A plane mirror does not invert left and right or top and bottom. You might think left and right are inverted, but they are not. Face the mirror and hold a meter stick with your left hand on the zero mark and your right hand on the one-meter mark. Look at the reflection of the stick. The image of the zero mark is also on your left (yours, not the image person's) and the one-meter mark is on your right. Nothing has been exchanged.

Left and right are exchanged in the image person. The zero mark is held by the right hand, the one-meter mark by the left. Thomas calls such an exchange a reversal. Plane and convex mirrors cause a reversal and one inversion (of

Stork's anamorphic photograph of a woman, made with a reflecting cone

NEW DATSUN MAXIMA

FOR THE LUXURY-MINDED WHO LONG TO BE DATSUN DRIVEN

There was a time when you felt proud to own a "luxury car." Today, if you're dragging around a big, bulky sedan, you're apt to feel embarrassed. Or guilty. Datsun's beautiful solution is the new 810 Maxima. Here is responsible luxury; pure performance in perfect harmony with the environment.

POWER OF THE 240-Z

The new Maxima wraps luxury around the fuel-injected engine of the 240-Z. Converts a one inch tap of the accelerator into hundreds of feet of instant thrust.

You control it... with power-assisted rack-and-pinion steering, fully independent suspension (sedan only), and power disc brakes.

LUXURY OF A MERCEDES

One look inside and you'll know where the word MAXIMA came from: Maximum luxury–so complete there are no options. Air conditioning. A four-speaker stereo. Cassette deck. Automatic transmission. Cruise control. And computer readout.

SOPHISTICATION OF A CADILLAC

No Cadillac owner will miss his power controls in a Datsun Maxima. Windows, door locks, dual mirrors, antenna and, in the sedan, a sky roof.

ABOUT THE SIZE OF A BMW 528i AT LESS THAN HALF THE PRICE.*

Maxima is the roomiest, most comfortable Datsun ever created. There's more front headroom and legroom in the sedan than the BMW 528i–plus a very generous trunk. For even more room, examine the new Datsun Maxima 5-door wagon.

QUALITY AND MILEAGE OF A DATSUN.

Maxima delivers 22 estimated MPG, 27 estimated highway. Use estimated MPG for comparison only. Your mileage may differ depending on speed, trip length, and weather. Highway mileage will probably be less.

*Manufacturer's suggested retail price excluding title, taxes and destination.

DATSUN
WE ARE DRIVEN

THE NAME OF QUALITY
NISSAN
NISSAN MOTORS

far and near). Any mirror that gives rise to an odd number of inversions reverses handedness.

I applied Thomas' analysis to the setup by Stork. In place of the slide I visualized a large letter *F* held by a small person. The variations in handedness could be followed by noting how the image of the *F* appears in the image of the person. For this purpose I darkened a square in one corner of the *F*. In my imaginary demonstration light rays traveled from the *F* and the person through the lens of the enlarger and were then reflected from the surface of the cone onto the print paper.

The orientation of the *F* is altered by the lens. (I shall continue to use for the lens the terminology employed by Thomas for mirrors.) The lens inverts left with right and top with bottom. The result is an image of the *F* rotated 180 degrees from its normal orientation. (I am not being fully precise here. The image is not really present in the space between the lens and the cone because the rays of light passing through the lens have not yet crossed one another to create a focused image. If the image were in focus on a card inserted into the light, you would see what I have described.) The image also has a reversal of handedness. The darkened square is held by the left hand of the person in the initial arrangement, but it is held by the right hand in the image of the person cast by the lens.

The reflection of the *F* by the cone is essentially the same as the reflection from a combination of a plane mirror and a convex one except in one important respect: the reflection falls on a flat surface before the image reaches you. The rules governing inversion and reversal of an image of yourself in a mirror must be modified. The final image is still inverted top with bottom compared with the original orientation. The left-right inversion caused by the lens, however, is canceled by the reflection from the mirror onto the photographic paper, as is the reversal of handedness. The final image has the person holding the darkened square with the left hand just as in the initial setup. In sum, the image has had one inversion (top-bottom) and no reversal.

When the final print is viewed by means of a reflection in the side of the cone, the reflection is in effect from a combination of a plane mirror and a convex one. The rules governing the reflection, however, are not the same as those that hold when you see a reflection of yourself with such a combination of plane and convex mirrors. The reflected image is a virtual one, since the rays that are reflected do not cross to form a real image. Instead they diverge when they leave the side of the cone.

The observer's eye collects the diverging rays and focuses them onto the retina. The virtual image created by the cone does not have the same orientation as the image in the photographic print. Left and right are inverted. The reflection also reverses handedness. In the print the person is holding the darkened square with his left hand. When the image is reflected in the side of the cone, the square is held by the image person's right hand.

