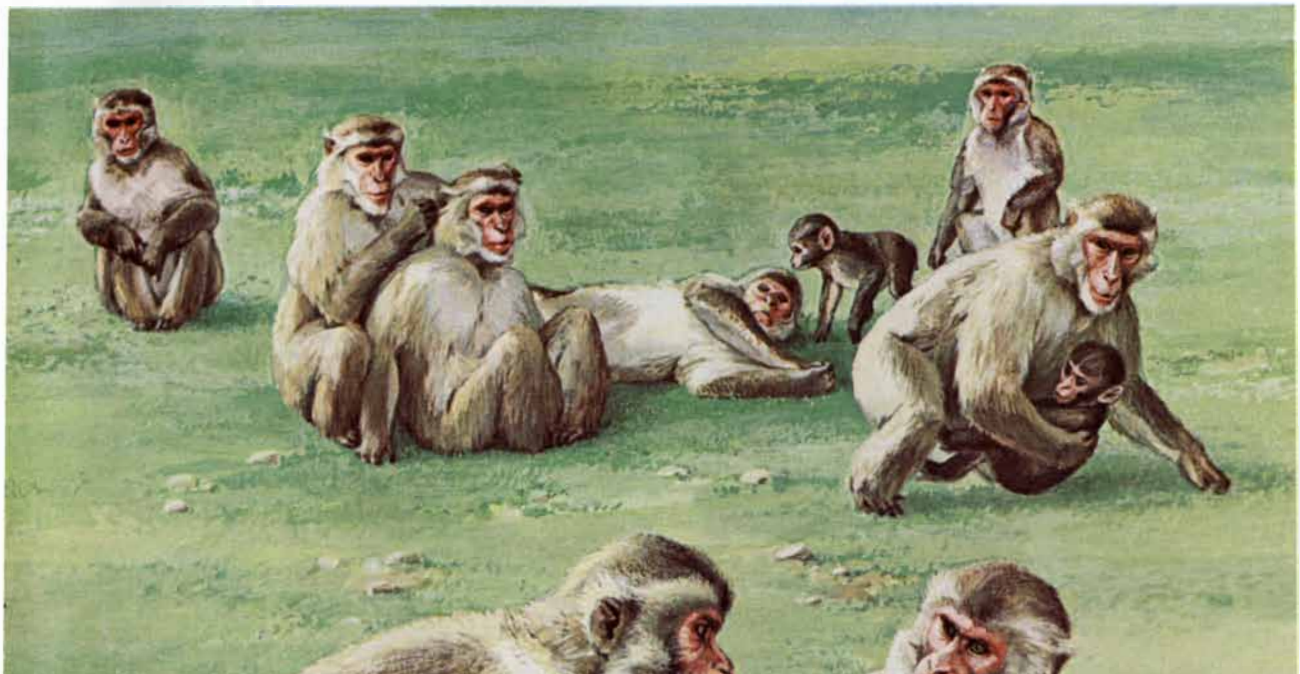
SCIENTIFIC AMERICAN



SOCIAL ORDER OF MACAQUES

$1.50

October 1976

WE'RE LOOKING FOR PEOPLE WHO LOVE TO DRIVE.

We realize that, for some of you, driving an automobile is about as exhilarating as riding an escalator. That's sad.

Because with the right kind of car in your hands, the act of driving can be one of the truly pleasant things you do each day.

Which brings us to Camaro. In fact it brings lots of us to Camaro.

People who love to drive love Camaro because it's definitely a driver's car. It sits low and stands wide and moves like it really means it. Camaro is quick, quiet, tight and tough. All of which translates to a very special "feel". The spirit of Camaro. The lift the car can give you, even just driving to work.

If you love to drive, or would like to, take a turn in a '77 Camaro one day real soon.

Your Chevy dealer has one all gassed up and ready to go.

Driving gloves are optional.

Chevrolet

CAMARO

GM

MARK OF EXCELLENCE

The Most Astonishing Scientific Of This Decade!

106 Complete Functions • Features Unequalled by Any Other Make Produced in USA by COMMODORE

A Limited Offering @

$59.88

The Commodore rechargeable SR5190R

A powerful preprogrammed multifunction calculator with 106 directly accessible keyboard functions.

The complete mini-computer for
Every Engineer
Physicist
Chemist
Geologist
Demographer
Ecologist
Mathematician
Statistician
Quality Control Analyst
Time and Motion Expert
Advanced Student
. . . and so many others

There has never been an electronic calculator with so extensive a performance range as the brand new Commodore SR5190R. Its operating capacity, combined with the speed and accuracy with which computations are performed is unequalled by any other calculator.

We are proud to offer it to you on a, "no-questions asked", full refund guarantee. More than its diversified applications, we want you to appreciate its many exclusive advantages for yourself.

Discover its unique **Hours-Minutes-Seconds** in digital clock format. Ambiguity is a thing of the past. And you can perform arithmetic operations, such as time study and motion analysis in this mode as easily as in the decimal mode.

Complex numbers, difficult on some calculators, impossible on others, are directly accessible from the SR5190R keyboard. Thanks to its extraordinary accuracy limit, **Combination and Permutation** operations are not hindered by the overflow which occurs when the factorial is larger than 10^{100}.

Among the 106 direct entry functions are **Poisson** and **Binomial Probability Densities; Gaussian Distribution; Linear Regression Analysis; Mean and Standard Deviation;** and so many more. Most important, all are on the keyboard. Several thousand internal preprogrammed steps put all of these functions at your fingertips . . . instantly. You needn't bother with preparing formulas, creating programs or maintaining an elaborate library. It's all right there. The broadest range of diversified applications is at your command.

commodore
A Name You Can Trust

Commodore is perhaps the oldest brand in consumer electronic calculators. The SR5190R was designed, developed and produced in Palo Alto, California. Other Commodore plants are located in Osaka, Japan, Toronto, Canada and Eaglescliff, England.

SATISFACTION OR YOUR MONEY BACK IN FULL

Order your SR5190R Rechargeable Scientific on a "No-Questions-Asked" trial basis. If not satisfied with the performance of your SR5190R, return it for a full refund.

A Brief Summary of the SR5190R

10-digit Mantissa, 2-digit exponent with Variable Exponent Integer Increase and Decrease: MANT, EE, EE↑, EE↓

Register Keys: STO 1, RCL 1, Σ1, STO 2, RCL 2, Σ2, x↔y, Xn, Xi, Yi, α, β, γ, ( ).

One Real Variable Function Keys: ln, log, e^x, 10^x, 1/x, $\sqrt{x}$, x^2, SIN, COS, TAN, INV SIN, INV COS, INV TAN sinh, cosh, tanh, INV sinh, INV cosh, INV tanh

Two Real Variable Arithmetic Function Keys: + − × ÷

Two Complex Variable Arithmetic Function Keys: j +, j −, j ×, j ÷

Two Real Variable Analytical Function Keys: →P, →R, P^n_m, y^x, $\sqrt[x]{y}$, %, Δ%, C^n_m

Statistical Function Keys: x↔y, SLOPE, INTCP, GAUSS, BINOM, POISS, x_s, y_s

Hours-Minutes-Seconds Mode: HMS

Unit Conversions: (°F) C, (d) dms, (d) gra, (gal) l, (oz) g, (lb) kg, (ft) m, (mi) km, (f oz) l, (in) cm, (BTU) J

Degree/Radian Conversion and Mode Keys: d/r, d↔r

Numerical Entry Keys: 0-9, π

3 Angular Units: radians, degrees, grads

Poisson and Binomial probability densities. Gaussian distribution.

Mean and Standard Deviation PLUS Linear Regression Analysis.

Power supply: Built-in rechargeable Ni-Cad batteries. AC adapter/recharger included.

One Year Warranty

Call Toll Free:
800-621-8318
(Illinois call: 312-595-0461)

Work With It For Two Weeks
No Obligation

Please send ________ Commodore SR5190R 106 Function Scientific Calculator(s) at the limited price offering of only $59.88 (plus $3.95 shipping and insurance) each. If not completely satisfied I can return it within two weeks for an immediate refund.

☐ Check or Money Order Enclosed
(Illinois residents add 5% sales tax)
☐ Charge My Credit Card Checked Below:
☐ American Express ☐ BankAmericard
☐ Diners Club ☐ Master Charge
☐ Carte Blanche

Credit Card #________

Master Charge Bank #____

Expiration Date________

Name________

Address________

City________ State________ Zip____

SCA-10/76

Signature________

Send Coupon To
Contemporary Marketing, Inc.
790 Maple Lane, Bensenville, Ill. 60106

Established 1845 SCIENTIFIC AMERICAN October 1976 Volume 235 Number 4

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $15 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $18 PER YEAR, ALL OTHER COUNTRIES

THE COVER

The painting on the cover shows nine members of a troop of 230 Japanese macaques living under seminatural conditions at the Oregon Regional Primate Research Center. In the foreground a fight is beginning between two adult male macaques, a common occurrence during the fall and winter breeding season. At right an adult female is about to carry her infant away from the area. Behind her and in the far left corner sit two low-ranking males. In the middle of the group an infant wanders past a dozing juvenile and a consort pair; the male member of the pair, startled by the screams of the fighting monkeys, has momentarily interrupted grooming the female. Macaque behavior is of special interest since it is not precisely defined by biological factors; instead social and developmental factors play a major role, giving rise to a flexible social structure that can adapt to changing situations and varied environments. G. Gray Eaton, who has been studying these monkeys for the past six years, describes his observations in his article "The Social Order of Japanese Macaques" (*page 96*).

THE ILLUSTRATIONS

Cover painting by Peter McGinn

Page	Source
25	World Health Organization
26–31	Lorelle M. Raboni
32	Fritz Goro (*top*), World Health Organization (*bottom*)
33	World Health Organization
34–35	Ralph Morse
36–43	Dan Todd
45	James A. Lake
46–54	Bunji Tagawa
66–78	Allen Beechel
81–84	Fritz Goro
85–94	George V. Kelvin
96	Kurt B. Modahl (*top*), Jon Brenneis (*bottom*)
97	Andrew Christie
98	Jon Brenneis
99	Andrew Christie
100–101	Sarah Landry
102–103	Andrew Christie
104–105	Sarah Landry
106	Andrew Christie
108–109	Bradshaw Photo, Big Spring, Tex.
110	Alan D. Iselin
111	Hamilton E. McRae III
113	Sherwood B. Idso (*top*); Con Keyes, EPA-Documerica (*bottom*)
114	National Oceanic and Atmospheric Administration (*top*); Theodore F. Tyler, U.S. Geological Survey (*bottom*)
116–118	Tom Prentiss
119	Tom Prentiss (*top*), Maurice Beresford (*bottom*)
120–125	Richard Daggett
126–127	Tom Prentiss
128	Maurice Beresford
132–136	Andrew Christie
139	Ramsay & Muspratt, Ltd.
140	*Folktales Told around the World.* © 1975, University of Chicago Press
141	*The Hohokam.* © 1976, University of Arizona Press
142	*Abyssal Environment and Ecology of the World Oceans.* © 1973, John Wiley & Sons
143	*Memory and Attention.* © 1976, John Wiley & Sons

WANTED

Five enterprising people to take on challenging project.

50 years of the Rolex Oyster

1976 is the 50th anniversary of the introduction of the Rolex Oyster case, which provides precise watch mechanisms with extraordinary protection against water, dust, dirt and shock.

Since 1926, Rolex has pioneered many other major developments in the wristwatch.

And Rolex owners have been pioneering, too. Lord Hunt climbing Mt. Everest, Sir Francis Chichester sailing alone around the world . . . Thor Heyerdahl, Jackie Stewart, Arnold Palmer . . . those are just a few.

The Rolex Awards for Enterprise

To provide financial help for projects that break new ground and capture the spirit of enterprise shown by Rolex and Rolex owners, we are offering five Awards for Excellence, totaling 250,000 Swiss Francs. Each winner will receive 50,000 Swiss Francs, plus a specially inscribed Rolex Chronometer.

Projects eligible

Applicants are invited to submit proposals for projects in any of these three general categories:

Science and invention —Projects involving a step forward in research, experiment or innovation in the general field of science and technology.

Exploration and discovery—Projects submitted need not match the scale of the examples mentioned earlier. This category covers a broad spectrum of activities—from original map research to an expedition.

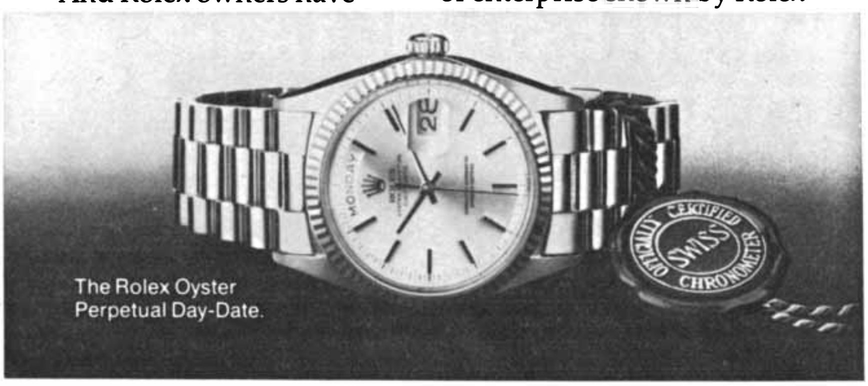
The environment—Projects that seek to protect and preserve the world around us—by fighting pollution, for example, or helping to save some of our threatened species of wildlife.

How to apply

Official entry forms, detailed rules and conditions are available from: The Secretariat, The Rolex Awards for Enterprise, P.O. Box 695, 1211 Geneva 1, Switzerland.

Include your name and address and state the general nature of the project you are considering for entry.

Entries should be submitted as soon as possible and must arrive no later than March 31st, 1977.



The Rolex Oyster Perpetual Day-Date.

ROLEX

LETTERS

Sirs:

I was amazed, as I am sure others were, by the responses of Messrs. Koenig and Kofahl to Dorothy Nelkin's article "The Science-Textbook Controversies" ["Letters," SCIENTIFIC AMERICAN, July]. In her reply Professor Nelkin refrains from counterargument, thereby underlining the fact that her discussion concerns the sociology of the issue, not who is right or wrong. Therefore allow me to come not to her defense (she needs none) but to science's.

For at least the past decade I think it has generally been agreed among philosophers of science that Karl Popper's criterion of falsifiability is as inadequate as the old positivist criterion of verifiability, and for quite symmetrical reasons. Theories do not by their nature admit of strict falsifiability (or verifiability), but that there can be confirming or disconfirming evidence sufficient unto rational belief is another matter, and the theory of evolution has of course plenty of confirming evidence. (If what the antievolutionists demand is logical certainty, I suggest they set evolution aside and first cut their teeth on the problem of proving the existence of the material world.) Dr. Kofahl never tells us why he takes the theory to be unfalsifiable. Whether it is or not, if a philosophical analysis does lead to the bizarre conclusion that the theory of evolution is not a scientific one, that is an eminently safe ground for dismissing the philosophical analysis.

A number of comments are in order regarding Reverend Koenig's letter. Whether or not a theory is true, and whether or not a given practical application of it is wise, are distinct questions, and the difference seems not to be fully appreciated by some parties to the dispute. As regards whether or not a theory is true, the only relevant consideration is the evidence pro or con: one can quite consistently hold that a theory is true, and that it would be better if people did not believe it to be so. However, if I may indulge in an *ad hominem* comment, the Bible itself is not neutral on this issue: we are enjoined to know the truth, since the truth shall make us free.

Reverend Koenig asserts that the "bases" of science are nonrational, and he cites a mixed bag of alleged examples. I fail to fathom the relevance of some; most do not have the required logical status of being presuppositions necessary to the scientific enterprise. I for one do not know why there is something rather than nothing. I also do not know whether the question has an answer, let alone a scientific one. The standard religious answer seems only to beg the question (Why then is there a God rather than no God?); attempts to meet this deficiency by arguments for God's self-sufficiency or necessary being seem to have fallen short of conviction. More important, we are not told in what sense this question, or its answer, is a "basis" of science.

Koenig goes on to list honesty, logic, memory and generalization as further "bases"; allegedly these are both assumed on faith and "proved to be valid by testing." Which is it? If it is the latter, and the testing is scientific, then we are in a vicious circle, relying on science to validate the presuppositions of science. But there is no such vicious circle. I confine myself to two remarks. First, to argue that "faith" in logic is nonrational sounds extremely puzzling. Second, as regards honesty, we clearly do have empirical means for checking into it, and science has armed itself with them. Specifically, one wonders whether Koenig wishes to imply that Darwin and his successors were attempting to perpetrate a fraud.

In conclusion, let me comment that I would probably take umbrage at Koenig's suggestion that philosophical thinking is "nonrational," if I could figure out what he meant by it. I am sure I speak for most philosophers. The remaining question is whether theology is nonrational. That one I leave to the theologians to answer.

EVAN FALES

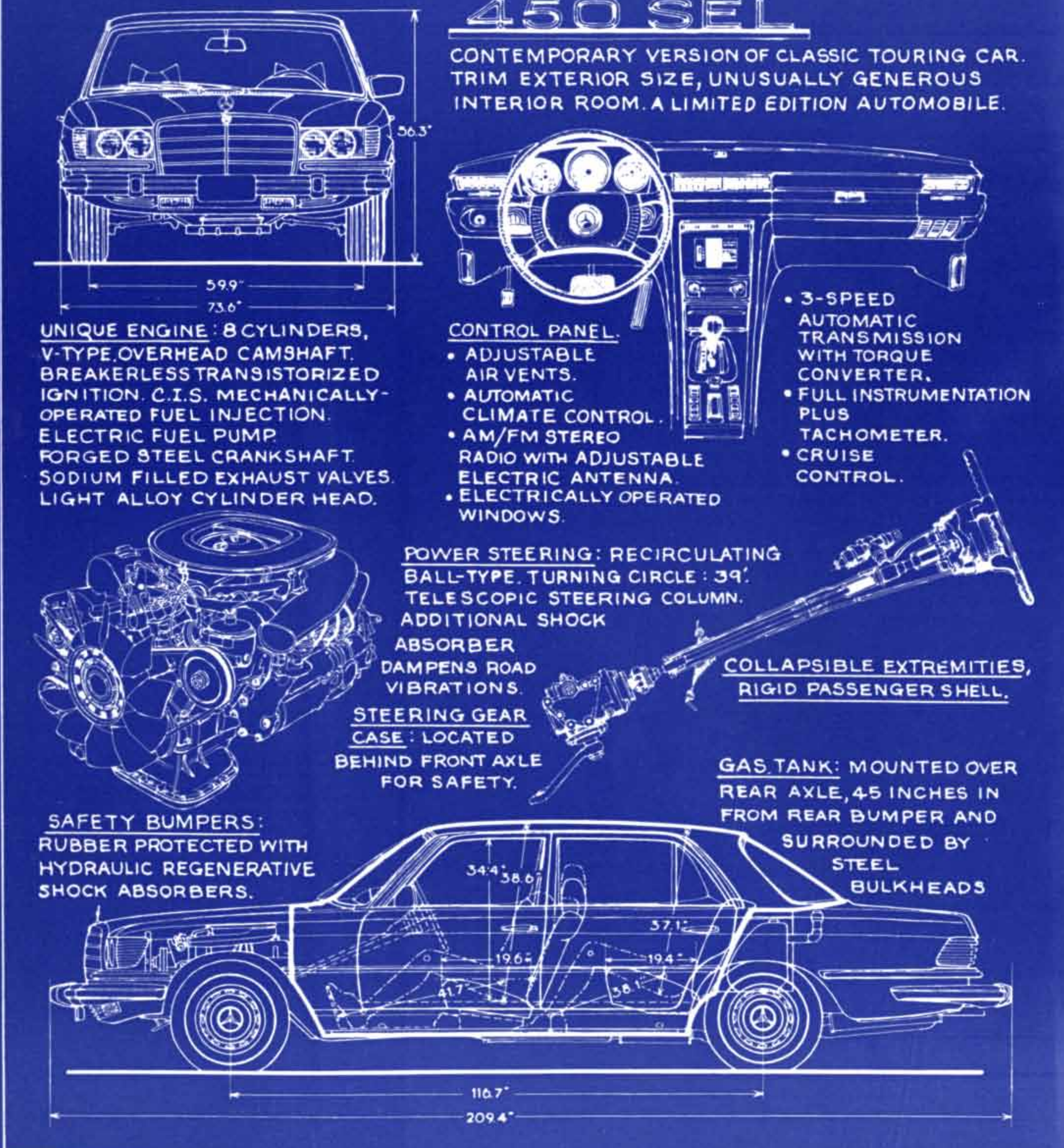
Department of Philosophy
University of Iowa
Iowa City

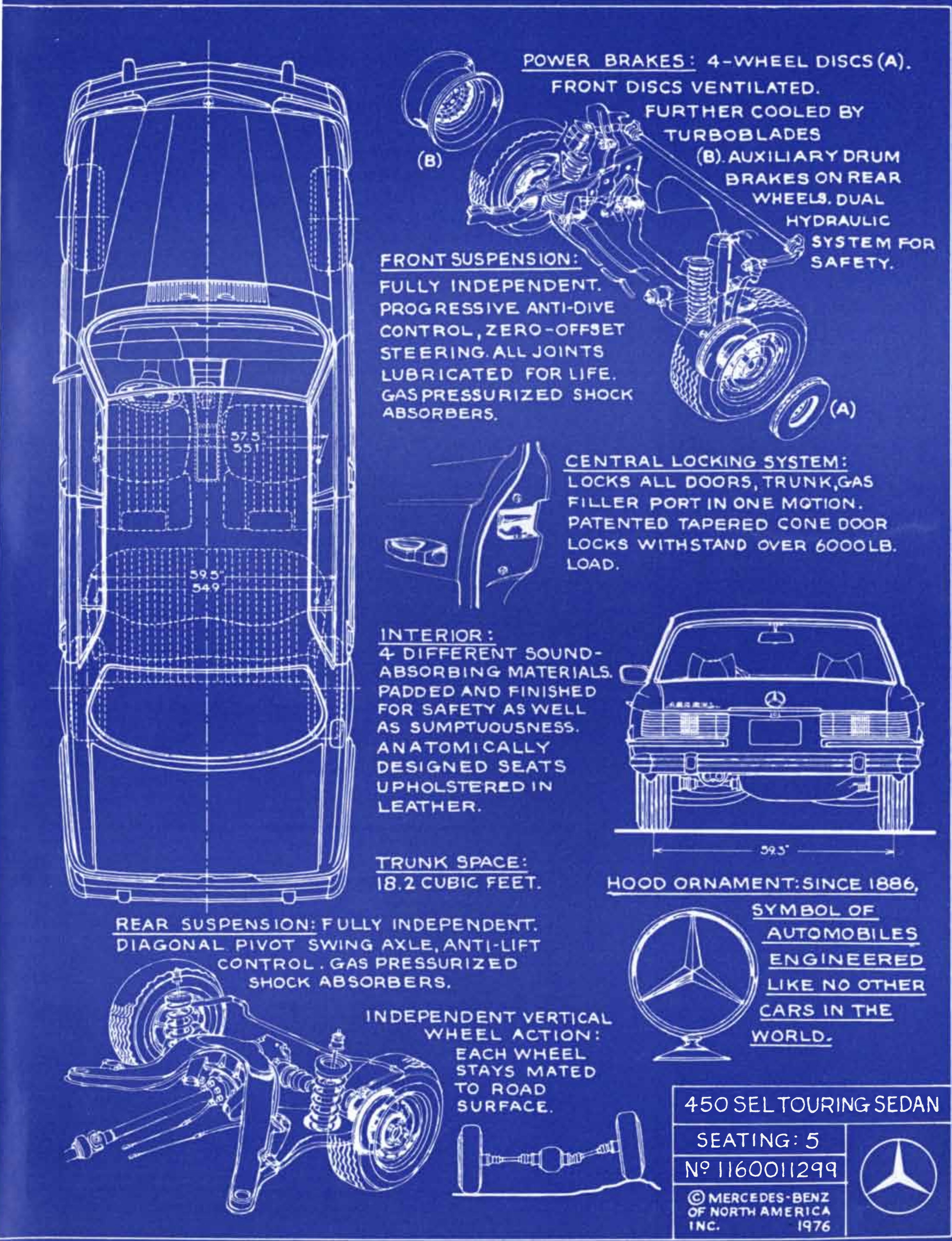
Sirs:

In his general conclusion that railways caused the end of the canal era, John S. McNown fails to mention in his article "Canals in America" [SCIENTIFIC AMERICAN, July] an interesting anomaly of the Erie Canal.

The Erie Canal reached its tonnage peak of traffic only in 1880, some 30 years after the parallel New York Central route had been opened. Considering that the freight rates were lower by about a factor of three for the canal, this is not surprising. When the canal tolls were abolished completely in 1882, the canal carriers were relieved of any right-of-way costs, courtesy of the state's taxpayers. This toll-free policy has been extended to our present extensive system of Federal waterways, bedeviling the railroads, which must support their own right-of-way costs, right up to the present. Indeed, William S. Vickrey of Columbia University has pointed out how the setting of rates by the Interstate Commerce Commission based on this cost inequity has led to an economic misallocation of haulage to the detriment of the railways.

How to tell a Mercedes-Benz 450SEL from all the rest.




POWER BRAKES: 4-WHEEL DISCS (A). FRONT DISCS VENTILATED. FURTHER COOLED BY TURBOBLADES (B). AUXILIARY DRUM BRAKES ON REAR WHEELS. DUAL HYDRAULIC SYSTEM FOR SAFETY.
(B)
(A)
FRONT SUSPENSION: FULLY INDEPENDENT. PROGRESSIVE ANTI-DIVE CONTROL, ZERO-OFFSET STEERING. ALL JOINTS LUBRICATED FOR LIFE. GAS PRESSURIZED SHOCK ABSORBERS.
CENTRAL LOCKING SYSTEM: LOCKS ALL DOORS, TRUNK, GAS FILLER PORT IN ONE MOTION. PATENTED TAPERED CONE DOOR LOCKS WITHSTAND OVER 6000 LB. LOAD.
INTERIOR: 4 DIFFERENT SOUND-ABSORBING MATERIALS. PADDED AND FINISHED FOR SAFETY AS WELL AS SUMPTUOUSNESS. ANATOMICALLY DESIGNED SEATS UPHOLSTERED IN LEATHER.
TRUNK SPACE: 18.2 CUBIC FEET.
HOOD ORNAMENT: SINCE 1886, SYMBOL OF AUTOMOBILES ENGINEERED LIKE NO OTHER CARS IN THE WORLD.
REAR SUSPENSION: FULLY INDEPENDENT. DIAGONAL PIVOT SWING AXLE, ANTI-LIFT CONTROL. GAS PRESSURIZED SHOCK ABSORBERS.
INDEPENDENT VERTICAL WHEEL ACTION: EACH WHEEL STAYS MATED TO ROAD SURFACE.
57.5"
59.5"
59.5"
450 SEL TOURING SEDAN
SEATING: 5
Nº 1160011299
© MERCEDES-BENZ OF NORTH AMERICA INC. 1976

If indeed the railways sounded the death knell of canals, it is surprising that some of the greater canal engineering feats were undertaken in this century, such as the canalization of the Mohawk River to replace the Erie Canal, begun in 1904. It seems to me that a generalization emerges from the competition of the railroads with the canals and then the competition of the automobile and the airplane with the railroads: The older mode has its useful place in the scheme of transport, but because the attractions of greater speed and flexibility cause us to carelessly proliferate new modes, we end up with a situation where public financial intervention is necessary to keep all modes operative. It is not clear that we have thus achieved the most efficient or economical transport system for ourselves.

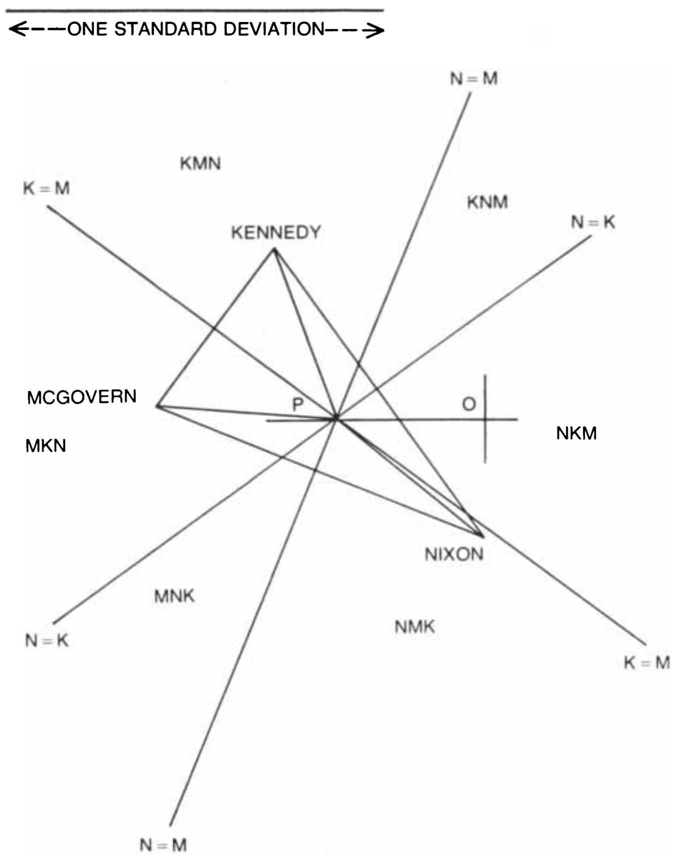
GARY NELSON

Transportation Planner
Troy, N.Y.

Sirs:

One of the worst features of ordinary plurality voting and of all the other voting systems described by Richard G. Niemi and William H. Riker ["The Choice of Voting Systems," SCIENTIFIC AMERICAN, June] was mentioned on the first page of their article. This feature is that votes may be split between two or more candidates whose platforms are similar, with the result that some other candidate wins even though his platform is less popular.

Earlier this year we published a voting system that is free from this feature ("From Individual to Collective Ordering through Multidimensional Attribute Space," *Proceedings of the Royal Society of London, Series A,* Vol. 347, pages 371–385). It uses the distribution of rank orders of preferences to estimate the positions of the candidates in "attribute space." The calculations, for three or more candidates, require the use of a computer, and we have written a program that deals with the case of three candidates. The accompanying illustration [*below*] shows an application of the program. It displays the relative locations of Nixon, McGovern and Edward Kennedy, based on the rank-order preferences of 2,521 voters. Zero is the mode of (normal) distribution of locations that most voters would prefer a candidate to occupy. The "winner" is the candidate whose representative point is closest to zero. Intransitivities can arise in this system, but they have an



A voting system in multidimensional-attribute space

interpretation as statistical error and are not at all paradoxical.

I. J. GOOD

T. N. TIDEMAN

Virginia Polytechnic Institute
and State University
Blacksburg, Va.

Sirs:

The system described by Professors Good and Tideman provides an excellent example of the difficulties involved in solving problems of voting. It solves some problems but creates others.

The Good-Tideman system is a device for generating a cardinal ordering from originally ordinal ballots. Although their cardinality minimizes intransitivities, their cardinalization (like others heretofore proposed) suffers from the defect that the winner under their calculations may not be the same as would have been the case if the voting data could have been collected originally in cardinal form. That is, there is no reason to suppose their cardinalization is the "true" one.

RICHARD G. NIEMI

WILLIAM H. RIKER

University of Rochester
Rochester, N.Y.

Scientific American, October, 1976; Vol. 235, No. 4. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

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

Offprint correspondence and orders should be addressed to W. H. Freeman and Company, 660 Market Street, San Francisco, Calif. 94104. For each offprint ordered please enclose 30 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 ____________________

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

OCTOBER, 1926: "The story of agriculture is the story of coal. The manufacture of modern farm machinery involves the use of coal in melting the iron out of the ore, in reducing the iron to steel and in fashioning the steel into the harrow and the plough. We hear much these days about nitrogen fertilizers, of which hundreds of thousands of tons are used every year on our farms. The nitrogen is a by-product of the coke oven, and 130 tons of coal have to pass through the ovens to produce 100 tons of coke. The movement of the country's enormous crops to market could be accomplished in no other way than by our railroads, which are run by coal. Coal drives the flour mill, and for every pound of bread it is necessary for the baker to burn up a third of a pound of coal. The United States Coal Commission says that between 40,000 and 50,000 tons of coal invisibly reach our table every day."

"When Guglielmo Marconi visited the United States in June, 1922, he predicted that the powerful trans-Atlantic stations then in existence would become obsolete. He emphasized the fact that he believed three-element tubes for transmitting purposes would replace the more expensive and cumbersome high-frequency alternators, and that short wavelengths were a field of vast possibilities. Four years have passed since Marconi sailed out of New York Harbor on his yacht *Elettra*. In the time that has elapsed since then radio development has followed the trend he outlined. The British Empire is linked together by Marconi short-wave beam transmitters. Many of the high-power long-wave commercial and naval transmitters are supplemented by comparatively simple short-wave installations, and thousands of American amateurs, many of them boys in high school, operating a home-made short-wave transmitter in the attic of their home, are talking around the world with less power than is required to operate an electric toaster or a flat iron."

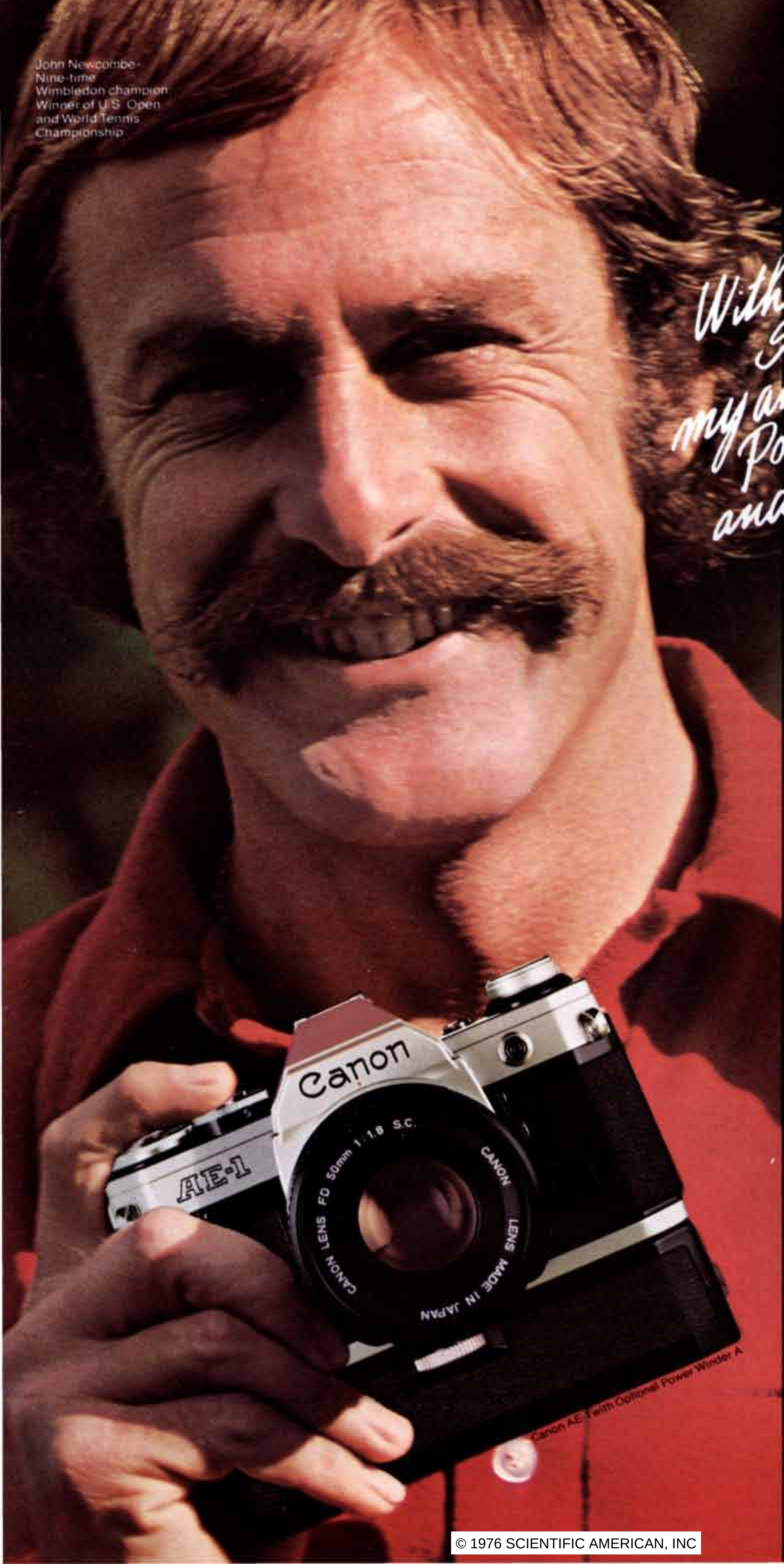
"The measurement of the velocity of light is nothing new. The world has known for generations that it is 186,000 miles per second—to the nearest round thousand. In A. A. Michelson's present work a beam of light, reflected from a rapidly rotating mirror, is sent to another mirror at a distant station and returns after a minute fraction of a second to the spinning mirror again, to find that this mirror has turned in the interval, so that the reflected ray is not sent back to the source as it would be if the mirror were stationary but is sent in a different direction. It is already known that the results of observations on many nights agree so well that we may hope for a final value that will be accurate to within 10 miles per second and perhaps better."

"In the past few years a number of attempts have been made to establish passenger-carrying air lines. Although some were successful from an operating point of view, they were unsuccessful from a financial one. It is therefore good to learn that a regular passenger service is again in being. This is the Philadelphia Rapid Transit Company's air line between Washington and Philadelphia. The flying distance between the two cities is 125 miles, and the flying time is one and a half hours. Two trips each way per day are scheduled, and the fare is $15 one way or $25 for the round trip."

"We have all seen the automatic toasting machine in restaurants and in the sandwich shops that are springing up all over the country. These toasters are too large for home use, but now we have a device that can be placed on the dining-room table. This little brother of the restaurant toaster turns out the same brand of golden-brown toast with no burning. You simply drop the bread into the oven slot and depress the two levers. When the toast is done, up it comes and the current is automatically cut off."

SCIENTIFIC AMERICAN

OCTOBER, 1876: "During Professor Huxley's recent sojourn in America he delivered three lectures (the only scientific discourses given by him in this country) titled 'The Theory of Evolution.' They were given at Chickering Hall in New York on the evenings of September 18, 20 and 22, before large audiences. There was a shade of disappointment visible on the faces of not a few of Professor Huxley's listeners as they left the hall at the close of his first lecture. It had not been at all what they had come to hear. The absurd reports that certain of the daily newspapers had circulated with regard to Professor Huxley's attitude toward religion had led many to anticipate something startling. As the 'arch enemy of Christianity' he could not do otherwise than run amuck with Genesis and say things very offensive to the orthodox. So that, when the speaker finished his calm, straightforward, logical and perfectly reasonable review of the three conflicting theories of the origin of animal forms, a discourse in which there had been no exciting language and no effort to say



With the AE-1, I've lost my amateur standing. Point and shoot, and you're a pro.