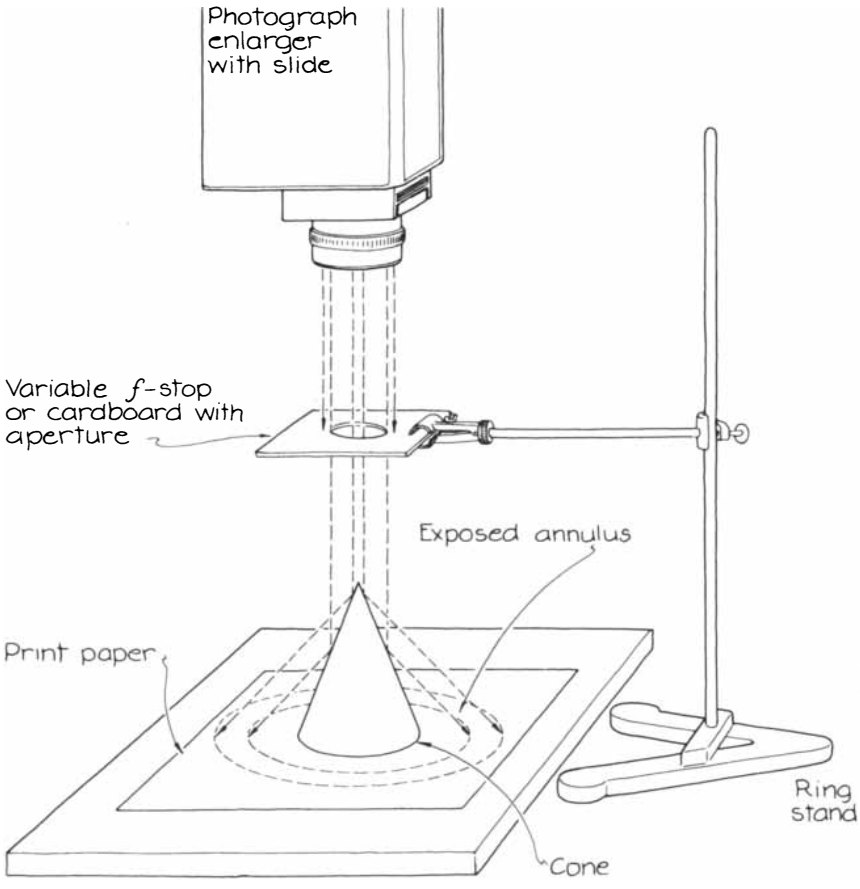
Stork has made anamorphic prints with two other setups. In one he reflects the light from the enlarger with a plane front-surface mirror so that the rays fall obliquely across the print paper, which is placed off to the side rather than directly under the enlarger. The result is a distorted print, similar to some anamorphic paintings, that must be viewed almost from the edge in order to re-create the proper perspective. For example, the hot-air balloons in the illustration on page 177 are clearer if the photograph is viewed (with one eye) from a slant angle or reflected in a plane mirror held tilted at one edge of the photograph. The sequence of inversion and reversal of an image in the exposure of the photograph is the same as it is with the cone.

To make the proper exposures on the anamorphic print made with a plane mirror you must adjust their duration across the width of the print paper. The areas farther from the mirror require more exposure than the areas closer to it. Stork adjusts the exposure by inserting a sheet of cardboard between the mirror and the paper. He begins the exposure with the cardboard fully blocking the light. Then he gradually withdraws the cardboard until finally the edge of the paper nearest the mirror is exposed. In this way the farther regions are exposed longer than the nearer ones.

Stork cautions that much experimentation and patience are required with both a reflecting cone and a plane mirror to make a satisfactory print. The focusing is never quite right. If one part of the paper is in focus, another part will not be. Stork generally adjusts the focus for the middle of the anamorphic print or for any particularly important feature of the print.

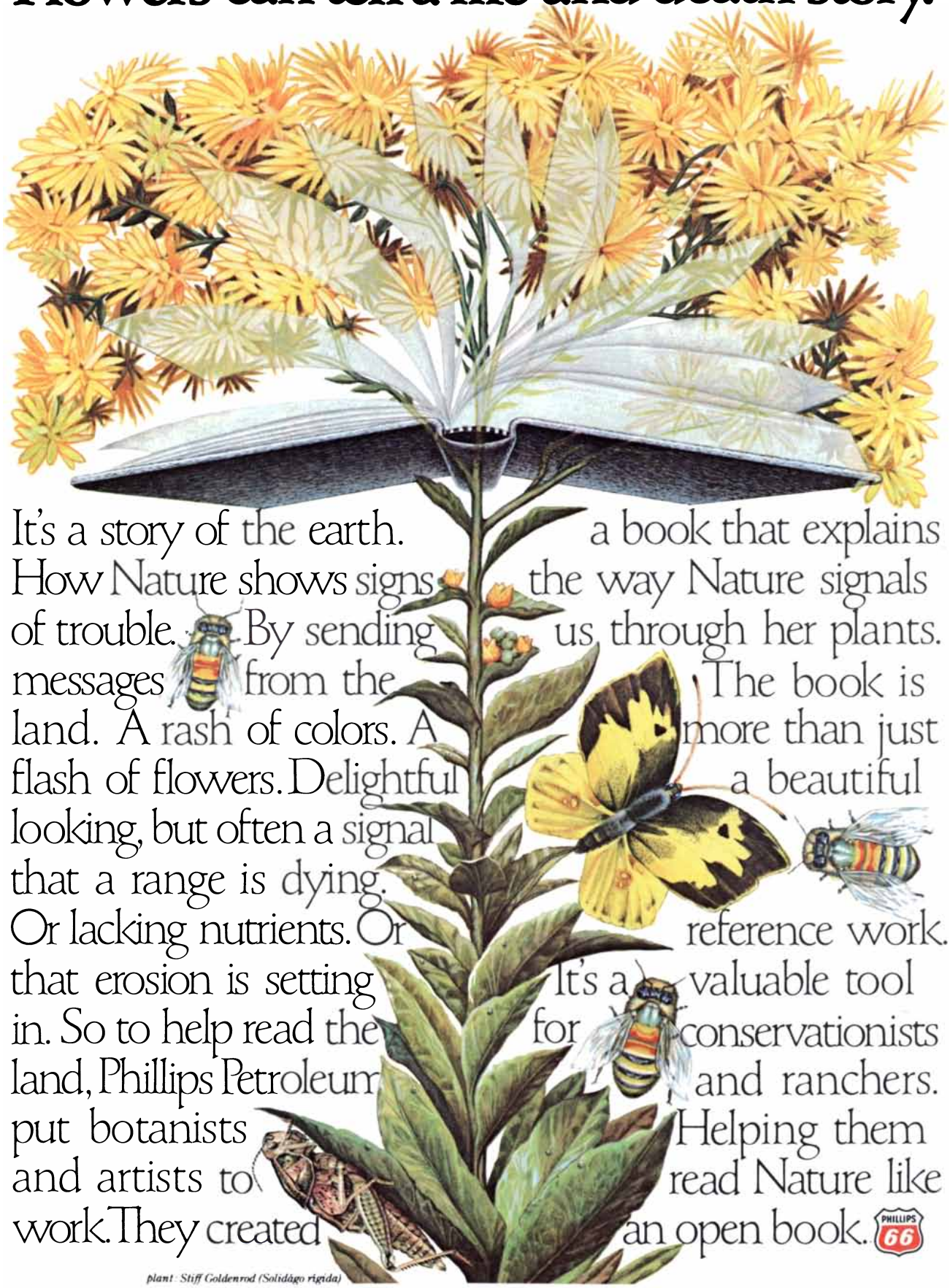
He also adjusts the *f*-stop diaphragm in the enlarger. A small aperture (a larger *f* number) gives more depth of focus. The smallness of the aperture means that longer exposures are needed because less light is reaching the print paper, but the improvement in focus is worth the extra effort.

Stork also employs a shiny cylinder to



A setup for making anamorphic prints

Flowers can tell a life and death story.



It's a story of the earth. How Nature shows signs of trouble. By sending messages from the land. A rash of colors. A flash of flowers. Delightful looking, but often a signal that a range is dying. Or lacking nutrients. Or that erosion is setting in. So to help read the land, Phillips Petroleum put botanists and artists to work. They created a book that explains the way Nature signals us through her plants. The book is more than just a beautiful reference work. It's a valuable tool for conservationists and ranchers. Helping them read Nature like an open book.

PHILLIPS 66

plant: Stiff Goldenrod (Solidago rigida)

Everything about David Plastow represents his company's philosophy. Which is why he wears a Rolex.

David Plastow is the custodian of a long and famous engineering tradition.

He is the Chief Executive of Rolls-Royce Motors, and indeed, his manner and personal appearance exactly reflect the ethos of that company. That of the skilled engineer.

Plastow takes a personal interest in any modification, however small.

"All our developments at Rolls-Royce are always evolutionary rather than revolutionary," he says.

"We are a highly personal business, and both our craftsmen and our customers have clearly defined ideas about what a Rolls-Royce should be. But while we don't tamper with those fundamental ideas, we are, of course, constantly searching for improvement. For instance, years ago, the gear selection on a Rolls-Royce car became completely electronic. But, a driver likes to 'feel' that the gear selection lever is doing something . . . so we engineered the 'feel' back into it – so it's satisfying to use."

David Plastow recognizes the similar philosophy behind the watch he wears.

"It's a Rolex Oyster Datejust. I'm told that the engineering concept of the Oyster case first appeared in 1926.

"Obviously this watch has changed and improved over the years but Rolex has stayed with the basic idea because it was a very good one. It's extremely tough, very reliable, and superbly engineered. After 50 years of development it's almost perfect".

Which, from the man who makes the finest cars in the world, is quite a compliment.

ROLEX

The Rolex Datejust Chronometer. Available in 18kt. gold, steel and gold combination or stainless steel, with matching bracelet.

Write for brochure. Rolex Watch, U.S.A., Inc., Dept. 294, Rolex Building, 665 Fifth Avenue, New York, N.Y. 10022.
World headquarters in Geneva. Other offices in Canada and major countries around the world.

reflect the light rays to the print paper. The light from the enlarger is reflected from a plane mirror and then from the cylinder before it falls on the paper. To make the light from the plane mirror fall only on the cylinder and not directly on the paper, Stork blocks off part of the light so that only the cylinder is illuminated.

The procedure for exposing the paper is similar to that for making an exposure with a cone. In darkness Stork mounts the paper in its holder and lays on the paper a placement mask that has a semicircular cutout into which he positions the cylinder. Then he lifts the mask off the paper before he illuminates the cylinder. If the mask is shaped so that it does not block any of the light reflected from the cylinder, it can be left on the paper during the exposure.

The geometry of the cone setup is simple. Consider a cone that has an apex with a known angle. I shall call the half angle of the apex ϕ (phi). It should be less than 45 degrees; a half angle of 30 degrees works well. Next consider a ray of light that is initially vertical. It strikes the side of the cone at an angle to the surface that is also ϕ. A light ray always reflects from a surface at the angle of its approach. Therefore in the cone setup the angle between the reflected ray and the surface is also ϕ.

Suppose the center of the cone's base is considered as being the center of a coordinate system. If the cone had not been in the way, the vertical ray would have struck the paper at a distance r from the center. With the cone in place the reflection redirects the ray so that it now strikes the paper at a distance greater than r from the center. The new distance is given by the equation in the top illustration on the next page.