Professionalism highlights everything John Newcombe does on the tennis court. But when it comes to taking pictures, he wants a camera that gives great results, but doesn't take a degree in math to operate. That's why he likes the Canon AE-1.

The AE-1 is a fine 35mm camera that has point-and-shoot simplicity, and still has the versatility even a pro can love. Its rapid-fire power winder is great for sequences—you won't miss a shot. And the AE-1's electronic flash is so automatic it's truly foolproof. Best of all, it does all this at a price that's just a little more than what you might spend on a camera that's a lot less. If you want to lose *your* amateur standing in photography, the Canon AE-1 is the way to go.

For a closer look at the Canon AE-1, see your local camera specialty dealer soon.

So advanced, it's simple.

Canon®

AE-1

Canon USA, Inc., 10 Nevada Drive, Lake Success, New York 11040
Canon USA, Inc., 140 Industrial Drive, Elmhurst, Illinois 60126
Canon USA, Inc., 123 Paularino Avenue East, Costa Mesa, California 92626
Canon USA, Inc., Bldg. B-2, 1050 Ala Moana Blvd., Honolulu, Hawaii 96814
Canon Optics & Business Machines Canada, Ltd., Ontario

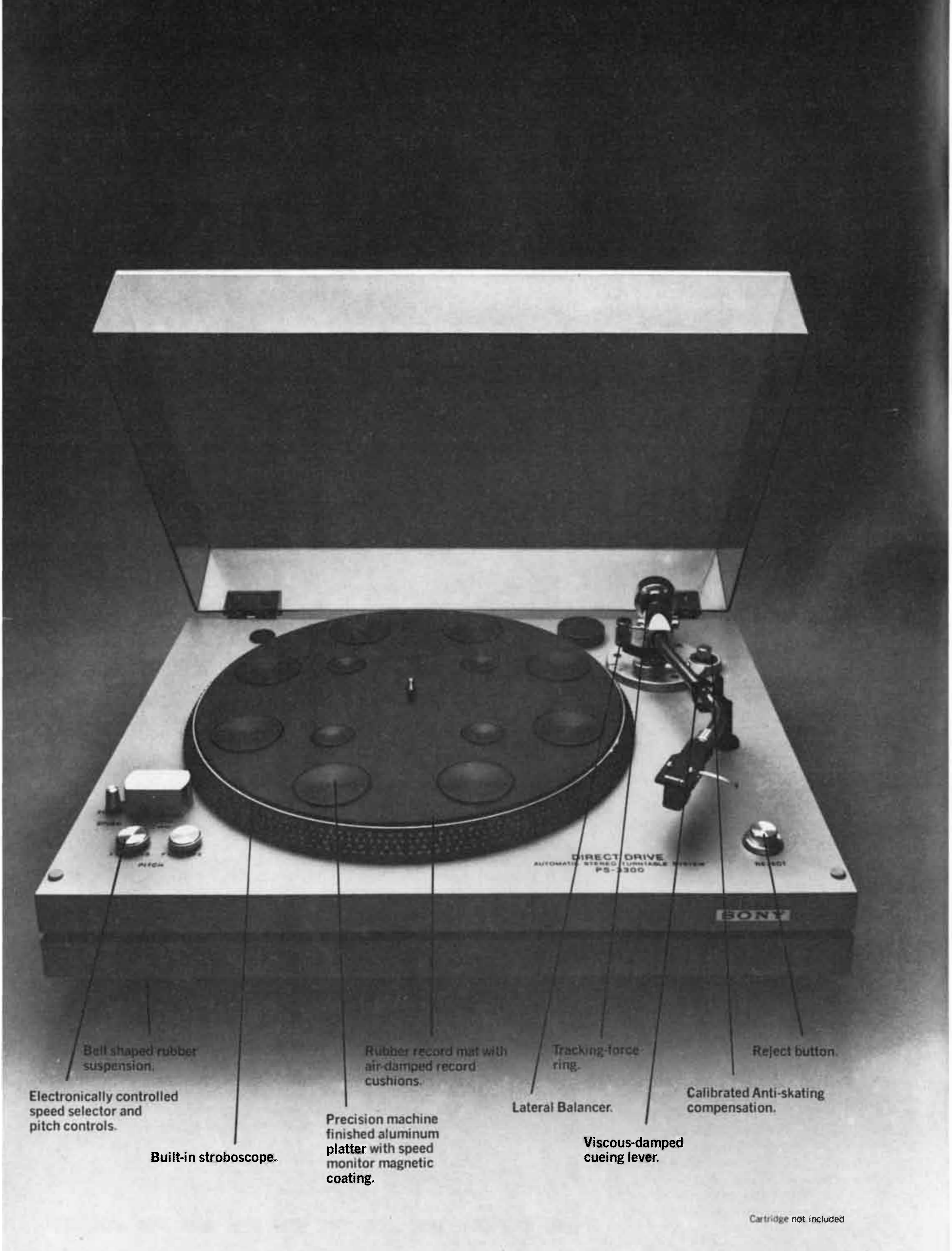
smart things, many in the audience were disappointed and were inclined to blame him for not being the sort of man they had come to see."

"Telegraphic reports from Savannah, on the 23rd September, state that more than 2,000 people are stricken with yellow fever in that city."

"At a recent sitting of the Academie des Sciences, M. Cornu exhibited specimens of photographs of the sun, the moon and the planets taken with a refractor of 15 inches' aperture, which he had specially adapted to photographic work by the device of separating the two lenses of the object glass. By this means the correction for achromatism is altered, so that the greenish-blue and ultra-violet rays, which are the most important for photographic purposes, are united instead of the scarlet and greenish-blue, the combination that gives the best result for optical observations."

"During recent years the British government has invested something like seventy millions of dollars in irrigation works in India, and it is proposed to spend thirty millions more for such purposes during the next five years. In almost every instance the works have proved immediately remunerative, and in some cases the profit has been enormous. The revenue returns from these great undertakings are not the only source of profit. In a country such as India a failure of seasonable rain is apt to be followed by loss of harvests and consequent famine, entailing great loss of life, loss of revenue to the government and sometimes the abandonment of thousands of square miles of fertile soil. All of this is prevented by irrigation. In 1860, when a large part of the North-West Provinces was baked as in an oven, the Ganges Canal preserved grain crops enough to feed a million of people who must otherwise have perished."

"Father Secchi, Director of the observatory of the Roman College, sums up the physical constitution of the sun as follows: The sun is formed of a fluid incandescent mass, enveloped in a highly luminous photosphere, above which there is yet an atmosphere of less density. The photosphere is a fiery mist, probably of gases that have become luminous through the effect of high temperature and high pressure. Immediately above this a very thin envelope of metallic vapors mixed with hydrogen is encountered. This is the chromosphere. Beyond the chromosphere again there is a vast envelope composed of hydrogen and of two unknown substances that produce the yellow spectrum line D_3, and the line 1,474, to one of which the name 'helium' has provisionally been given."



DIRECT DRIVE
SONY
Bell shaped rubber suspension.
Electronically controlled speed selector and pitch controls.
Built-in stroboscope.
Rubber record mat with air-damped record cushions.
Precision machine finished aluminum platter with speed monitor magnetic coating.
Tracking-force ring.
Lateral Balancer.
Reject button.
Calibrated Anti-skating compensation.
Viscous-damped cueing lever.
Cartridge not included

THE AUTHORS

DONALD A. HENDERSON ("The Eradication of Smallpox") is chief medical officer in charge of smallpox eradication at the World Health Organization in Geneva. A graduate of Oberlin College, he received his M.D. from the University of Rochester School of Medicine and Dentistry in 1954 and his M.P.H. from the Johns Hopkins University School of Hygiene and Public Health in 1960. After an internship in medicine he worked for several years in the Epidemic Intelligence Service at the Communicable Disease Center of the Department of Health, Education, and Welfare, serving as chief of the Surveillance Section from 1961 to 1966. In November, 1965, when President Johnson decided to offer assistance in smallpox eradication to 18 countries of western and central Africa, Henderson was designated the program's director. Six months later the World Health Assembly asked him to come to Geneva to organize an intensified global eradication program.

BRUCE CHALMERS ("The Photovoltaic Generation of Electricity") is professor of metallurgy at Harvard University and a consultant to the silicon solar-cell development program at the Mobil Tyco Solar Energy Corporation. He attended the University of London, receiving his B.Sc. in 1929 and his Ph.D. in 1932, both in physics. From an original research interest in the physics of plastic deformation, he undertook to learn how to grow single crystals of metals and why they are almost always imperfect. These studies "led to a broader interest in the behavior of metals and alloys and a metamorphosis from physicist to metallurgist, as teacher, researcher, author, editor and consultant." A lifelong interest in energy sources has culminated in a series of courses on the subject at Harvard and his participation in the development of the silicon solar cell.

DONALD M. ENGELMAN and PETER B. MOORE ("Neutron-scattering Studies of the Ribosome") are both associate professors in the department of molecular biophysics and biochemistry at Yale University. Engelman received his B.A. from Reed College in 1962 and his Ph.D. in biophysics from Yale in 1967. He spent the following year as a postdoctoral fellow in Walther Stoeckenius' laboratory at the University of California Medical Center in San Francisco and then left the country for a two-year fellowship at King's College of the University of London, where he worked with M. H. F. Wilkins and John Randall. He returned to Yale in 1970 as assistant professor. In addition to playing classical and flamenco guitar and enjoying tennis and the outdoors, Engelman is a professional sculptor whose kinetic constructions have been exhibited in Los Angeles, San Francisco, New Haven and London. Moore received his B.S. summa cum laude from Yale in 1961 and his Ph.D. from Harvard University in 1966, both in biophysics, and did his thesis research in James Watson's laboratory. He then spent a postdoctoral year at the Institute of Molecular Biology at the University of Geneva, moving on in 1968 to the Medical Research Council Laboratory of Molecular Biology at Cambridge in England, where he worked with H. E. Huxley. In 1969 he returned to Yale as assistant professor. His recreational activities include soaring and sailing.

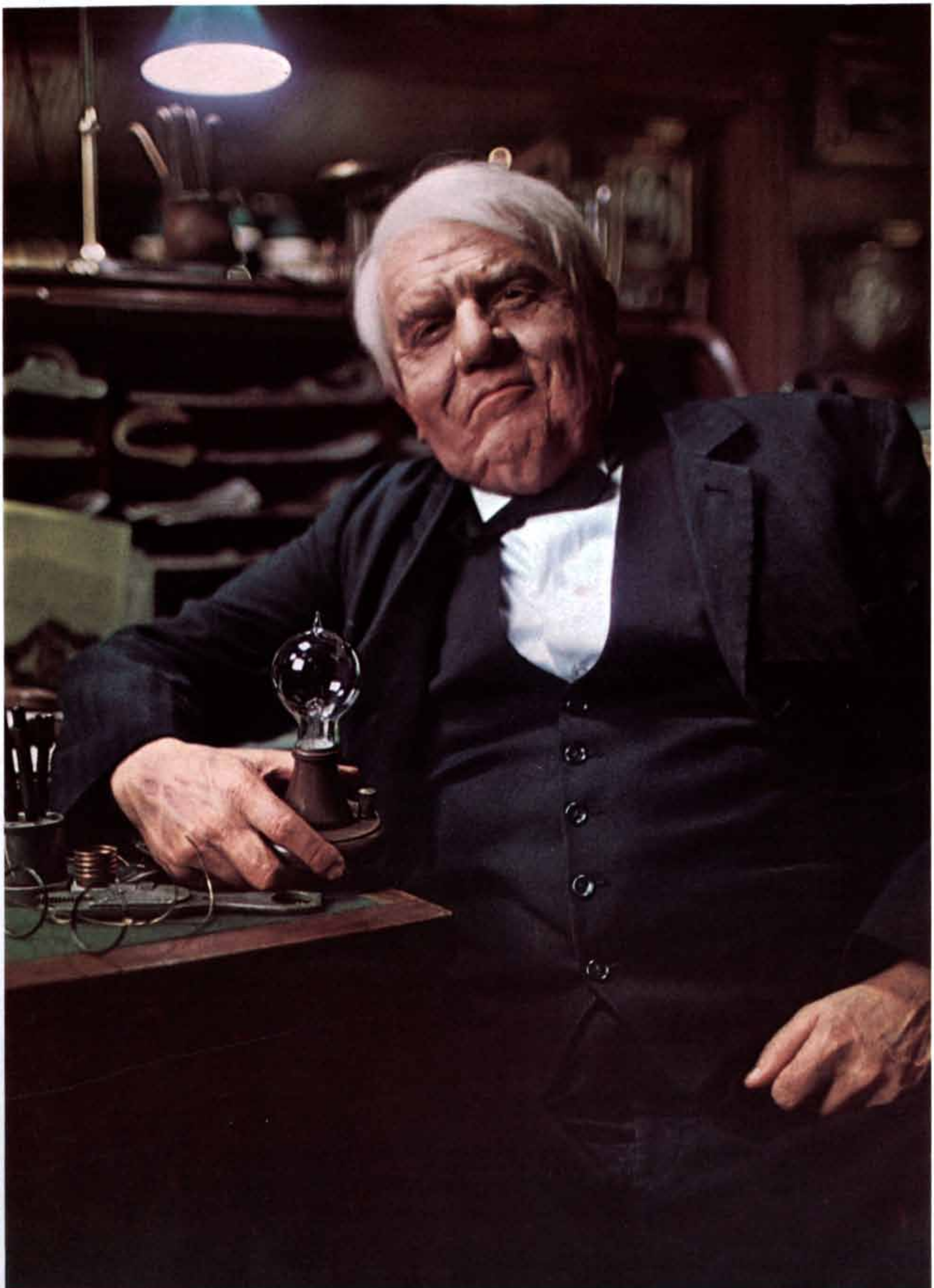
IAN B. STRONG and RAY W. KLEBESADEL ("Cosmic Gamma-Ray Bursts") are on the staff of the Los Alamos Scientific Laboratory. Born in the state of New York but raised in Scotland, Strong attended the University of Glasgow, receiving a B.Sc. in physics in 1953. He then returned to the U.S., and after a few years on the staff of Bell Laboratories he obtained a Ph.D. in nuclear physics from Pennsylvania State University. A member of the high-altitude-physics group at Los Alamos since 1961, Strong lives with his wife and 11-year-old daughter in pine and sagebrush country overlooking the Rio Grande. He and his family grow much of their own food and have a large experimental vineyard. Klebesadel grew up on a small dairy farm in Wisconsin. After serving four years in the Air Force he attended the University of Wisconsin, receiving a B.A. in electrical engineering in 1959. He then helped to oversee the installation of the tandem Van de Graaff accelerator at Wisconsin, meeting the requirements for an M.S. in physics. He has been at Los Alamos since 1960, when he accepted a position to help develop instrumentation for the Vela satellite program.

EMMETT N. LEITH ("White-Light Holograms") is chief scientist at the Willow Run Laboratories of the University of Michigan. He received his B.S. from Wayne State University in 1949, staying on to obtain an M.S. in physics in 1952. That same year he joined the Willow Run Laboratories. In 1955, when he and his colleagues were working on a radar imaging system, they discovered a new way to record microwave images on photographic film, a technique similar in many respects to Dennis Gabor's early work in holography, which was not known to them at the time. Some years later Leith and a colleague, Juris Upatnieks, first demonstrated the capability of holography for forming high-quality three-dimensional images. Leith became professor of electrical engineering at Michigan in 1968. He was cited by the Nobel Prize Commission in 1971 for his contributions to holography.

G. GRAY EATON ("The Social Order of Japanese Macaques") is associate scientist at the Oregon Regional Primate Research Center in Beaverton, Ore. A native of Canada, he received a B.A. in psychology from the University of Victoria in 1964. He then moved on to graduate school at the University of California at Berkeley, initially working on brain-lesion experiments in nonhuman primates. "The following year," he writes, "Professor Frank Beach accepted me as one of his research assistants and my interests changed from neural to ethological (I like to think from reductionist to holist) while I studied social behavior in dogs, birds, rodents and elephant seals. I remained intrigued with primates, however, and accepted a postdoctoral position at the Oregon Primate Center in 1969, where I have been studying Japanese macaques since 1970." Eaton's major recreation is "learning about life" from his infant daughter Jessica. He and his wife Barbara also raise Appaloosa horses.

SHERWOOD B. IDSO ("Dust Storms") is a research physicist with the U.S. Department of Agriculture stationed at the U.S. Water Conservation Laboratory in Phoenix, Ariz. He received a B.S. in physics from the University of Minnesota in 1964, staying on to obtain his Ph.D. in 1967 with "a major in soil science and a split minor in meteorology and mechanical engineering." That broad background has enabled Idso to publish more than 100 scientific papers on topics ranging from basic micrometeorology to the effects (and noneffects) of man on world climate. His article derives from an avocational interest in severe weather phenomena. "I specialize in dust storms and dust devils," he writes. "For both of these activities I live in an ideal location for eyewitness studies."

MAURICE BERESFORD ("A Deserted Medieval Village in England") is professor of economic history at the University of Leeds. After studying history at the University of Cambridge he became a lecturer at Leeds, where a chair was created for him in 1957. He has just completed a year as Harrison Visiting Professor of History at the College of William and Mary. Now back in England, he is engaged in a social history of building in Leeds during the Industrial Revolution. In his spare time Beresford, long concerned with the education of prisoners, is chairman of the parole board of a local prison.



Thomas Edison, reminiscing in 1928 at age 81, is portrayed by actor Pat Hingle in a unique series of television commercials. One is reprinted here. You can see many of the others on the General Electric Theater in late December on CBS.

Established 1845 SCIENTIFIC AMERICAN October 1976 Volume 235 Number 4

The Eradication of Smallpox

A 10-year campaign led by the World Health Organization is right now at the point of success. The world may have seen its last case of the most devastating disease in human history

by Donald A. Henderson

Smallpox, the most devastating and feared pestilence in human history, is making its last stand in two remote areas of Ethiopia, one in the desert and one in the mountains. As of the end of August only five villages had experienced cases in the preceding eight weeks. More important, the onset of the last known case was on August 9. Because man is the only known reservoir for the smallpox virus, the disease should be eliminated forever when the last infected person recovers. Right now more than 1,000 Ethiopian health workers, together with 10 epidemiologists of the World Health Organization, are combing the countryside to make sure no more cases exist. If they discover one, the victim will be isolated under 24-hour guard and everyone who has been in contact with him will be vaccinated. An effort will be made to trace the chain of infection back to a previously known, contained outbreak. For two years after the last case is recorded the search will continue for additional outbreaks. If none is found, and if a WHO international commission can be satisfied that the search has been thorough, smallpox will be declared to have been eradicated from the earth. It will be the first such achievement in medical history.

The interruption of smallpox trans-

SEARCH FOR SMALLPOX CASES is pressed in the remote Simyen Mountains of Ethiopia's Begemdir Province. Two-man teams are assigned to cover an area on foot or on muleback, looking for new cases and vaccinating any possible contacts. Supervisory personnel (in this case a World Health Organization epidemiologist and his Ethiopian counterpart) maintain contact with the teams by helicopter.