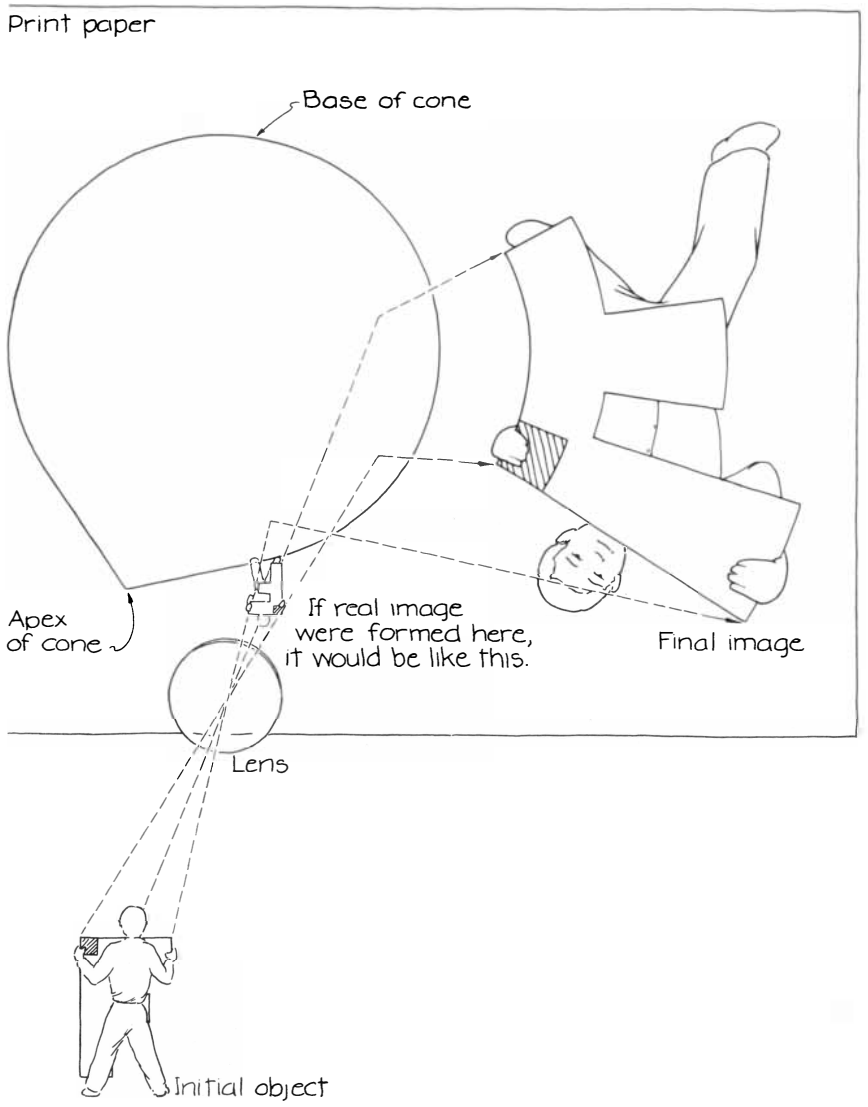
One can predict the appearance of the anamorphic print by plotting the new positions of each section of the original scene in the slide. With the cone displaced focus the slide on a sheet of ordinary paper. Choose a small area of the image and measure its distance r from the center point (where the center of the cone's base will be once it is positioned on the paper). Then use the equation to determine the new distance r' for that section of the scene. A simple procedure for measuring the half angle of the cone's apex is to take the inverse sine of the radius of the base divided by the distance along the side of the cone (from the apex to the base).

To map the original scene into the anamorphic one completely position a circular grid on the paper. Choose a section of the original scene that is at a distance r from the center and lies within a certain angle around the center. That area of the picture should be reproduced at a new distance r' and within the same angular extent around the center. If the new location is adjacent to the base of the cone, the new area will resemble the original area. If the new location is away from the base, the new section will appear to be stretched out along a section of a circle around the center. This procedure for mapping normal drawings into anamorphic ones has been described in these pages by Martin Gardner [see "Mathematical Games," SCIENTIFIC AMERICAN, January, 1975].

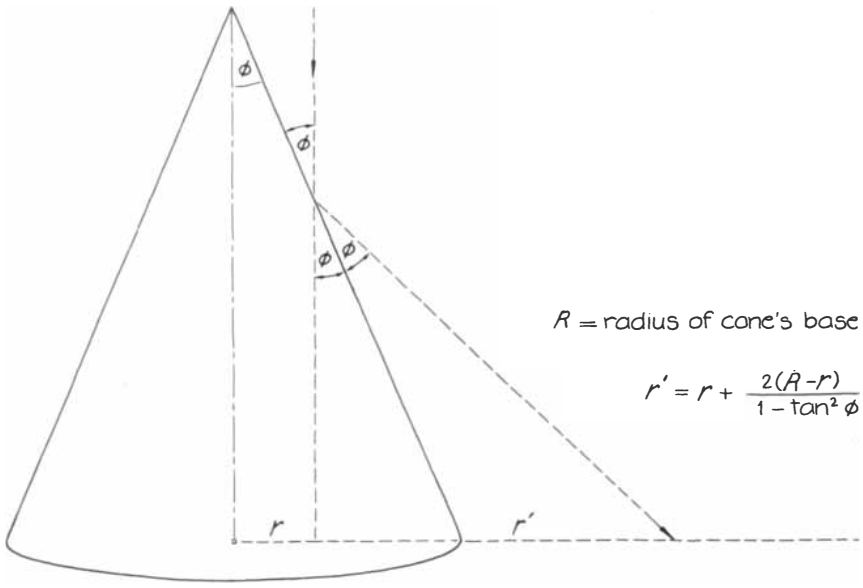
A faster way is to have a computer plot the anamorphic pattern. The input must include the coordinate positions of all the sections of the original scene: the initial radius R and the angular extent of the section, say one degree. The computer will compute the new positions and (if a plotter is attached) plot the results within one degree at the new radius.