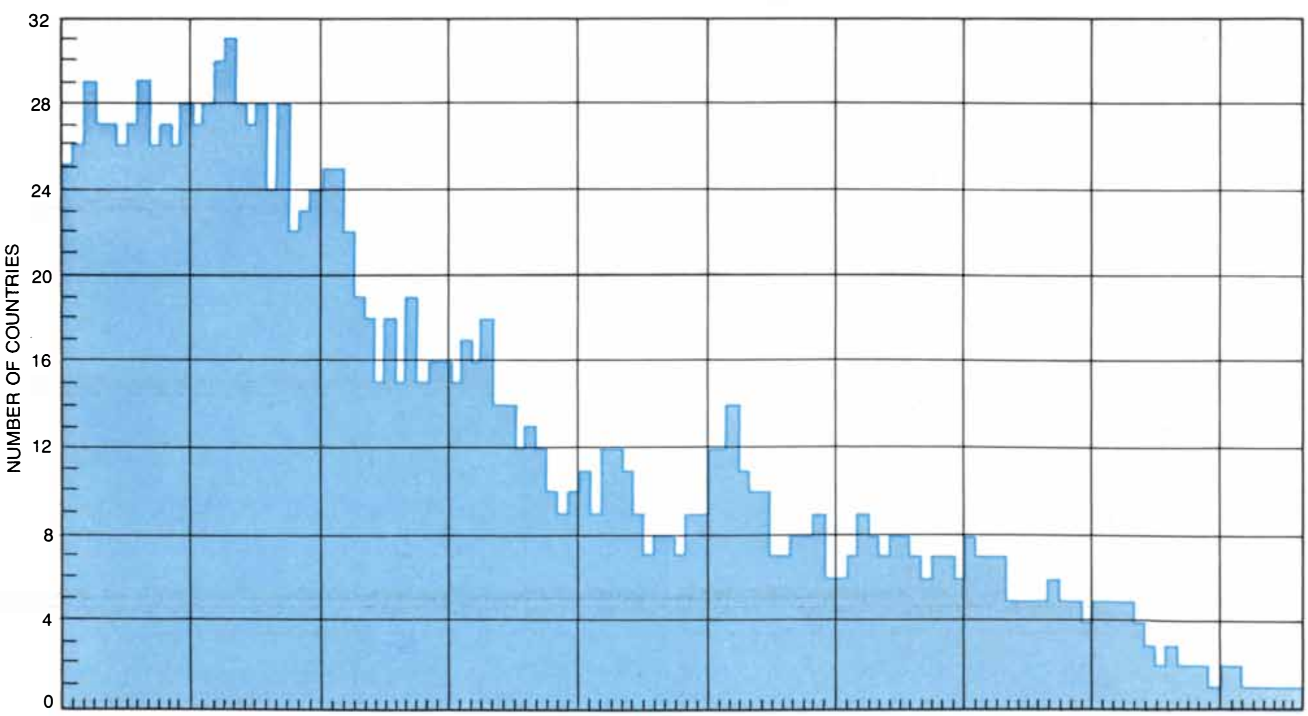
mission in 1976 was the objective of a 10-year global campaign voted by the World Health Assembly (the WHO's controlling body) in 1966 and launched in 1967. When the campaign began, smallpox was considered to be endemic—an indigenous, ever present illness—in more than 30 countries, and "imported" cases were regularly reported every year in perhaps a dozen other countries. The program moved forward steadily, with major campaigns eliminating smallpox successively in western and central Africa, Brazil, Indonesia, southern Africa, Pakistan, India and Bangladesh. In Ethiopia, the last infected country, the campaign has been complicated by some of the most rugged and inaccessible terrain in Africa; it is estimated that more than half of the country's 28 million people live more than a day's walk from any road. Fighting between government forces and various dissident groups has been a recurrent problem. Two Ethiopian health workers have been shot and killed, and the search teams have had to withdraw from some districts for weeks at a time. Yet ever since 1971, when the Ethiopian campaign began, both the extent of the infected areas and the number of cases have been steadily reduced, and now final success seems to be within reach.

Smallpox is caused by a virus that spreads from person to person in minute droplets discharged from the mouth or the nose. (Virus particles in clothing or bedding have sometimes infected people who have not had face-to-face contact with a patient, but such indirect infection has been infrequent.) About 10 or 12 days after inhaling the virus the infected person becomes sick, with a high fever and aching sensations resembling those of acute influenza. After two to four days a rash develops on the face, and within a day or two it spreads over the entire body. Usually it has a "centrifugal" distribution: it is densest on the face, arms and legs and less dense on the trunk. The small, red, pimplelike papules quickly become enlarged vesicles filled first with a clear serum and then, by the fifth day of the rash, with pus. In severe cases the pustules may be so close together, particularly on the face and eyelids, that there is no normal skin; the face is swollen and the patient, now acutely ill, may be unrecognizable. By the 10th day scabs begin to form, and by the third week they fall off, leaving depigmented areas that become pitted, disfiguring scars. Some patients are left blind. Among those afflicted by the virulent Asian form of the virus, variola major, from 20 to 40 percent die. In Ethiopia, where the less virulent variola minor is prevalent, the death rate is 1 percent. Once smallpox has been contracted there is no effective treatment for it.

The origin of smallpox antedates written history. The mummified head of the Egyptian pharaoh Ramses V, who died about 1160 B.C. of an acute infection, shows lesions that appear to be those of smallpox. Chinese and Sanskrit texts indicate that smallpox was also present at least that early in China and India. On the other hand, there is no mention of a disease resembling smallpox in either the Old Testament or the New or in Greek or Roman literature, and it would seem that such a serious disease would almost certainly have been described if it had been prevalent.

A plausible explanation lies in the same epidemiological characteristics of smallpox that lead us to believe it can now be eliminated. Since there is no known animal or insect reservoir of the virus, for infection to persist in a population one afflicted person must transmit the virus to a susceptible contact, and that contact in turn must transmit it to another in an unbroken chain. The smallpox victim can transmit the disease only from the time his rash appears until the scabs drop off, a period of about four weeks. After that he is immune to reinfection. In isolated villages and among scattered populations a point is therefore reached where so few people remain susceptible that the chain of transmission is broken. In such an area smallpox dies out, and it does not recur unless it is reintroduced.

It seems reasonable to speculate that in ancient times only the more densely populated areas of India and China

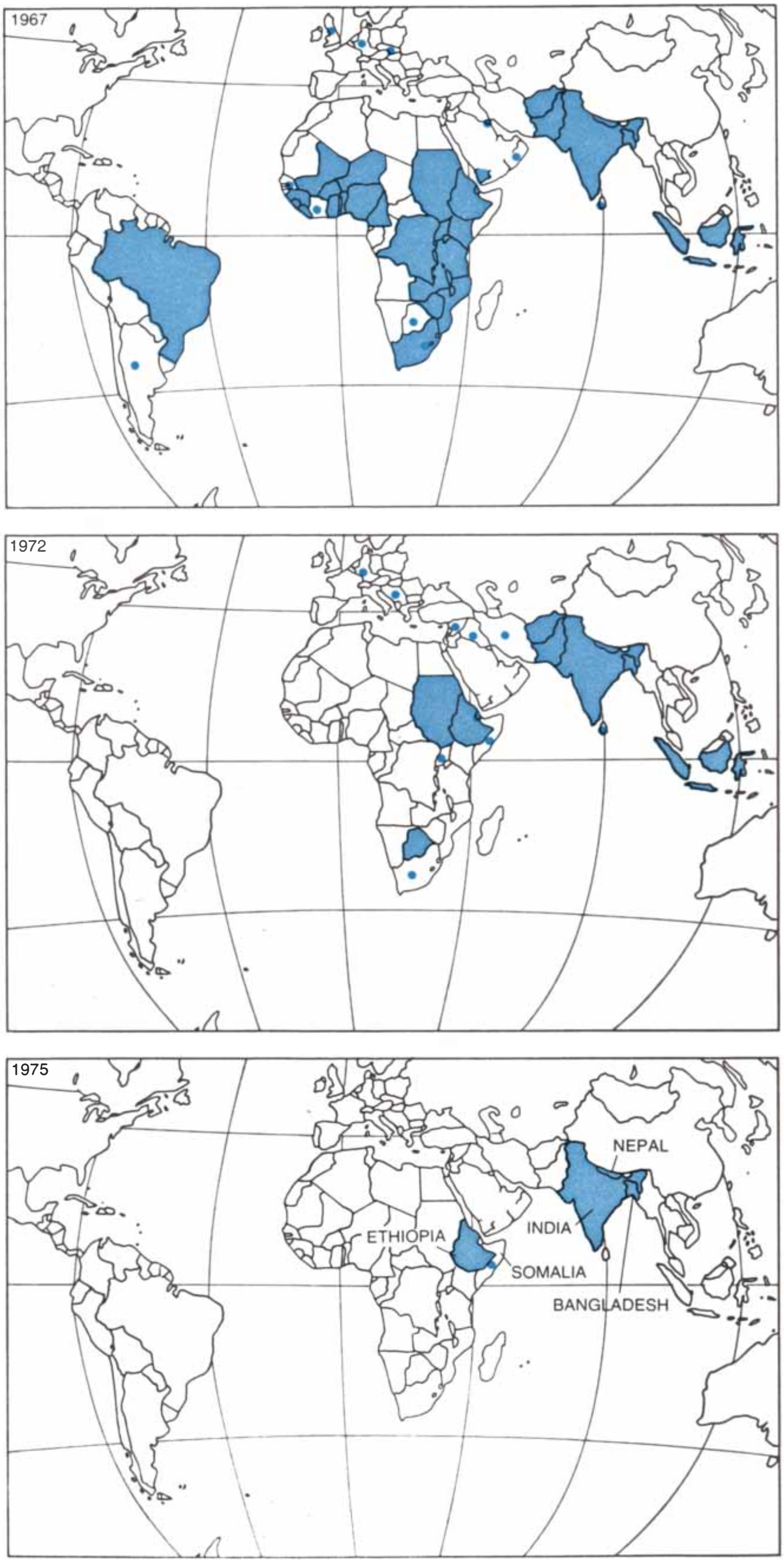


TEN-YEAR TREND in the incidence of smallpox is traced here by a histogram showing the number of countries reporting one case or more each month. That includes cases in countries where the disease was endemic and also cases "imported" into nonendemic countries.

were able to sustain the continued transmission of smallpox. If it was occasionally introduced into less populated central and western Asia, Europe and Africa, it might have persisted for a time in such a region as the Nile delta, but eventually the chain of transmission would have been broken. The slow pace of travel in pre-Christian times and the infrequency of long journeys could have served to confine smallpox largely to Asia during that period. In the early Christian Era, however, descriptions of a disease that was almost certainly smallpox appear increasingly in historical accounts of western Asia and, beginning in the sixth century, of Europe and Africa. In 1520 Spanish conquistadors brought the disease to the Americas.

Increasing population densities provided enough susceptible individuals to sustain the chain of transmission. So pervasive was the disease that, like measles in our time, virtually everyone contracted it. English parish records of the early 18th century indicate that about 20 percent of all smallpox victims died but that the rate was higher in many epidemics. In one well-documented instance in Iceland in 1707, 18,000 people, or 31 percent of a population of 57,000, died of smallpox. Since part of the population was immune because of previous smallpox, the death rate in that epidemic may have approached 50 percent. In Mexico 3.5 million indigenous people are believed to have died of smallpox shortly after it was introduced in the 16th century. Until the advent of vaccination smallpox played a major role in inhibiting population growth.

Even before Edward Jenner's day some people had been protected against smallpox by variolation: material from the pustules of a patient was scratched into the skin of a healthy person, who, if all went well, developed a mild case and was subsequently immune [see "The Prevention of Smallpox before Jenner," by William L. Langer; SCIENTIFIC AMERICAN, January]. Variolation spread from Asia via the Near East into Europe as a folk practice, and in the early 18th century it was taken up and popularized by physicians in England. The death rate from variolation was only about a tenth as high as the mortality from the naturally acquired disease, but a variolated person could transmit virulent smallpox to others. (Recent outbreaks in Ethiopia have been traced to the practice.) When variolation became widespread in England, it was common knowledge that dairymaids who had contracted the illness called cowpox would not "take the smallpox" when they were variolated and also appeared to be immune to natural smallpox infection. In 1796 Jenner inoculated material from a dairymaid's

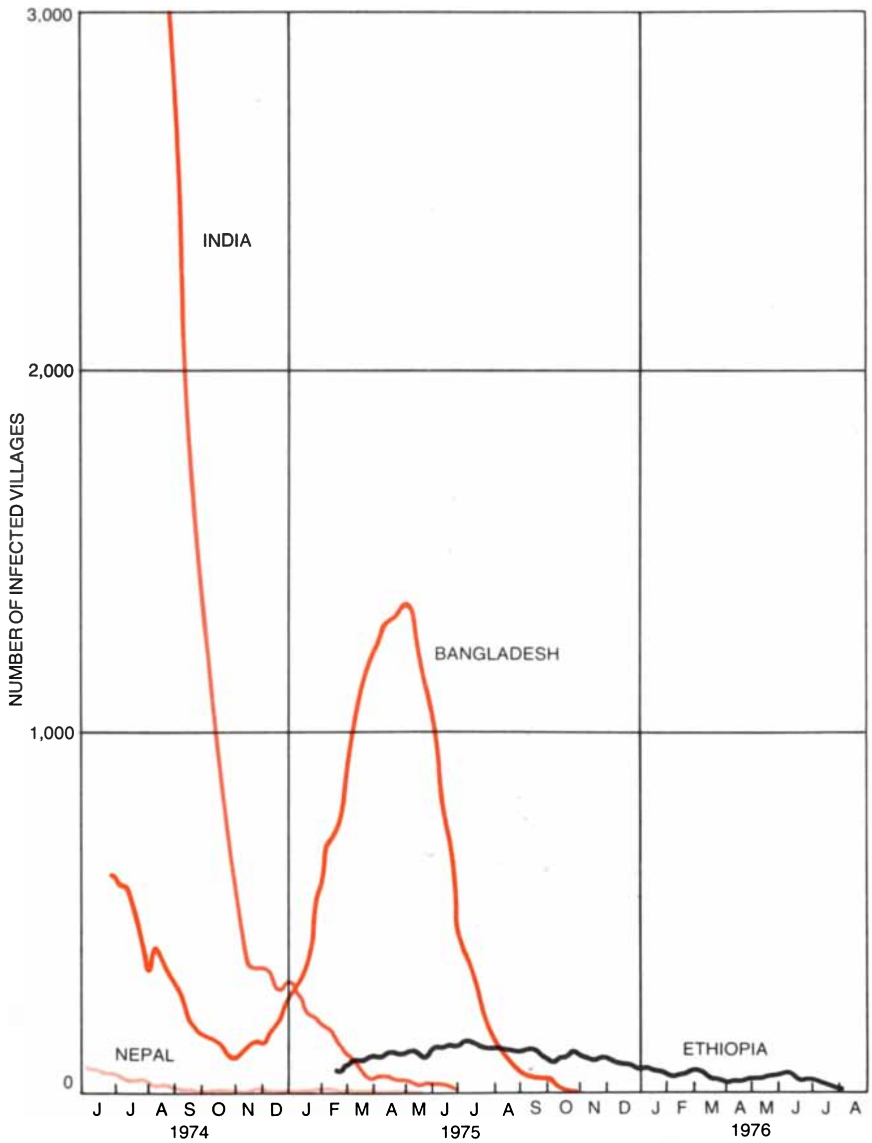


AREA AFFLICTED BY SMALLPOX has shrunk as shown on maps for 1967 (*top*), 1972 (*middle*) and 1975 (*bottom*). The countries where the disease was considered endemic, or indigenous, are shaded in color; those reporting imported cases are designated by a colored dot.

cowpox lesion into the arm of an eight-year-old boy, who was subsequently immune to the effects of smallpox variolation. Jenner called the new procedure vaccination, and he predicted that "the annihilation of smallpox must be the final result of this practice." The new technique spread rapidly, but it is only now, as the result of a concerted international public-health effort, that Jenner's prediction is being fulfilled.

Until late in the 19th century virus for vaccination was obtained from the pustular lesion of one vaccinated person and was scratched into the arm of other individuals to be vaccinated. It was an inefficient system, and not infrequently the syphilis spirochete and hepatitis virus were transferred along with the vaccine virus. The discovery that large amounts of vaccine virus could be obtained by scarifying and inoculating the shaved flank of a calf was an important advance. (That procedure is still followed to obtain sufficient quantities of virus for vaccine production.) Preservation of the virus was difficult, however, particularly in tropical climates. Sometimes the calf was led from door to door and a bit of pus was scraped from its flank for each vaccination, but usually the pustular material scraped from the calf was suspended in a 50 percent solution of glycerol in order to reduce bacterial contamination and was distributed in small glass capillaries. In this form and without refrigeration the vaccine remained potent for perhaps a few days.

Jenner himself had noted that dried vaccine lasted longer than liquid vaccine, but for years there was no way to produce dried vaccine in bulk. Methods for the commercial freeze-drying of biological preparations were developed just before World War II, and in the early 1950's a technique for producing a remarkably heat-stable freeze-dried smallpox vaccine was devised by Leslie H. Collier of the Lister Institute in England. Instead of drying the crude pulp scraped from the animal, Collier produced a partly purified suspension of virus by differential centrifugation and freeze-dried the material in the ampules in which it was to be distributed. It was reconstituted with a glycerol solution for the vaccination. Most batches of Collier's vaccine were still potent after storage for two years at body temperature. Potent dried vaccine was an essential tool that could make the eradication of smallpox possible.



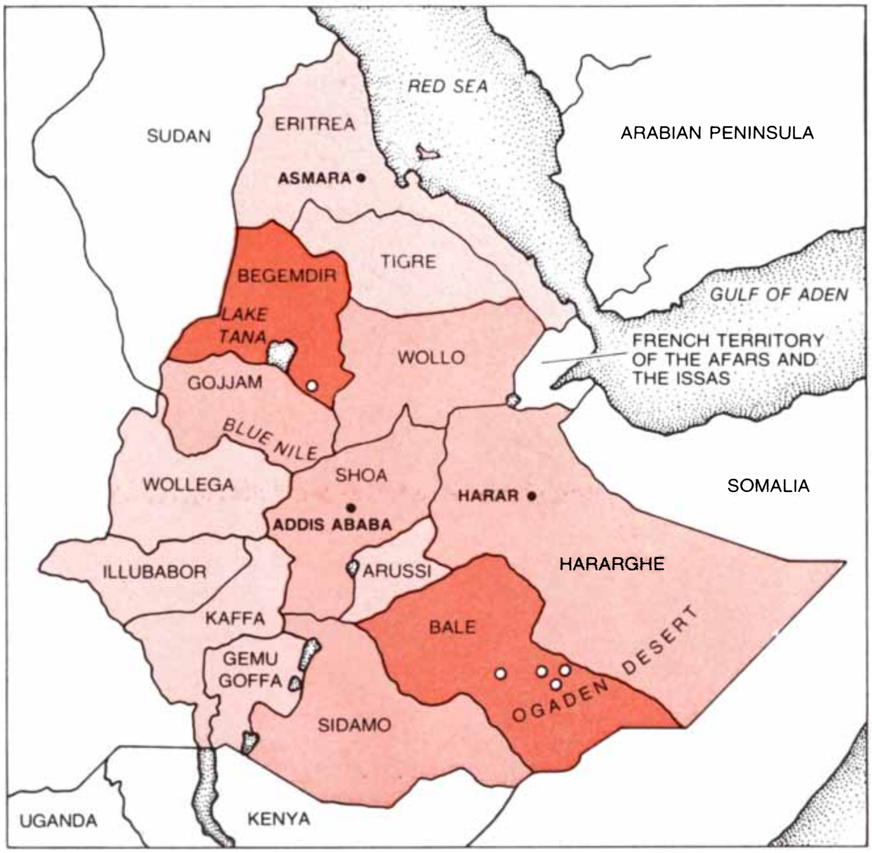
FINAL STAGES of eradication campaigns in the last four endemic countries have been monitored by a weekly count of "infected villages," defined as those reporting a case within the past six weeks. The countdown was instituted in Ethiopia only early in 1975 and it was not until the summer of that year that it was considered to represent a reasonably complete national total.

In 1926, when the Health Section of the League of Nations began publishing a weekly bulletin on disease prevalence around the world, smallpox was made a reportable disease, but only against opposition. When a Japanese delegate to the International Sanitary Conference suggested that smallpox be made reportable, a Swiss delegate maintained that "smallpox has, in reality, no place in an international convention. It is not a pestilential disease in the proper sense of the term; it is, in effect, a disease that exists everywhere. There is probably not a single country of which it can be said that there are no cases of smallpox."

When the WHO was established in 1948 as a specialized agency of the United Nations, it approached the smallpox problem gingerly at first; its interim commission had stated in 1946 that it was "impracticable as yet" even to standardize smallpox vaccines. By the late 1940's, however, the concept of disease eradication in general and the idea of eradicating smallpox in particular was acquiring a growing number of supporters, one of the most persuasive being Fred L. Soper, director of what was then the Pan American Sanitary Bureau and is now the WHO Regional Office for the Americas. They pointed to the very practical demonstration of the elimination of smallpox from North America and Europe in the 1940's and also—a more notable achievement—from a few countries with less highly developed health services, such as the Philippines and some countries in Central America. A campaign to eradicate smallpox in the Americas was undertaken in 1950, with the Pan American Sanitary Bureau providing technical assistance. The results were encouraging: by 1959 smallpox had been effectively eliminated from all American countries except Argentina, Brazil, Colombia, Ecuador and Bolivia.

In that year, on the initiative of the U.S.S.R., the World Health Assembly called for global smallpox eradication and vaccination "in foci where the disease exists." The resolution pointed out that "funds devoted to the control of and vaccination against smallpox throughout the world exceed those necessary for the eradication of smallpox in its endemic foci."

The WHO and the UN Children's Fund (UNICEF) helped a number of countries to establish vaccine-production centers. Donations of vaccine were received from several countries, and mass-vaccination programs were started by some. Progress was slow, however. A global program to wipe out malaria had captured the interest and energies of some governments and health workers, and when it faltered, many came to doubt the feasibility of any disease-eradication program. Lack of skilled personnel, vehicles and other necessary assistance kept many countries from undertaking campaigns. Those that did make the attempt were often beset by cases of smallpox imported from adjacent countries. Vaccine donations were far too small to meet the need, and there was no mechanism for regular monitoring of the quality of vaccine produced in newly established production centers. Vaccine quality declined, and countries that found themselves administering poor vaccines became disillusioned.



IN ETHIOPIA all 14 provinces were infected with smallpox in 1971, when the WHO-directed eradication program began. At the end of 1974 seven provinces (*medium and dark color*) still remained infected. At the end of August this year there were infected villages (*white dots*) in only two provinces (*dark color*), in the gorge of the Blue Nile and in the Ogaden Desert.

In 1966 delegates to the World Health Assembly argued that either enough money should be provided for a fully coordinated program or the idea of global smallpox eradication should be given up. After lengthy debate and in spite of many misgivings the delegates voted a special budget of $2.5 million for an intensive program to begin on January 1, 1967. The objective was to eliminate smallpox from the world by the end of 1976.

In 1967, the year the intensified eradication program began, smallpox was considered to be endemic in 33 countries, and 11 others reported cases attributable to importations. In the Americas, Brazil was the only endemic country, but it alone constituted half a continent. In the rest of the world there were three major reservoirs of smallpox. One, in Asia, extended from what is now Bangladesh through India, Nepal and Pakistan into Afghanistan. A second reservoir comprised virtually all of Africa south of the Sahara. The third reservoir was the Indonesian archipelago.

At the start of the program the urgent need was for adequate supplies of potent and stable freeze-dried vaccine. Little satisfactory freeze-dried vaccine was being produced anywhere. Some vaccines lacked potency when they were reconstituted and others quickly lost their potency under field conditions; some that were being administered contained no detectable virus whatever. Almost no vaccine in the endemic countries met the WHO's basic standards, and there was no central laboratory for testing vaccines.

At the WHO's request two major laboratories agreed early in the campaign to serve as international vaccine reference centers: the Rijksinstituut voor de Volksgezondheid in the Netherlands and Connaught Laboratories, Ltd., of Toronto. It was estimated that 250 million doses of vaccine a year would be needed for the endemic areas. Buying that amount would have cost more money than had been set aside for the entire program, and so a decision was made not to buy any vaccine but to ask for donations of it, and to provide equipment and technical advice to laboratories within the endemic regions so that at least the most populous endemic countries could produce good vaccine on their own.

In the first years most of the vaccine—more than 140 million doses a year—was given by the U.S.S.R. and 40 million doses came from the U.S. Eventually donations were received from more than 20 countries. Gradually vaccine production in the developing countries increased and the quality improved, until by 1970 all the vaccine in the program met accepted international standards of potency and stability.

Another early objective was to simplify and improve the vaccination techniques. The commonest technique throughout the world when the program began was some form of the scratch method: a drop of vaccine was placed on the skin and scratched into the superficial layers, sometimes with instruments that made severe wounds. The technique was wasteful of vaccine and achieved a lower success rate than an alternative method in which a needle is held parallel to the skin and the tip is pressed into the skin repeatedly. This multiple-pressure procedure is difficult to follow and to teach, however. A more efficient method that consumed less vaccine was needed.

A jet injector, originally designed for a deeper, subcutaneous injection but modified to accomplish the more superficial smallpox vaccination, had been developed by the U.S. Army. Under high pressure it injected a tenth of a milliliter of vaccine into the superficial layers of the skin. The gun required a more highly purified vaccine than the scratch methods but less than a third as much virus, and more than 1,000 people an hour could be vaccinated with the gun (if that many could be gathered together). Introduced in 1967 for programs in Brazil and Zaïre and in 20 countries of western and central Africa assisted by the U.S. Center for Disease Control, the jet guns proved to be effective, but they were expensive and required considerable maintenance and repair, making them impractical in places where trained technicians were unavailable. And they were of limited benefit in Asia, where vaccination traditionally was done on a house-to-house basis.

The ultimate solution, the bifurcated needle, was developed at Wyeth Laboratories in the 1960's. It resembles a large, blunt sewing needle with part of the eye ground off to leave two small prongs. The bifurcated end of the needle is dipped into the vaccine, and one drop, enough for one vaccination, clings between the prongs. The needle is held perpendicular to the skin and 15 quick punctures are made. Only one-fourth as much vaccine is required as for the scratch technique. Wyeth waived patent charges for the design for any manufacturer producing the needles exclusively for the WHO program. The needle had originally been designed to be used once, but the WHO had them made of special steel so that they could be boiled or flamed more than 200 times without becoming dull.

Vaccination technique was further simplified after field studies revealed that there was no difference in the incidence of bacterial infections whether or not the vaccination site had first been swabbed with acetone, alcohol or soap. That made it possible to dispense with cotton swabs and bottles of soap and alcohol. The vaccinator needs only a vial of diluent for the vaccine, a container of sterile needles and another container for used needles, all of which fit handily into a shirt pocket.

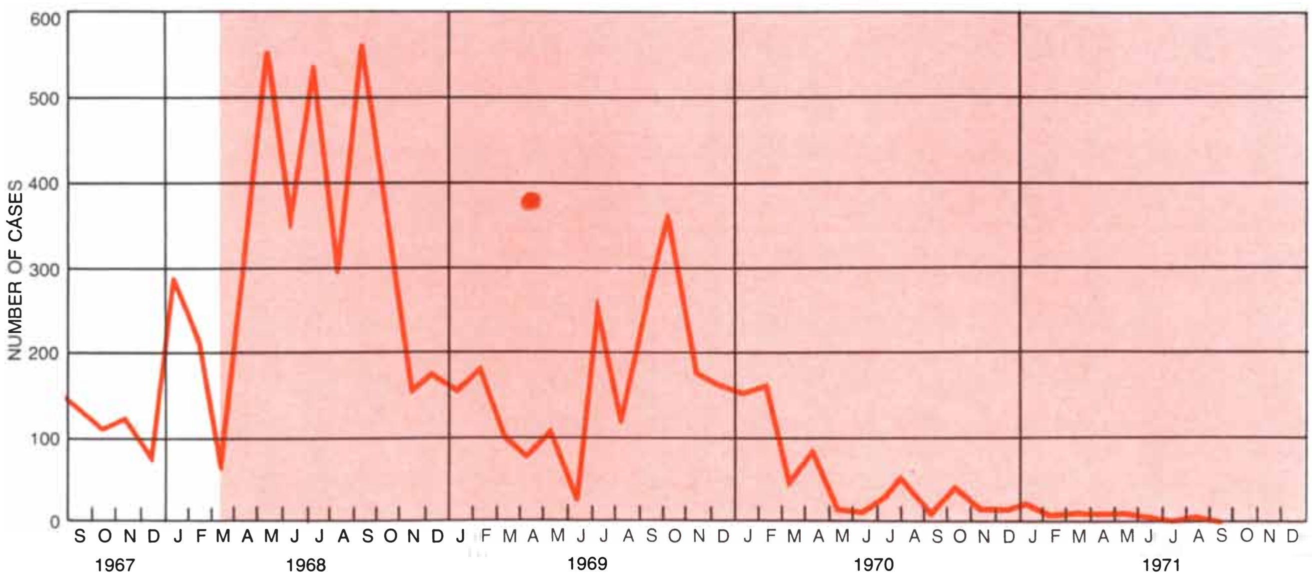
At the beginning of the program there were heavily endemic areas, but it was not at all clear how prevalent the disease was, or how extensive. It quickly became apparent that routine reporting of smallpox cases, particularly in the endemic parts of the world, was far worse than had been supposed; it was so bad that for certain years some large areas were reported as having more smallpox deaths than they had cases. One study in northern Nigeria indicated that the efficiency of reporting was 8.1 percent in an urban area and 1.3 percent in rural areas; in West Java it appeared that not more than 6 percent of all cases had been reported. In the light of nine years of experience we now believe that when the program began not more than 1 percent of all cases were actually being reported. Although 131,418 cases of smallpox worldwide were reported to the WHO in 1967, there may have been that many cases in northern Nigeria alone. An accurate figure for the world in 1967 might be on the order of 10 to 15 million cases.

A major problem was that even when cases were detected and reported by health units they were not necessarily included in national reports. At the time the campaign began in Indonesia only about half of the cases reported at the provincial level were being reported nationally; in Niger, of 325 cases reported to local government units between December, 1966, and February, 1967, only 14 were finally reported to the ministry of health. A first step in national smallpox-eradication campaigns was therefore to increase the quality and regularity of reporting from all fixed medical units. One approach was to appoint a surveillance team of two to four people for each administrative division with two to five million in population. These teams circulated among the medical units encouraging them to report, looking for cases, distributing vaccine, counseling and cajoling.

When the intensified eradication program began, the basic strategy called for a two- to three-year mass-vaccination campaign throughout a country, during which an improved reporting and surveillance system would be developed. It was felt that if 80 percent of the population could be vaccinated, smallpox incidence would be reduced to fewer than five cases per 100,000 of population; thereafter, with improved surveillance, the remaining foci would be identified quickly and measures could be taken to eliminate them.

A development early in the campaign resulted in a change in that strategy. There was a delay in the delivery of supplies for the mass-vaccination program in eastern Nigeria, and an energetic U.S. adviser, William H. Foege, organized an interim program: he searched out smallpox cases and vaccinated thoroughly in a limited area surrounding each case [*see illustration on opposite page*]. The mass-campaign supplies arrived only a few months later, but by then there was no detectable smallpox in eastern Nigeria. And less than half of the population had ever been vaccinated.

This result and similar experiences in other places led to an emphasis, even in the early stage of the campaigns, on what became known as surveillance-containment: improved search and detection as speedily as possible, isolation of patients and vaccination of every known or suspected contact around



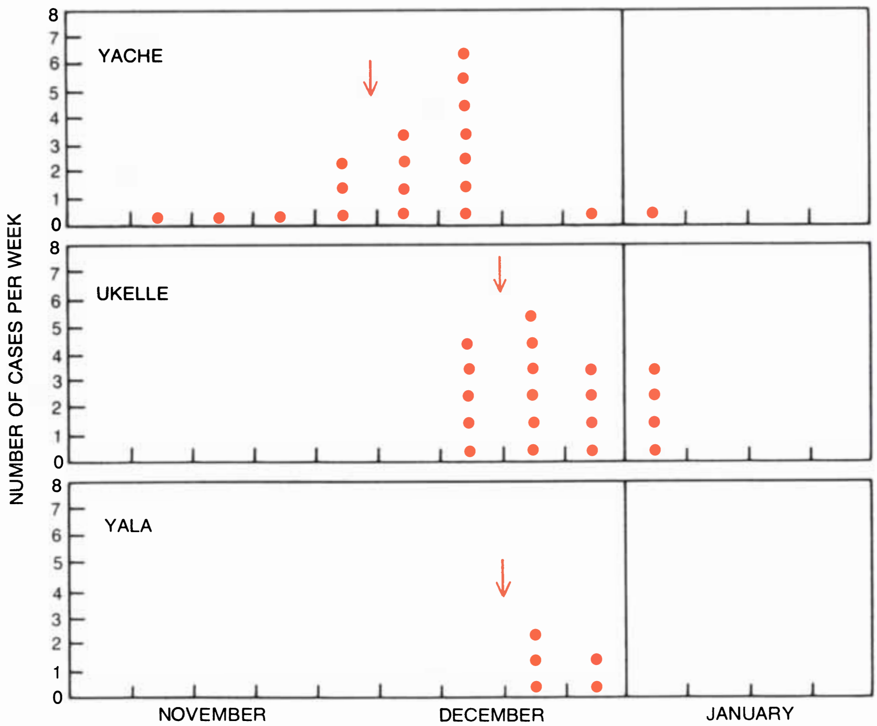
IN ZAÏRE (then known as the Democratic Republic of the Congo) an intensified eradication program began in March, 1968. The curve of reported smallpox cases is typical: as the program got under way improved reporting produced what appeared to be an increased incidence of smallpox. After six months a fairly steady but irregular decline began, and the last case was recorded during August, 1971.

them. The procedure sealed off outbreaks from the rest of the population. That proved to be effective partly because in most areas smallpox turned out to spread slower than had been supposed. One person generally infected no more than two to five others; cases were usually clustered in one part of a town or in adjacent villages rather than being widely and evenly distributed through a country. The containment teams could focus on problem areas and had somewhat more time than had been anticipated in which to seal off an outbreak. When it became clear that smallpox in a previously vaccinated person was uncommon even if the vaccination had been done many years before, heavy emphasis was placed on primary vaccination. (In Asia 80 percent of all cases were reported among the 20 percent of the people who had no vaccination scar.)

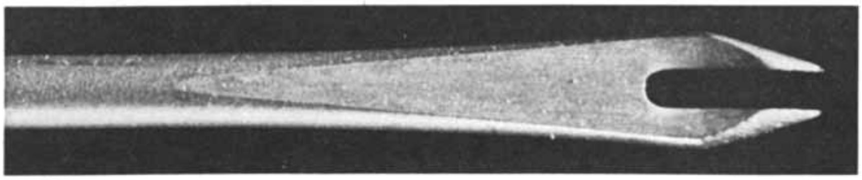
Substantial progress was made in the four years after 1967. In 1971 only 16 countries reported smallpox cases. In the 20 countries of western and central Africa where the U.S. Center for Disease Control was providing support the last case was recorded (in Nigeria) in June, 1970, a full year earlier than had been expected. Elsewhere in Africa endemic smallpox was limited to Ethiopia, whose program did not begin until 1971, to the southern Sudan, where civil war had made a program impossible, and to Botswana, which had been reinfected at about the time of the last cases in neighboring South Africa. Brazil's last case developed in April, 1971, and Indonesia's in January, 1972. One major reservoir of smallpox remained, and it was in Asia.

In 1972 an extensive area still heavily endemic with smallpox stretched from Bangladesh through northern India and Nepal into Pakistan. Here case detection was inadequate and reporting systems were archaic; the importance of surveillance and containment was not appreciated, and when containment was attempted, it was usually done poorly. Support for the program by health authorities was lukewarm; so many efforts to control smallpox had failed over so many years that the disease was widely considered inevitable and its elimination impossible.

During the 1972–1973 smallpox season in India a WHO epidemiologist, tracing the source of infection of a case in the southern state of Andhra Pradesh, followed the trail into an adjacent state that was thought to be smallpox-free. There in one district he discovered a major epidemic. It had been known to district health workers but had been concealed. Rapid containment was crucial, but that required the prompt detection of all cases throughout a district of two



OUTBREAK IN NIGERIA at the outset of the campaign was dealt with by containment, which proved so effective that it later largely replaced mass immunization in Asia. The outbreak was reported on December 4, 1966, in Yache, a village in the Ogoja district (*map at top*). Vaccination began that day in the area (*dark color*) immediately surrounding the known cases and was extended, as new cases were reported, to successively implicated areas (*medium color and light color*). The table (*bottom*) relates cases by week of onset to the initiation of vaccination. In each village the outbreak was ended within two or three weeks after vaccination began (*arrows*), except for two individuals in Yache who were not successfully immunized. The experience demonstrated that an outbreak could be extinguished by vaccinating people in a limited area around each new case even if the general area contained many unvaccinated people.



BIFURCATED NEEDLE was the principal tool of the vaccination program. The simple instrument was developed by Wyeth Laboratories. The pronged end of the needle is enlarged about 10 diameters in this photograph; the actual length of the needle is about two inches.

million people. All available health personnel were mobilized for a house-to-house search, which was two weeks in preparation and one week in execution. It revealed numerous previously undisclosed cases in addition to those recently discovered. Containment vaccination around each outbreak eliminated smallpox from the district within weeks. The experience underscored the danger of poor reporting, but it also showed what could be gained by a thorough search and demonstrated that vast numbers of health workers were deployed in India and could be mobilized quickly. Plans were made to undertake the same kind of search throughout India.

In October, 1973, the smallpox-eradication program in India conducted the first of a series of week-long searches that were scheduled once a month and involved more than 100,000 health workers. The results of the first search were ominous. The state of Uttar Pradesh, with a population of 90 million, had regularly been reporting from 100 to 300 cases of smallpox per week. The first one-week search in the state found more than 7,000 cases. And that was not all: a later assessment showed that only half of the villages officially reported as having been searched had actually been visited.

After that the techniques of case search were steadily improved. Between search weeks special surveillance teams asked questions at markets and in schools to uncover rumors of cases. A reward was offered to anyone who reported a case of smallpox and to the health worker receiving the report. As the incidence of smallpox declined the amount of the reward was increased, always with extensive publicity. Containment measures were tightened. Special books were prepared in which every person in an infected village and for one mile around it was listed. Guards hired locally were stationed day and night at the homes of smallpox patients to keep the patient from leaving and to see to it that all visitors were vaccinated.