Waldo Tobler of the University of California at Santa Barbara created an anamorphic computer-graphical picture from a drawing of a woman Stork had made on graph paper. Patricia S. Irle, one of Tobler's students, entered 1,450 data points for the drawing into a Tektronix 4051 computer. The original drawing was printed out by the computer. Then the program effectively reflected the image from a cone by applying the transformation equation to determine the new distance of each point from the center of the picture. Finally the distorted picture was printed out by the computer. When Stork positions his cone at the center of the printout and peers down into the reflecting surfaces of the cone, he sees an image of the original drawing. Without the aid of the cone he would see only the anamorphic version.

The same kind of transformation can be done with a home computer. If you do not have a printer, the transformation can be displayed directly on the monitor screen. Write a program that "draws" a design close to the center of



Reversals and inversions of an image



The geometry of the reflection

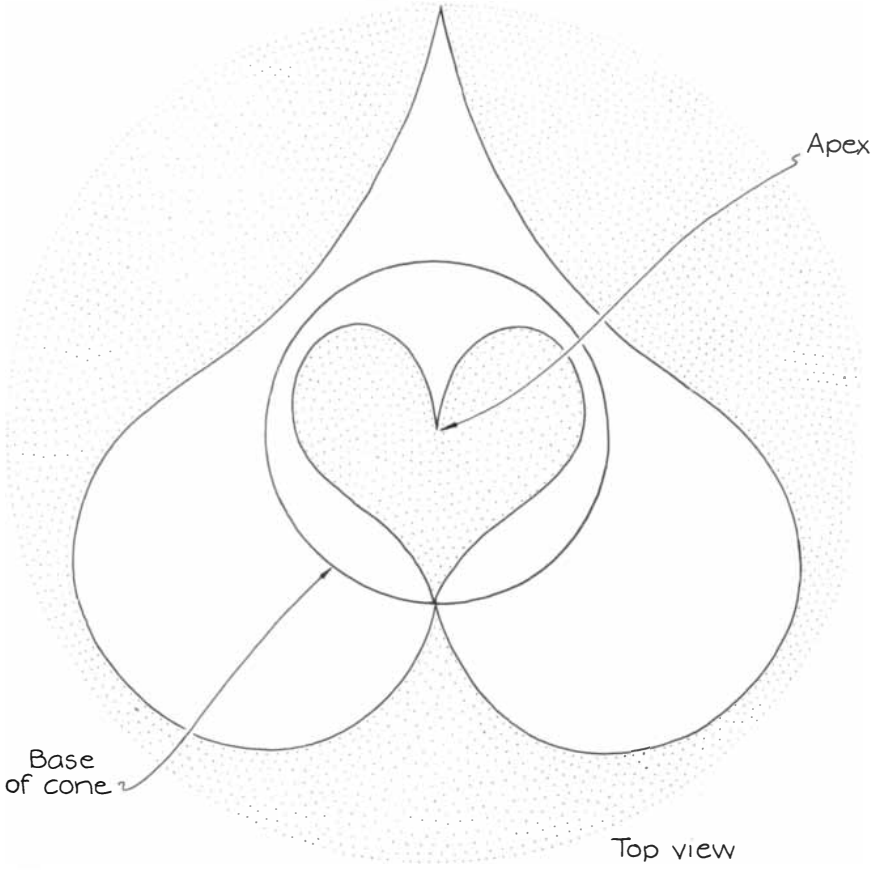
the screen. The pattern will actually be composed of many small rectangular elements that are turned on by the computer. Since the positions of these elements will have to be remembered by the computer, you must store the positions of the rectangles.

For each rectangle have the program apply the equation for the transformation by a cone. New rectangles, farther from the center of the screen, will be turned on. This new set constitutes the anamorphic picture. Place a cone at the center of the screen and look into its sides. In the reflections you will see a likeness of the original drawing.

The reproduction is not complete because the monitor lights up rectangles, not small dots. Some of the details of a curve are lost. The rectangularity of the elements also means that the angular extent of part of the design will not be mapped faithfully at the new radius. Still, the speed of the computer makes it possible to analyze many patterns. Other anamorphic pictures can be studied if the equation for the transformation is altered. Examples can be found in the article by Andy A. Zucker listed in the bibliography for this issue [*page 188*].

Some types of anamorphic distortion are particularly interesting. For example, Stork sent me the results of three simple drawings that had been distorted anamorphically in a cone. In one he had drawn a heart (in Valentine style). He colored the heart red and the area around it white. After the transformation what he saw was an inverted white heart with a red surround. The color and orientation were reversed because the cone exchanged the center of the sketch for the perimeter and vice versa.

For another example Stork drew a "happy face": a circle with two dots for eyes and a semicircle for the mouth. The transformation results in another happy face of the same orientation. I noted that if the mouth was originally almost a straight line, the transformation to the anamorphic version made the face happier. Can you think of any design for which transformation by a cone results in a recognizable design with a transformed meaning?



A heart and an anamorphic version of it

Last November I described a visual illusion that arises when you scan a television screen horizontally. When you scan the screen properly, several images of the picture on it can be seen floating in the air off to the side of the set. The illusion depends primarily on two things: the persistence of vision and the rapid creation and disappearance of the picture on the screen.

Brian Glassner, a student at Cleveland State University, pointed out that the same illusion can be seen in the electronic displays on some types of calculator and on many of the newer pinball machines and other electronic games. If the display consists of light-emitting diodes, it is likely to be pulsed. The pulsation is too fast for the eye to detect, but it does mean the display is constantly appearing and disappearing.

If you sweep your eyes rapidly across your field of view, the flashing light will excite a series of areas on your retina. After the sweep each area in the series provides a separate image of the display. As a result you perceive the images as floating to the side of the actual display.

I described another visual illusion in

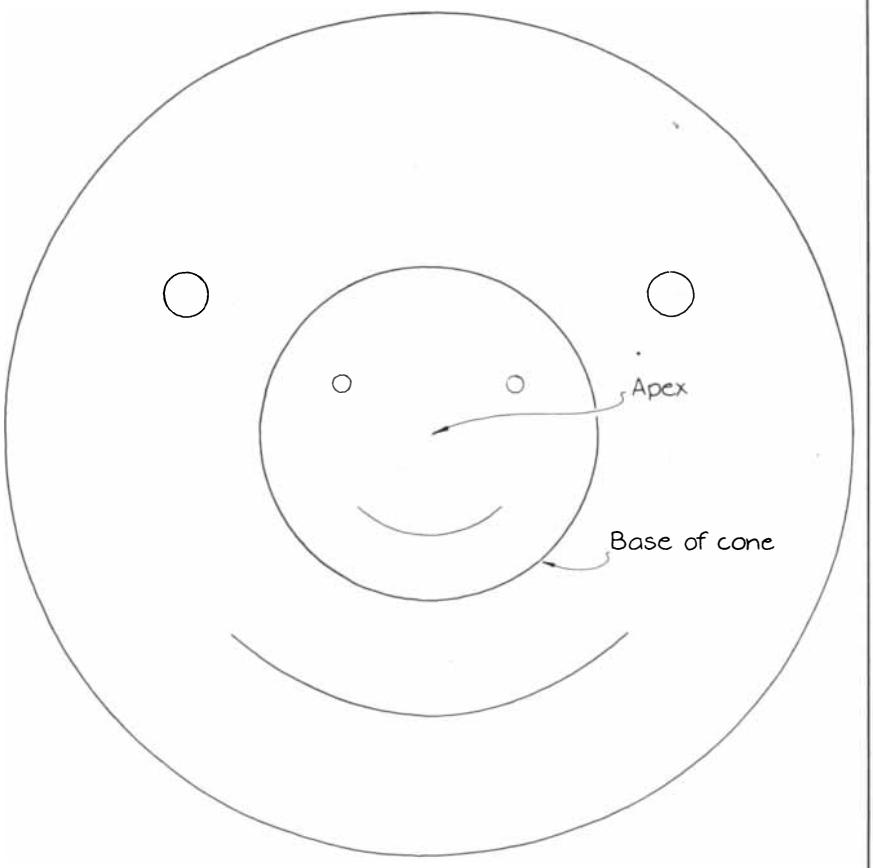
this department for March, 1978. If a swinging pendulum is viewed while one eye is covered with a dark but not opaque filter, the pendulum appears to move in an ellipse. This illusion, known as the Pulfrich illusion, has been attributed by some investigators to the delay of perception by the covered eye. Apparently the reduced intensity of the light reaching that eye results in a delay of the signal transmitted to the brain. The resulting discrepancy of information from the two eyes forces the observer to interpret the pendulum as being closer or farther away than it actually is. The difference between the apparent distance and the true distance varies with the speed of the pendulum, being less when the pendulum slows as it reaches the end of a swing.

Recently Jerry Harris of Worthington, Ohio, described a related illusion that can be produced with a display of a Lissajous figure on an oscilloscope. He connected a signal generator to the vertical input of the oscilloscope and another signal generator to the horizontal input. Sine-wave signals from the generators swept the oscilloscope trace vertically and horizontally.

When the frequencies of the signals were adjusted properly, the trace on the screen formed a Lissajous figure. If the frequencies were slightly displaced from the optimum values, the figure still appeared, but it gradually varied. An observer will often interpret the varying display as a three-dimensional object revolving slowly about a fixed axis. The perspective of the seemingly three-dimensional object appears to vary.

When the frequencies match, the figure might be a circle. With a slight mismatch the circle appears to revolve about an axis on the screen. Sometimes the full plane of the circle is seen. At other times the circle is seen on edge. The direction of the rotation is arbitrary; some observers can switch rapidly from one interpretation of the direction to the other.

Harris modified this classic demonstration by holding a dark filter (a pair of partly crossed polarizing filters) over one eye to delay that eye's perception of the oscilloscope display. The delay introduced depth into his interpretation of the object, presumably as it does in the Pulfrich illusion. The appearance of depth removed the ambiguity of the rotation of the apparent three-dimensional object, which was seen to rotate about a vertical axis in the plane of the screen. When Harris tilted his head so that his eyes were along a vertical axis, the rotational axis of the object appeared to be horizontal. If he shut either eye, the three-dimensional effect disappeared.



A "happy face" and an anamorphic version of it

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

THE CHARACTER OF PHYSICAL LAW. Richard P. Feynman. The MIT Press, 1967.

THE CONCEPTUAL DEVELOPMENT OF QUANTUM MECHANICS. Max Jammer. McGraw-Hill Book Company, 1973.