As case detection and notification improved, the number of reported cases climbed steeply. More than 218,000 cases were reported in India in 1974, the highest total recorded since 1958. Some newspapers called it a disaster, and more than one wrote an obituary of the global eradication program. Those involved in the program, however, felt that in the search tactic they had at last the key to smallpox eradication on the subcontinent.

In June, 1974, a new unit of measurement was adopted: the "infected village." A village (or a ward in a city) was listed as infected if even one case was reported there; it remained on the list until six weeks after the onset of the last case and until an epidemiologist had certified (after a special search) that transmission had been stopped. The epidemic in India peaked in July, 1974, at which time there were more than 7,000 infected villages. The number fell steadily until in November it reached 350, but then the decline stopped. That caused concern because a new post-monsoon season with more rapid transmission of infection was beginning. Additional epidemiologists joined the fight, containment was made more rigid and in January the incidence of smallpox again began to decrease. May 24, 1975, saw the onset of the last known case of smallpox in India.

Afghanistan had become smallpox-free in 1972, and Pakistan's last case was in October, 1974. Nepal, plagued by importations from India, recorded its last case in April, 1975. Bangladesh proved to be more difficult. There was great optimism late in 1974, when the number of infected villages in the country declined to only 91. Unfortunately nearly all the infected areas were in a region struck by famine after the most devastating floods in decades. Thousands of infected refugees began moving through Bangladesh, and in spite of heroic efforts it was impossible to stem the spread of disease. (In two days in one marketplace one beggar infected 52 people from 18 villages.) A further blow came in January of 1975, when slums in Dacca, the capital city, were bulldozed and about 500,000 people were displaced. Many of these people, some of them infected with smallpox, fanned out through the country, producing dozens of satellite outbreaks; the number of infected villages rose to a peak of 1,208 in late April. Then, under a presidential directive, a national emergency program was launched and the number of infected villages decreased sharply and steadily. On October 16, 1975, the onset of the last case in Bangladesh was reported. The patient, three-year-old Rahima Banu, was the world's last known case of the severe form of smallpox, variola

REWARD FOR REPORTING a smallpox case is announced by a wall poster in Bangladesh. The Bengali writing offers a "250-taka prize" (about $17) to anyone who reports a smallpox case to a health office. Reports and even rumors of cases were followed up by health workers.

TWO SMALLPOX PATIENTS, both in Bangladesh, are pictured with their mothers. In the severe case (*left*) the pustules are so closely spaced on the face that they touch one another; the rash is less dense on the trunk. The three-year-old girl recovering from smallpox (*right*) was the last known victim in Bangladesh and thus the world's last known case of the more virulent form of the disease, variola major.

major. After that 12,000 health workers supervised by nearly 100 epidemiologists repeatedly searched Bangladesh house by house. They found no cases, and it is unlikely that any more will be discovered, but surveillance will continue for two years.

There has been some concern that smallpox, after having been declared eradicated, might emerge again from an unknown animal reservoir, from dormant virus in old scabs or from some other source. The most impressive evidence against the likelihood of such recurrence is the fact that for nine years no cases of smallpox have been discovered in the vast areas that have been smallpox-free, except for those imported from known endemic areas. Although 20 cases of an illness closely resembling smallpox have been detected in regions of Africa declared smallpox-free, in each such case "monkeypox" virus has been isolated, or identified by antibodies in victims' blood serum, as the causative factor. The reservoir of monkeypox is not known, but it is probably among rodents rather than monkeys. The disease is related to smallpox but is caused by a different virus, and its capacity for spreading from one human being to another appears to be almost nil.

What of the cost of the 10-year campaign? Approximately $83 million has been spent in international assistance for the smallpox-eradication program since 1967. The endemic countries themselves have spent roughly twice that amount, but few of them have spent much more than they were already spending on smallpox control. The total amount of money spent in international assistance is little more than half what was computed in 1968 to be the yearly expenditure for smallpox control in the U.S. alone; worldwide expenditures for smallpox vaccination and quarantine measures have been estimated as being in the range of from $1 billion to $2 billion a year. With the eradication of the disease smallpox vaccination will no longer be required, nor will international certificates of smallpox vaccination. Apart from the alleviation of human suffering, the savings have already repaid the small investment many times over.

The eradication of smallpox will represent a major milestone in the history of medicine. It will have demonstrated what can be achieved when governments throughout the world join an international organization in a common purpose. In perspective, however, the campaign must be seen as representing only a small first step toward achieving a tolerable level of public health throughout the world.

A logical next step would be to apply what has been learned in this campaign to immunization programs for controlling diphtheria, whooping cough, tetanus, measles, poliomyelitis and tuberculosis. Effective vaccines are available for all these diseases, but there has been little immunization outside the developed world. Now a number of the developing countries, given the confidence of achievement and recognizing how much can be accomplished with very little, have embarked on new immunization programs. With such campaigns one begins to see a perceptible—although far from adequate—shift from curative medicine for the rich to preventive medicine for all.

The Photovoltaic Generation of Electricity

One way to harness the energy of the sun is to convert it into electricity with solar cells of the kind carried by spacecraft. Such cells are expensive, but new kinds of fabrication may make them economically competitive

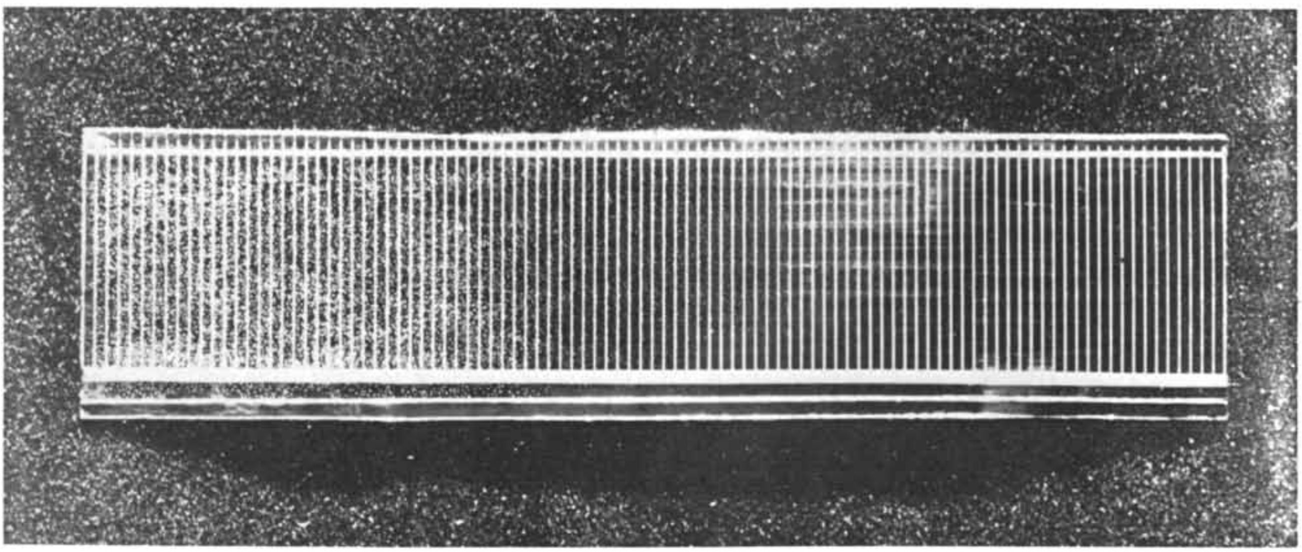
by Bruce Chalmers

It is paradoxical that Americans should be concerned about their supply of energy even as the U.S. is receiving energy from the sun at 500 times the rate at which they use it for all purposes, but perhaps the paradox can be resolved. Men have only just begun to turn their scientific and technological ingenuity to the task of converting their most abundant and reliable energy source, sunlight, into electricity, the most convenient and adaptable form of energy.

When the sun is high in a clear sky, the energy it radiates reaches the surface of the earth at the rate of a little more than a kilowatt per square meter. That is a very substantial amount of energy; while the sun is shining the energy equivalent of a gallon of gasoline (36 kilowatt-hours) falls on an area the size of a tennis court about every 10 minutes, and an area of only 80 square meters in the least sunny parts of the 48 contiguous states of the U.S. receives in the course of a year enough energy to supply the needs of an average American family. The sun is also a reliable energy source, in the sense that a given place will receive about the same amount of energy in the form of sunshine every year. It is an unreliable source, however, in the sense that sunshine is intermittent: it is not available at night or in unfavorable weather.

Before solar energy can provide electricity on a large scale there are two technical problems to be solved. The first is the problem of converting the energy of the sun's radiation into electricity. The second has to do with the fact that the electricity must be available whenever and wherever it is needed, regardless of the weather, the season or the time of day. It must also be kept in mind that in the real world both objectives must be achieved at prices that are competitive with those of other sources of energy.

It has been known for well over 100 years that light can generate electricity; the utilization of this effect to measure the strength of light in the photographic exposure meter is familiar to almost everyone. The efficiency of such devices, however, is very low. They could never be considered as a practical means of generating useful amounts of electricity. Then the discovery in the 1950's of the extraordinary electrical properties of



EXPERIMENTAL SOLAR CELL made of silicon grown by a new process known as edge-defined film-fed growth (EFG) was made at the Mobil Tyco Solar Energy Corporation in Waltham, Mass. The EFG process is one method that shows promise of lowering the cost of building photovoltaic power stations to a level comparable to the cost of building fuel-burning stations. This particular solar cell is 10 centimeters long and 2.5 centimeters wide. In full sunlight it can generate between a quarter and a third of a watt of electricity. The white lines at regular intervals are the wires of the current collector on the cell's front surface. Fine speckles are dust on surface of cell.

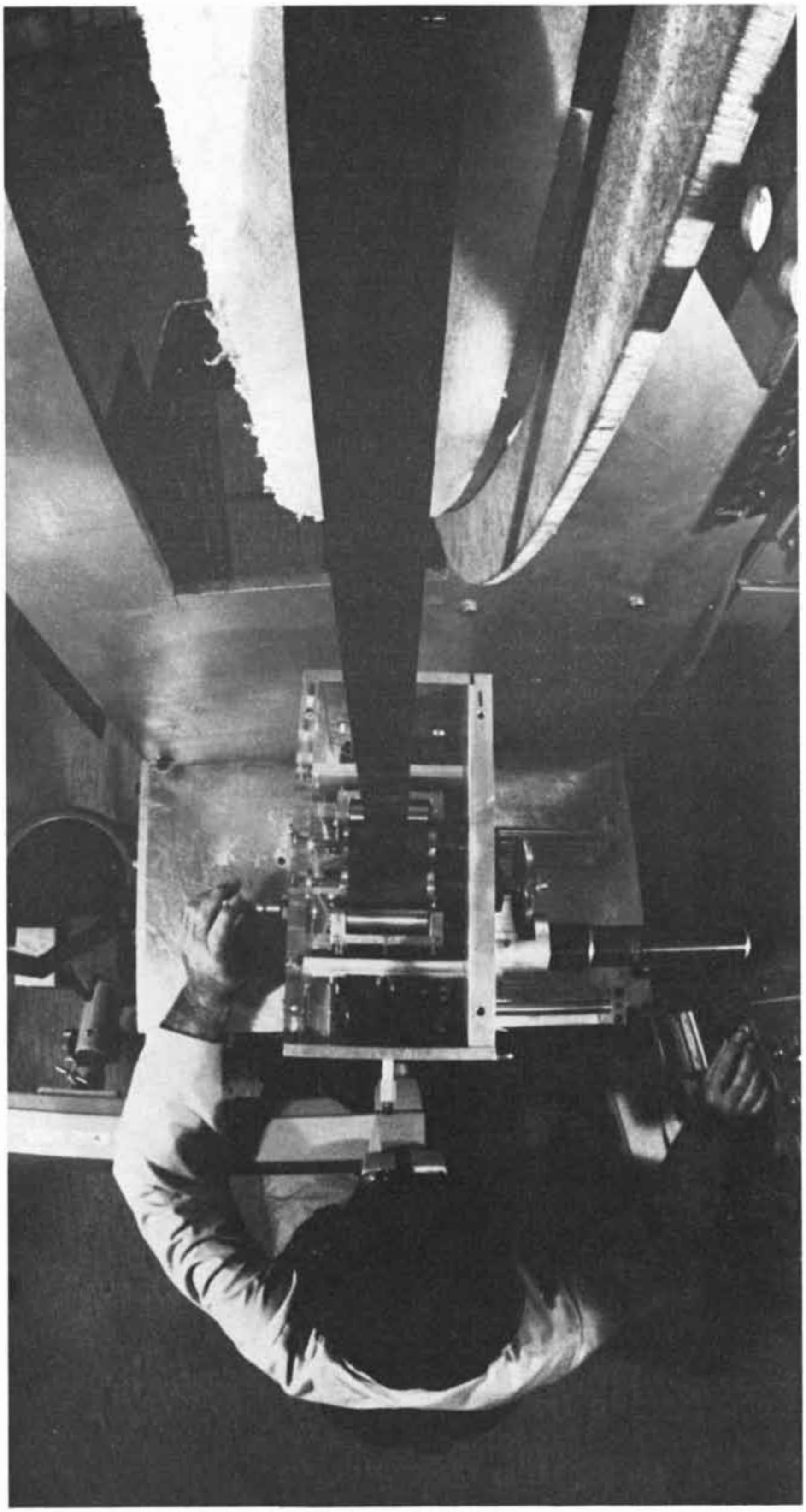
the group of materials that came to be called semiconductors led the way to the development of devices in which the efficiency of conversion of light into electricity is much higher. As has happened in a number of other branches of technology, the space program provided the incentive for the development and production of the first practical photovoltaic devices, which have provided electric power in many space vehicles, manned and unmanned, with extremely high reliability and for an apparently unlimited time.

The solar cells that have been employed so far in the space program are made of silicon. This material is not the only one that is suitable for photovoltaic devices, and in the future some alternative material may be found to be superior to silicon. Since the principles that underlie all photovoltaic devices are the same, however, it will be convenient to describe their construction and operation in terms of the silicon solar cell.

An electric current is synonymous with a flow of electrons in a circuit. We are therefore concerned with the behavior of electrons in a crystal of silicon. Silicon has a chemical valence of four, that is, of the 14 electrons that are normally found in each silicon atom four are available to interact with other atoms. They can combine with the atoms of other elements to form chemical compounds or they can interact with other silicon atoms to stabilize the structure of a crystal.

When silicon atoms are free to take up their lowest energy configuration, as when molten silicon is solidifying into a crystal, each atom surrounds itself with four equidistant neighbors. It shares two electrons with each neighbor, one of its own and one of the neighbor's, and therefore it lies at the center of a tetrahedron, each corner of which is the site of another silicon atom. The pattern can be extended ad infinitum, so that each atom, not just the central one, is surrounded by four equidistant neighbors. The result of this geometrical arrangement is the cubic crystal structure characteristic of diamond.

If the structure of the silicon crystal were perfect, each electron would be strongly held in position by the electrostatic forces between it and the two atoms it helps to bind together. Such a crystal would be a perfect insulator because no electrons would be free to move if a voltage were applied to it. In metals, which are good conductors of electricity, the atoms are also bound together by electrons, one electron or more being supplied by each atom. In a metal, however, the electrons are not localized in such a way that they bind specific pairs of atoms together; they are free to move within the crystal as an



RIBBON OF SILICON CRYSTAL is grown from a pool of molten silicon at Mobil Tyco. The ribbon is 2.5 centimeters wide and .15 millimeter thick, and it emerges through a slot in a graphite die above the melt at the rate of 2.5 centimeters per minute. For the manufacture of solar cells it is cut into pieces the same length as the cell shown in the photograph on the opposite page. The silicon crystal produced by the EFG process is not as chemically pure and crystallographically perfect as that grown by Czochralski process, used to make crystals for microelectronic devices. It is, however, adequate for making solar cells of good efficiency and is cheaper.

electric current whenever a voltage is applied.

In a crystal of silicon a substantial amount of energy is needed to break the bond between an electron and the two atoms it binds together so that the electron can become available for conducting electricity. There are several ways in which this energy, which amounts to about 1.1 electron volts, can be supplied. In one of them it is supplied spontaneously whenever an electron receives a sufficiently high concentration of thermal energy. As a result at any temperature except absolute zero there are always some electrons in silicon available for conducting electricity, and the number of electrons increases as the temperature rises. Thus a crystal of silicon is not a perfect insulator but has an intrinsic conductivity. The conductivity is low compared with that of a metal, but it is an important property of the crystal.

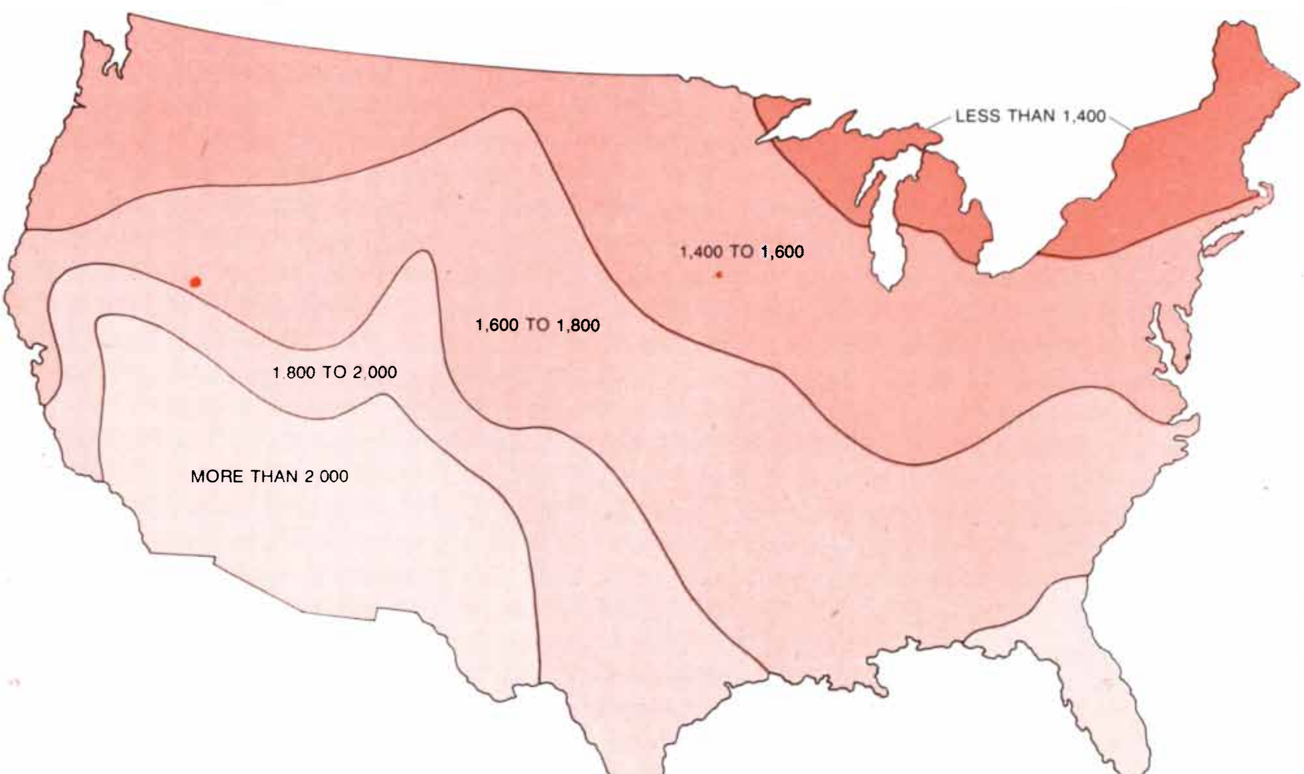
When an electron is promoted into the energy level known as the conduction band, where it can play a role in the conduction of electricity, it leaves behind a "hole": a location that lacks an electron. A bound electron from a neighboring site can move into a hole and so exchange places with it. Both electrons and holes can therefore move within the crystal, and their joint motion constitutes an electric current. When a voltage is applied, the electrons, which are the basic unit of negative charge, move toward the positive potential and the holes move toward the negative potential, as if they were positively charged.