THE MANY-WORLDS INTERPRETATION OF QUANTUM MECHANICS: FUNDAMENTAL INTERPRETATION. Edited by Bryce S. DeWitt and Neill Graham. Princeton University Press, 1973.

THE FEYNMAN LECTURES ON PHYSICS. Richard P. Feynman, Robert B. Leighton and Matthew Sands. Addison-Wesley Publishing Co., 1973–75.

OTHER WORLDS. Paul Davies. Simon and Schuster, 1980.

VALUES AND ATTITUDES OF THE POLISH PEOPLE

EGALITARIAN ATTITUDES OF WARSAW STUDENTS. Stefan Nowak in *The American Sociological Review,* Vol. 25, No. 2, pages 219–231; April, 1960.

ROLE AND LIMITS OF THE FUNCTIONAL APPROACH IN FORMULATION OF THEORY OF ATTITUDES. Stefan Nowak in *Understanding and Prediction: Essays in the Methodology of Social and Behavioral Theories.* D. Reidel Publishing Co., 1976.

LIKE FATHER, LIKE SON? Stefan Nowak in *Polish Perspectives,* Vol. 19, No. 7–8, pages 10–20; July–August, 1976.

THE SALINITY OF RIVERS

REPORT ON WATERLOGGING IN RELATION TO IRRIGATION AND SALINITY PROBLEMS, LAHORE, PAKISTAN, 16–28 NOVEMBER, 1964: REPORT NO. 1932. Edited by Gasp van 't Leven. Expanded Program of Technical Assistance, Food and Agriculture Organization of the United Nations, 1965.

SALINITY PROBLEMS AND MANAGEMENT IN RIVER SYSTEMS. Arthur F. Pillsbury and Harry F. Blaney in *Journal of the Irrigation and Drainage Division, American Society of Civil Engineers,* No. IR1, Proceedings Paper 4733, 1966.

WATER QUALITY CRITERIA: REPORT OF THE NATIONAL TECHNICAL ADVISORY COMMITTEE TO THE SECRETARY OF THE INTERIOR. Federal Water Pollution Control Administration, 1968.

THE ATMOSPHERE OF VENUS

SCIENCE, Vol. 203, No. 4382; February 23, 1979. (Special issue devoted to Pioneer Venus.)

SCIENCE, Vol. 205, No. 4401; July 6, 1979. (Special issue devoted to Pioneer Venus.)

STRUCTURE AND CIRCULATION OF THE VENUS ATMOSPHERE. G. Schubert, C. Covey, A. D. Del Genio, L. S. Elson, G. Keating, A. Seiff, R. E. Young, J. Apt, C. C. Counselman III, A. J. Kliore, S. S. Limaye, H. E. Revercomb, L. A. Sromovsky, V. E. Suomi, F. Taylor, R. Woo and U. von Zahn in *Journal of Geophysical Research,* Vol. 85, pages 8007–8025; December, 1980.

THE THERMAL BALANCE OF VENUS IN LIGHT OF THE PIONEER VENUS MISSION. M. G. Tomasko, P. H. Smith, V. E. Suomi, L. A. Sromovsky, H. E. Revercomb, F. W. Taylor, D. J. Martonchik, A. Seiff, R. Boese, J. B. Pollack, A. P. Ingersoll, G. Schubert and C. C. Covey in *Journal of Geophysical Research,* Vol. 85, pages 8187–8199; December, 1980.

GENETIC ENGINEERING IN MAMMALIAN CELLS

REPLICATION AND EXPRESSION OF THYMIDINE KINASE AND HUMAN GLOBIN GENES MICROINJECTED INTO MOUSE FIBROBLASTS. W. French Anderson, Lillian Killos, Linda Saunders-Haigh, Peter J. Kretschmer and Elaine G. Diacumakos in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 77, No. 9, pages 5399–5403; September, 1980.

ALTERING GENOTYPE AND PHENOTYPE BY DNA-MEDIATED GENE TRANSFER. Angel Pellicer, Diane Robins, Barbara Wold, Ray Sweet, James Jackson, Israel Lowy, James Michael Roberts, Gek Kee Sim, Saul Silverstein and Richard Axel in *Science,* Vol. 209, No. 4463, pages 1414–1422; September 19, 1980.

GENE THERAPY IN HUMAN BEINGS: WHEN IS IT ETHICAL TO BEGIN? W. French Anderson and John C. Fletcher in *New England Journal of Medicine,* Vol. 303, No. 22, pages 1293–1297; November 27, 1980.

FIBER BUNDLES AND QUANTUM THEORY

A COMPREHENSIVE INTRODUCTION TO DIFFERENTIAL GEOMETRY. Michael Spivak. Publish or Perish, 1970.

MAGNETIC MONOPOLES, FIBER BUNDLES, AND GAUGE FIELDS. Chen-Ning Yang in *Annals of the New York Academy of Sciences,* Vol. 294, pages 86–97; November 8, 1977.

EINSTEIN'S IMPACT ON THEORETICAL PHYSICS. Chen-Ning Yang in *Physics Today,* Vol. 30, No. 6, pages 42–49; June, 1980.

MIMICRY IN THE SEXUAL SIGNALS OF FIREFLIES

BIOLUMINESCENCE AND COMMUNICATION. James E. Lloyd in *How Animals Communicate,* edited by Thomas A. Sebeok. Indiana University Press, 1977.

INSECT BIOLUMINESCENCE. James E. Lloyd in *Bioluminescence in Action,* edited by Peter J. Herring. Academic Press, 1978.

SEXUAL SELECTION IN LUMINESCENT BEETLES. James E. Lloyd in *Sexual Selection and Reproductive Competition in Insects,* edited by Murray S. Blum and Nancy A. Blum. Academic Press, 1979.

BINOCULAR DEPTH INVERSION

THE EFFECT OF TOUCH ON A VISUALLY AMBIGUOUS THREE-DIMENSIONAL FIGURE. Charmian Shopland and Richard L. Gregory in *The Quarterly Journal of Experimental Psychology,* Vol. 16, Part I, pages 66–70; February, 1964.

ART AND ILLUSION: A STUDY IN THE PSYCHOLOGY OF PICTORIAL REPRESENTATION. E. H. Gombrich. Princeton University Press, 1969.

THE INTELLIGENT EYE. R. L. Gregory. McGraw-Hill Book Company, 1970.

FOUNDATIONS OF CYCLOPEAN PERCEPTION. Bela Julesz. University of Chicago Press, 1971.

THE ARCHITECTURE OF CHRISTOPHER WREN

ST. PAUL'S CATHEDRAL: OBSERVATIONS ON WREN'S SYSTEM OF BUTTRESSES FOR THE DOME, PIERS, AND ON SOME OTHER THINGS. Somers Clark in *Sir Christopher Wren, A.D. 1632–1723.* Royal Institute of British Architects, 1923.

THE ARCHITECTURE OF SIR CHRISTOPHER WREN. Viktor Fürst. Lund Humphries, 1956.

THE STRUCTURAL ANALYSIS OF GOTHIC CATHEDRALS. Robert Mark in *Scientific American,* Vol. 227, No. 5, pages 90–99; November, 1972.

THE AMATEUR SCIENTIST

HIDDEN IMAGES: GAMES OF PERCEPTION, ANAMORPHIC ART ILLUSION. Fred Leemann. Harry N. Abrams, Inc., 1976.

ANAMORPHIC ART. Andy A. Zucker in *Creative Computing,* Vol. 3, No. 4, pages 137–140; July/August, 1977.

ANAMORPHOSCOPES: A VISUAL AID FOR CIRCLE INVERSION. Philip W. Kuchel in *The Mathematical Gazette,* Vol. 63, No. 424, pages 82–89; June, 1979.

Renault Presents The Remarkable 18i Sportswagon.

Renault, maker of Europe's best selling cars, presents a remarkable new wagon for America...the Renault 18i Sportswagon.

Remarkable, because it is, at once, a wagon that provides comfortable seating for 5, *and* muscle to move you crisply from 0 to 50...

...a wagon with up to 65.5 cubic feet of load space, *and* the good manners to go where you aim it through precise, quick rack and pinion steering (lock to lock in just 3 turns)...

...a wagon that takes care of the whole family, *and* carefully meters out fuel for outstanding efficiency*...

...a wagon that bristles with innovative technology of the 80's.

38 HWY EST* (25) MPG EST

Bosch L-Jetronic Fuel Injection
This is the Electric Multi-Point Bosch system also used by Porsche 928 and Jaguar XJ-S. It precisely measures out the fuel required for optimum performance and efficiency from the 18i's enthusiastic 1.6 liter (1647 cc, aluminum block and head) 4-cylinder engine.

20 Years Experience in Front-Wheel Drive Design.
The 18i front-wheel drive system is a study in compact, lightweight design. Front drive transaxles, engine and 4-speed manual transmission (5-speed, avail.) are deftly mounted to provide a 60/40 weight ratio, front to rear. It also allows quick access for service at any of the more than 1300 Renault and American Motors Dealers.

Slip Stream Aerodynamics
The 18i Sportswagon silhouette is notably slippery. A wide front modesty panel—more discreet than add-on air dams—encourages turbulence to slide beneath the 18i. Slip stream styling effects promote excellent fuel efficiency* as the wagon carves the air.

Road Adhesion
The 18i Sportswagon is masterfully tuned for the road. With 155SR x 13 Michelin steel belted radials, box section "live" rear axle (it flexes in the turns), front and rear sway bars, and beefy helical coil springs (variable-flex in the rear) surrounding long-travel shock struts.

Inner Space Geometry
The 18i's elegantly tailored seats are bio-formed with special support for the lumbar region and upper thighs.

Controls and gauges are strategically angled and positioned for driver access.

A wide rear bench seat folds forward, providing a lushly carpeted flat load bed more than 5½ feet long.

The Renault 18i Sportswagon. Remarkable? We think so. It provides what wagon users ask for...generous space and great mileage. And something many have been missing...the sheer joy of commanding a responsive, nimble, sensitive, exciting road machine.

*EPA estimated at (25) mpg. 38 mpg highway est. Remember: Compare this estimate with estimated mpg for other cars. Your mileage may differ depending on speed, trip length and weather. Your highway mileage will probably be lower.

Renault 18i

More than just economy

At Renault and American Motors dealers.

THE TASTE BEYOND 12-YEAR-OLD SCOTCH

Discover more.

Let your mature taste lead you from the finest premium Scotch on up to the most expensive 12-year-old Scotch in the world. The Glenlivet. The ultimate in Scotch.

Most premium Scotch is blended and depends on several whiskies for taste and smoothness.

Yet the unblended character of The Glenlivet is noble enough to stand alone. In this distinctive Scotch whisky you'll experience an exceptional smoothness and full-bodied richness, unmatched in all other premium Scotch.

Try the taste beyond premium Scotch.

The Glenlivet
Unchanged since 1824.