A second way to introduce conduction electrons or holes into a crystal is to build into it atoms that supply either more or fewer electrons than the four that are required for the ideal structure. Here the energy that is needed to create holes or conduction electrons comes from the electrical misfit between the foreign atom and the silicon crystal. For example, phosphorus is an element that has a valence of five, that is, five of its electrons can interact with other atoms. A silicon crystal in which a small proportion of the atomic sites are occupied by phosphorus atoms instead of silicon atoms still has the basic structure of silicon. Nevertheless, it will have extra electrons that are in the conduction band because there are no bonding sites available to them. A phosphorus-doped silicon crystal is called an *n*-type semiconductor because it has an excess of the negatively charged electrons. Conversely, a silicon crystal can be doped with boron, which has a valence of three. Then the crystal is called a *p*-type semiconductor because it has an excess of the positively charged holes.

Doping a crystal not only increases its electrical conductivity but also makes the crystal preferentially receptive to electrons or to holes. An *n*-type crystal, which already has more electrons than it needs for bonding atoms (but not more than it needs for electrical neutrality) can readily absorb extra electrons. Conversely, the *p*-type crystal has an affinity for holes.

Light falling on the crystal is another source of the energy that can move electrons into the conduction state, creating electrons and holes in equal numbers. A photon of light striking any material is absorbed when its energy is transferred to an electron. If the electron absorbs the photon's energy near the surface of a metal, the electron may be emitted from the surface; that is the well-known photoelectric effect. If the photon penetrates within a semiconductor, and if its energy is equal to or greater than the amount of energy needed to move the electron into the conduction band, it gives rise to a conduction electron and a hole.

Thus if light of sufficient energy impinges on a perfect crystal of silicon, electrons and holes are produced and are free to move within the crystal. If the



AMOUNT OF SOLAR RADIATION received each year in different areas in the U.S. is given in terms of kilowatt-hours per square meter. The amount received in the New England states is about two-thirds the amount received in the southwestern states. In every area, of course, the amount received varies with the weather, the time of day and the season. For that reason any large-scale solar electric-power system would have to be integrated with conventional power systems or would have to have a substantial storage facility of its own.

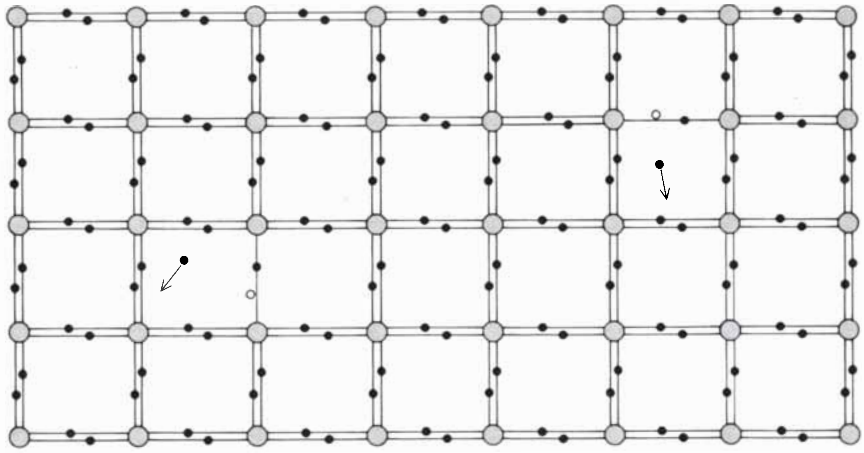
crystal is left to itself and no external voltage is applied to it, both the electrons and the holes wander through it randomly. When in its travels a conduction electron encounters a hole, it "falls" into it. The electron and the hole annihilate each other as the electron drops out of the conduction band and is again bound into the structure of the crystal. The energy that had been absorbed in originally creating the electron-hole pair is released as heat that slightly raises the temperature of the crystal.

In order to employ the electrons and holes as a source of electricity, it is necessary to arrange matters so that an electron cannot recombine with a hole until it has traveled through an external circuit, doing some useful work along the way. That can be achieved by taking advantage of the contrasting properties of *n*-type and *p*-type crystals. Consider a crystal that is composed of a layer of *n*-type silicon and a layer of *p*-type silicon. If pairs of electrons and holes are created near the junction of the two types of crystal, the affinity of the *n*-type crystal for electrons and that of the *p*-type crystal for holes will reduce the randomness of their movements through the crystal so that there is a net flow of electrons from the *p*-type crystal to the *n*-type one and a net flow of holes in the opposite direction.

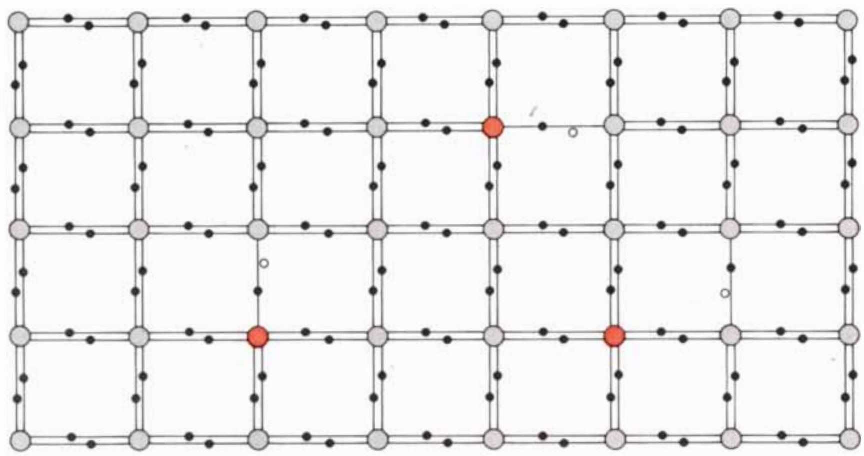
In a silicon solar cell the layer of *n*-type silicon is of a thickness such that light falling on the surface penetrates far enough into the crystal to create electron-hole pairs in the vicinity of the junction with the *p*-type silicon. That thickness is typically half a micron. Therefore when light falls on the cell, electrons will collect in the *n*-type layer and holes will collect in the *p*-type layer until there is a voltage built up within the crystal sufficient to push any further electrons back into the *p*-type layer. In a silicon solar cell that voltage is about .65 volt.

A current can be drawn from the cell through a circuit that makes electrical contact with both the front surface of the cell and the back one. It is through the external circuit that the electrons trapped in the *n*-type layer find their way back to the *p*-type layer and can then recombine with holes. If the external circuit has a very low resistance, the current that flows through it is a measure of the rate at which the electrons are separating from the holes. That rate depends on the intensity of the light falling on the surface of the cell and on the rate at which electrons and holes are lost through their recombination.

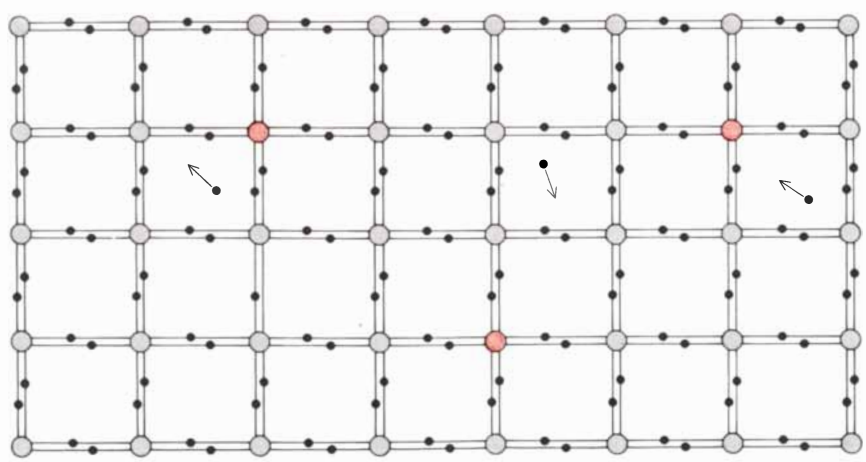
What is the efficiency of such a silicon photovoltaic device? Under optimum conditions solar energy arrives at the earth's surface at the rate of about 1.15 kilowatts per square meter. How much



CRYSTAL OF PURE SILICON has a cubic structure, shown here in two dimensions for simplicity. The silicon atom (*gray*) has four valence electrons. Each atom is firmly held in the crystal lattice by sharing two electrons (*black*) with each of four neighbors at equal distances from it. Occasionally thermal vibrations or a photon of light will spontaneously provide enough energy to promote one of the electrons into the energy level known as the conduction band, where the electron is free to travel through the crystal and conduct electricity. When the electron moves from its bonding site, it leaves a "hole" (*white*), a local region of net positive charge.



CRYSTAL OF P-TYPE SILICON can be created by doping the silicon with trace amounts of boron. Each boron atom (*dark color*) has only three valence electrons, so that it shares two electrons with three of its silicon neighbors and one electron with the fourth. Hence the *p*-type crystal has same structure as pure silicon, but it contains more holes than conduction electrons.



CRYSTAL OF N-TYPE SILICON can be created by doping the silicon with trace amounts of phosphorus. Each phosphorus atom (*light color*) has five valence electrons, so that not all of them are taken up in the crystal lattice. Hence *n*-type crystal has an excess of free electrons.

of this energy can be converted into electrical energy? In more precise terms, what efficiency can we expect in photovoltaic devices? As in determining the efficiency of a fuel-burning power station there are three component questions to be considered: (1) What would the efficiency of the device be if it had the theoretically ideal characteristics? (2) To what extent does a photovoltaic device that can actually be built fall short of the performance of the ideal? (3) How closely can a device manufactured in large numbers at an acceptable cost approach the best device that can be made when cost is no object?

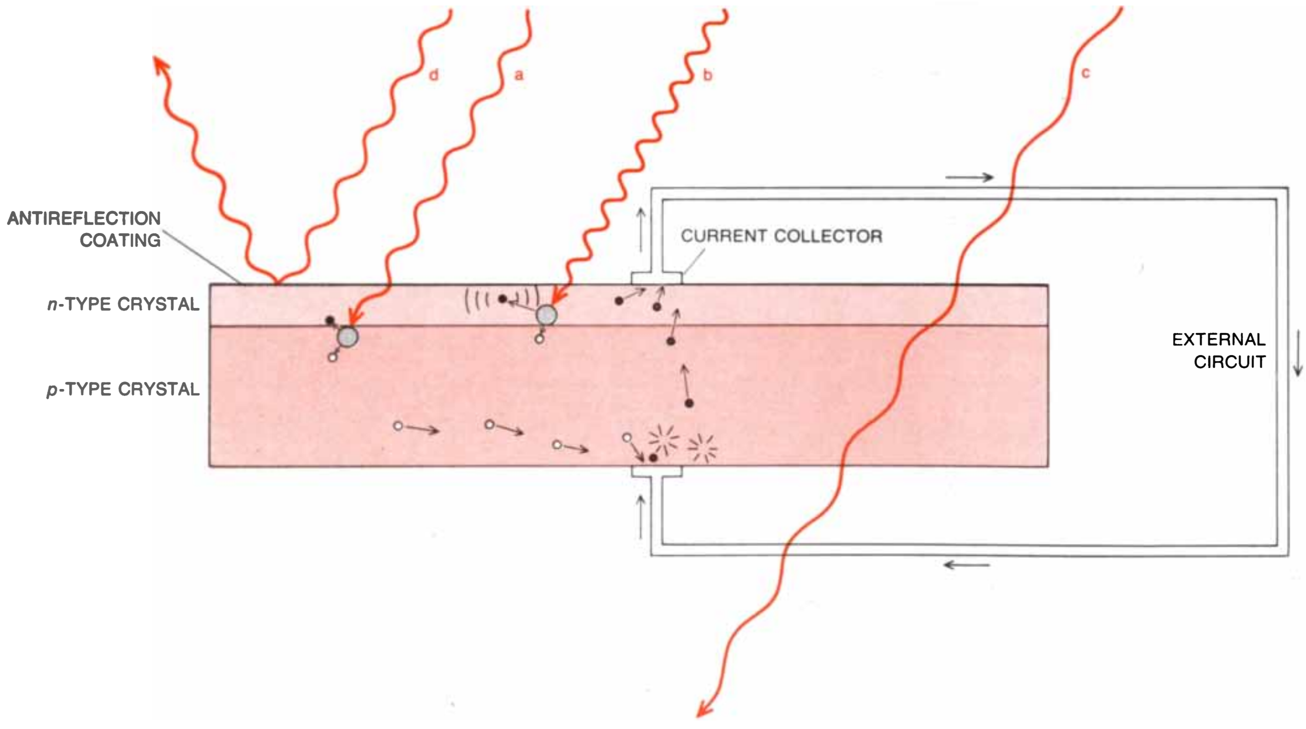
In considering the theoretical efficiency of a photovoltaic device let us look more closely at the process by which a photon is absorbed by an electron in a semiconductor. That process is the transfer of the quantum of energy to a single electron. If the photon is sufficiently energetic, the energy received by the electron will release it from its normal function as a bond between two neighboring atoms in the crystal; it becomes free to move in the crystal as a conduction electron, and a hole is simultaneously created.

The energy required to promote an electron from its bound state in a crystal into the conduction band is known as the band gap. The band gap varies considerably from one semiconductor to another; it can be expressed as electron volts or as the wavelength of the light that has the required quantum energy. Silicon has a band gap of 1.12 electron volts; thus in a silicon photovoltaic device an electron needs an energy of 1.12 electron volts in order to be moved into the conduction band. That energy corresponds to a wavelength of 1.1 microns, which is in the infrared region of the spectrum.

Radiation with a wavelength longer than 1.1 microns does not have enough energy to move electrons into the conduction band of a silicon photovoltaic device. Nearly half of the energy the earth receives from the sun arrives in the form of radiation with wavelengths longer than 1.1 microns. Hence that energy can play no part in generating electricity in a silicon photovoltaic cell. If the energy of a photon is greater than the band gap, that is, if in the case of silicon the wavelength of the photon is shorter than 1.1 microns, the energy is entirely absorbed by an electron. The electron cannot retain more than 1.12 electron volts of the photon's energy, however, and so any excess energy is immediately converted into heat. In principle if the photon had enough energy, it could promote two electrons into the conduction band. In actuality all such energetic photons from the sun are absorbed by the earth's atmosphere before they reach the ground.

The result of these physical limitations is that only about half of the photons reaching the ground can create pairs of electrons and holes in a silicon photovoltaic cell, and a substantial fraction of the energy of many of these photons is converted into heat instead of electrical energy. To be somewhat more precise, the fundamental limitations set by quantum physics restrict the maximum possible efficiency of a silicon solar cell on the ground to about 21 percent. Other semiconductors have different band gaps, and their maximum theoretical efficiencies are therefore different.

A curve can be drawn relating theoretical efficiency and band gap. The de-



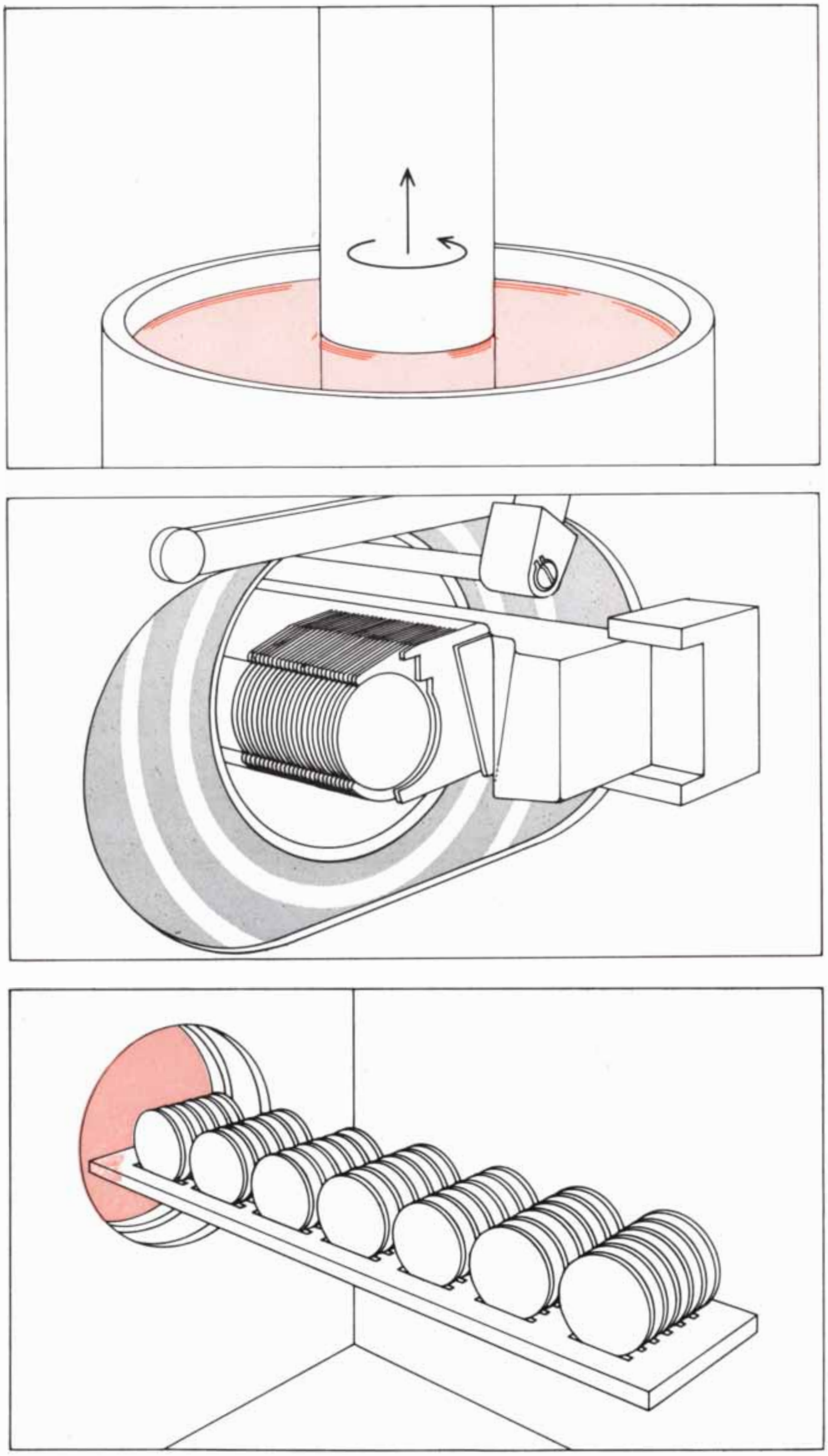
SILICON SOLAR CELL is a wafer of *p*-type silicon with a thin layer of *n*-type silicon on one side. When a photon of light with the appropriate amount of energy penetrates the cell near the junction of the two types of crystal and encounters a silicon atom (*a*), it dislodges one of the electrons, which leaves behind a hole. The energy required to promote the electron into the conduction band is known as the band gap. The electron thus promoted tends to migrate into the layer of *n*-type silicon, and the hole tends to migrate into the layer of *p*-type silicon. The electron then travels to a current collector on the front surface of the cell, generates an electric current in the external circuit and then reappears in the layer of *p*-type silicon, where it can recombine with waiting holes. If a photon with an amount of energy greater than the band gap strikes a silicon atom (*b*), it again gives rise to an electron-hole pair, and the excess energy is converted into heat. A photon with an amount of energy smaller than the band gap will pass right through the cell (*c*), so that it gives up virtually no energy along the way. Moreover, some photons are reflected from the front surface of the cell even when it has an antireflection coating (*d*). Still other photons are lost because they are blocked from reaching the crystal by the current collectors that cover part of the front surface. All these losses mean that a real silicon cell cannot convert more than about 18 percent of the solar energy it receives into electrical energy.

tailed shape of such a curve depends on the distribution of energy in the spectrum of the radiation that is received; the curve on page 41 is for sunlight outside the earth's atmosphere, which is the relevant condition for space vehicles. Where devices on the ground are concerned, however, we must remember that the sunlight is filtered by the atmosphere before it reaches them. By the time it has passed through "air mass 1," which would be its vertical path through the atmosphere, it has lost a substantial part of its shortest-wavelength (ultraviolet) radiation. The curve is therefore somewhat different in detail.

Although maximum theoretical efficiency is an important characteristic of a semiconductor, there are many reasons, less fundamental than the quantum restriction I have mentioned but nonetheless real, that rule out any hope of reaching more than about four-fifths of these efficiencies. Lest it be thought that an efficiency of less than 20 percent is unworthy of our technical sophistication, it should be remembered that the epitome of advanced industrial technology, the automobile, performs its energy conversion at an efficiency of less than 20 percent, and it performs its people-moving function at a far lower figure.

The consideration that leads to a maximum theoretical efficiency of 21 percent for silicon solar cells ignores two problems that can be minimized but never eliminated completely and a third problem that can be reduced only insofar as it is economically worthwhile. The first problem lies in the fact that the figure of 21 percent refers to the utilization of the energy of radiation that has actually entered the crystal, whereas some of the sunlight is reflected from the crystal's surface. There is no way to eliminate this reflection, but it can be minimized with an antireflection coating: a layer of transparent material of a thickness such that rays of light reflected from the surface of the crystal interfere destructively with the rays reflected at the front surface of the coating. The rays that would have been reflected thus reinforce the rays that penetrate into the crystal. Even the best antireflection coating, however, can be designed to match only a single wavelength in the solar spectrum. It seems unlikely that the loss of solar energy by reflection can ever be reduced below 5 percent.

The second problem is the resistance of the semiconductor, through which the current must pass in traveling from the vicinity of the junction between the n-type and p-type layers, where the voltage is developed, to the points where the current enters the conductors of the external circuit. The ideal collector of current would be layers of metal covering

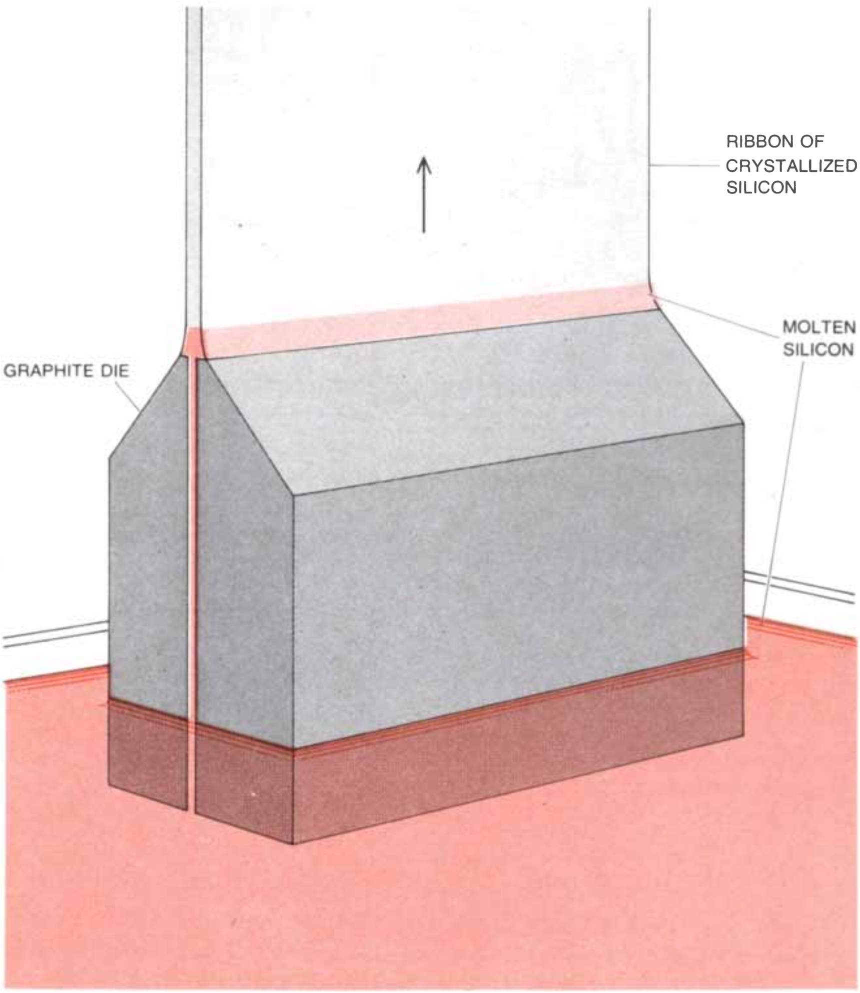


CZOCHRALSKI METHOD OF GROWING SILICON CRYSTALS is based on dipping a rotating seed crystal into a crucible of molten silicon and then slowly withdrawing it (*top*). The result is a massive cylindrical crystal three or four inches in diameter and several feet long. If an appropriate amount of boron is first added to the melt, the crystal is uniformly doped with boron. In making a silicon solar cell from such a crystal the crystal is cut into thin wafers (*middle*), a process in which a substantial portion of the crystal is lost as sawdust. One surface of each wafer, which is of p-type silicon, is converted into a layer of n-type silicon by being exposed to phosphorus (*bottom*) at a temperature high enough for the phosphorus atoms to diffuse a short distance into it. The electrical contacts are attached, the antireflection coating is applied and each cell is encapsulated in a protective skin. The finished product is a photovoltaic cell that can operate with an efficiency of between 15 and 18 percent and costs about $13.

the entire front and back of the solar cell and making good electrical contact with it. Such a collector presents no serious problem for the back of the cell. The conductor covering the front of the cell, however, would have to be transparent to sunlight. For fundamental reasons good conductors are not transparent, and so there is no way to cover the front surface with a layer that is sufficiently conductive and at the same time allows the light to reach the crystal. At present one must compromise by making the conductor cover as little of the cell's front surface as possible while minimizing the distance between any point where current is generated and the nearest point where it is collected.

That compromise presents its own set of problems. If one attempted to make a conductor out of a network of extremely thin metal wires in contact with the crystal's surface, only a small percentage of the cell's area would be obscured. The resistance of the collector would be very high, however, and energy would be lost because the wires would be heated by the current. In order to reduce that resistance in an actual solar cell about 10 percent of the front surface is covered by the collector, which of course prevents 10 percent of the available light from reaching the cell itself. Hence the performance of the cell is further reduced from its ideal performance to that extent. There are still other losses attributable to the resistance the current encounters within the crystal as it travels to the collector, even though the maximum distance it must travel is typically no more than three millimeters.

If a solar cell consisted of a perfect crystal of silicon, with the ideal quantities and distribution of the *n*-type and *p*-type dopants and with no other impurities, it should be about 18 percent efficient, and cells with that efficiency have been reported. Any departure from perfection, in terms of chemical purity or of the regularity of the structure of the crystal, leads to decreased efficiency. Impurities can cause leakage of current through the junction in the reverse direction, and crystallographic imperfections can act as traps where electrons and holes can recombine (and so be lost) and as highly conductive paths through which short-circuiting can occur.



EFG PROCESS FOR GROWING SILICON CRYSTALS is shown in greater detail. The graphite die (*dark gray*) rests in a pool of molten silicon (*color*). The silicon rises by capillary action through a narrow slit in the die and forms a layer of liquid silicon on the top of the die. A seed crystal (*not shown*) is dipped into the liquid at the top of the die and is drawn upward at the same rate as that at which the crystal (*light gray*) is growing downward, so that interface between crystal and liquid silicon remains a fraction of an inch above top of die. Silicon crystal grown in this way crystallizes into ribbon that can be cut into wafers with minimum of waste.

Solar cells of the type I have been describing were developed for the space program, and they provide the electric power for all our space vehicles. They are a completely satisfactory solution to the problem of supplying energy without fuel of any kind, but they are expensive. In comparing the cost of electric power generated by solar cells with the cost of the power generated by the methods employed by electric utilities, the first criterion is the capital cost per kilowatt of output. It is reported that in 1959 the capital cost of solar cells was $200,000 per kilowatt. They can now be bought for nonspace purposes for about $20,000 per kilowatt, but production is very limited (amounting to the equivalent of 100 kilowatts per year). The cost of building a fuel-burning power station is about $500 per kilowatt of electric-power output. Obviously the cost of solar cells is far too high to outweigh the fact that no fuel is needed.

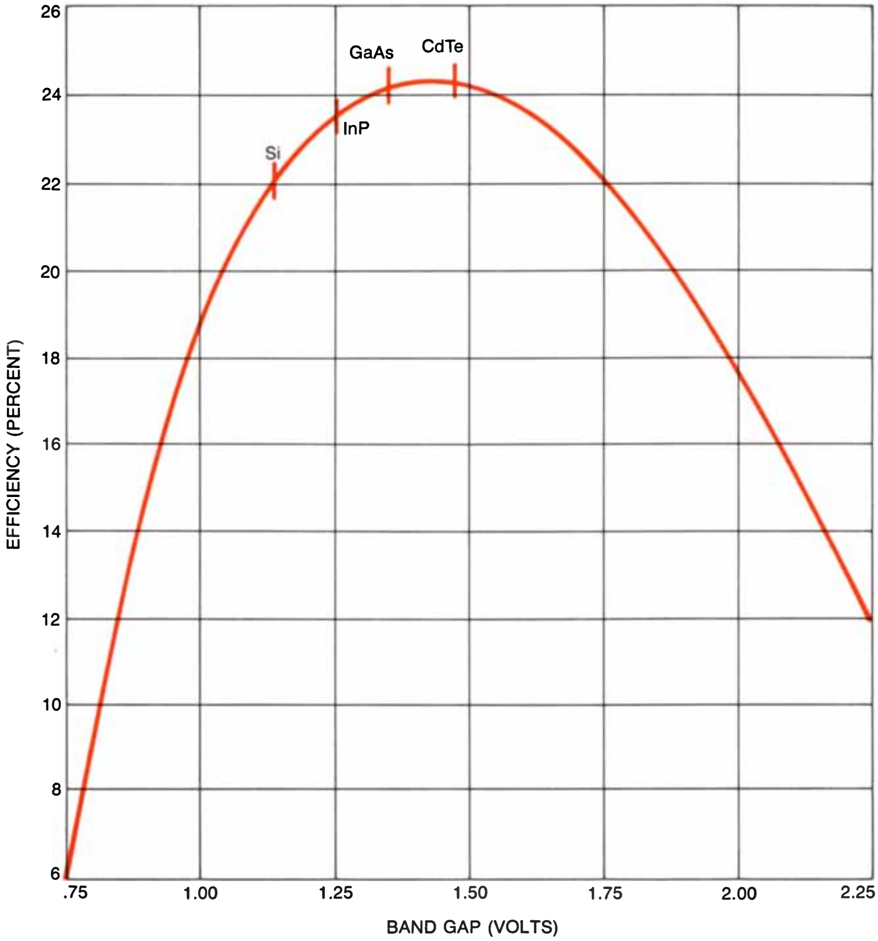
Why is the cost of photovoltaic devices so high? The reason is the chemical purity and crystallographic perfection that are needed to enable the solar cell to perform at its highest efficiency. The processes currently employed for attaining chemical purity and crystallographic perfection are inherently expensive. They are the same processes by which crystals are made for many purposes, such as microelectronic devices, where the highest possible quality is an absolute requirement. If photovoltaic cells are ever to be harnessed on a large scale for converting solar energy into electricity, good solar cells will have to be made cheaply. In the light of that prerequisite it is reasonable to ask: Is the manufacture of a solar cell intrinsically expensive, or is there any possibility of reducing the cost by a factor of 40, which would bring it into a competitive price range?

The process by which solar cells are currently made requires as a starting material silicon of "semiconductor grade," a very expensive material, costing about $70 per kilogram. This is not because of any scarcity of the element (silicon is one of the most abundant elements in the earth's crust) but because of the complexity of the purification process. Perfect crystals of silicon can be grown by the Czochralski method, a process in which a seed crystal is rotated while being slowly withdrawn from a crucible of molten silicon. If the temperature and the rates of rotation and pulling are controlled with sufficient precision, the silicon in the crucible is converted into a single perfect crystal. If a suitable quantity of boron has been added to the melt, the crystal is uniformly doped to the required extent. Such a silicon crystal is quite massive: it is typically a cylinder three or four inches in diameter and several feet long.

The next step in manufacturing a solar cell is to slice the silicon crystal into thin wafers, each of which will eventually be an individual cell. To avoid damaging the crystal the wafers are sawed from it quite slowly. The typical wafer is a quarter of a millimeter thick, and at this point it is a uniform crystal of *p*-type silicon. A thin layer of it is converted into *n*-type silicon by exposing one surface of the wafer to phosphorus at a temperature high enough for the phosphorus atoms to diffuse into the wafer to a depth of about half a micron. Finally the electrical contacts are attached to the front and back of the wafer, the antireflection coating is applied and the entire cell is encapsulated in a protective skin.

The product of this series of operations is a cell that will generate between 10 and 15 milliwatts per square centimeter. Since the area of a typical cell is 45 square centimeters, if the cell operated at an efficiency of, say, 15 percent, it would generate up to two-thirds of a watt of electricity when it was exposed to full sunlight. At current prices such a silicon solar cell would cost about $13. The high cost is partly due to the fact that a large proportion of the initial crystal of silicon is wasted as sawdust when the wafers are cut. The principal reason the cell is so expensive, however, is that such a precise manufacturing process calls for much time and labor on the part of the skilled workers who grow the crystals, slice the wafers and fabricate the finished cell.

The output of a solar cell of a given size is proportional to the intensity of the radiation falling on it. It might therefore be economically sound to concentrate sunlight onto a relatively small area of solar cells. That could be done by focusing the light with a parabolic mirror. Such a mirror, however, has two disadvantages. First, it would focus the radiation in such a way that it would strike the cells at various angles, whereas the cells perform best if the radiation strikes them at right angles to their front surface. Second, it would be necessary to rotate the reflector to keep it pointing at the sun throughout the day, which would require a fairly precise automatic tracking device.

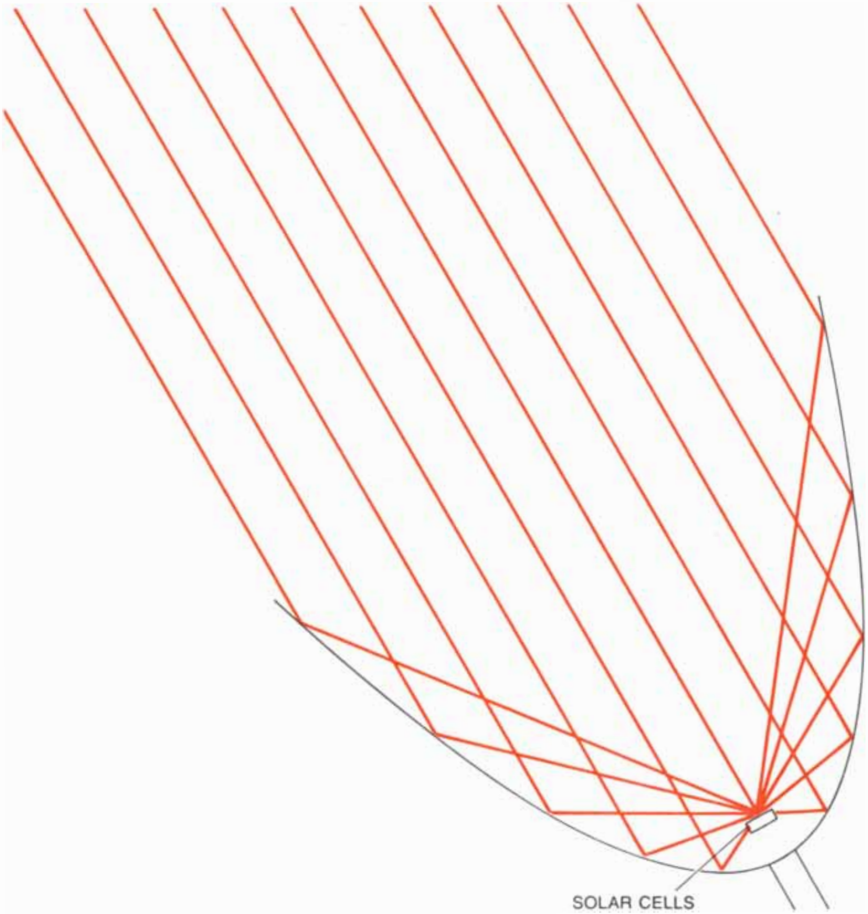


THEORETICAL MAXIMUM EFFICIENCY of a photovoltaic cell depends on the band gap of the semiconductor of which it is made. The band gap is given in terms of electron volts. In this diagram the efficiency of four different semiconductors is shown: silicon (*Si*), indium phosphorus (*InP*), gallium arsenide (*GaAs*) and cadmium tellurium (*CdTe*). Fundamental restraints of physics limit efficiency of even an ideal photovoltaic cell to less than 25 percent.

A promising alternative to a parabolic mirror is the trough-shaped reflector known as a Winston collector. The collector is designed to reflect a high proportion of incident sunlight onto a small area even when the sun is not directly in front of it, and so it is not necessary to have the collector track the sun. It has been demonstrated that a concentration ratio of eight to one is feasible, and that the performance of the cells does not suffer as long as they are not allowed to get too hot. It is therefore possible to increase the output of a given area of solar cells by a factor of at least eight, reducing the capital cost of the cells per kilowatt by the same factor. Whether or not this is economically desirable will depend on the cost of Winston collectors in relation to the cost of solar cells.

Is it necessary to maintain such extremely high standards of quality in the silicon crystals for solar cells? It is true that if a crystal is intended for a microelectronic device, even the slightest imperfection can render it useless. The requirements for a solar cell, however, are much less stringent, and the desire for high efficiency must be weighed against the high cost of perfection. As we have seen, a perfect crystal would have an efficiency of about 18 percent. Crystals of somewhat lower quality could have an efficiency of between 10 and 12 percent for a fraction of the cost. The capital cost per kilowatt of output could thus be much lower. A substantial effort, financed partly by the Energy Research and Development Administration (ERDA) and partly by private companies, is being concentrated on the development of less costly methods of making sufficiently pure silicon and of converting it into wafers from which solar cells can be made.

One of the most promising methods calls for continuously growing a silicon ribbon of the width and thickness required for solar cells. The process, known as edge-defined film-fed growth (EFG), was invented at Tyco Laboratories and is being developed by the Mobil Tyco Solar Energy Corporation. In this



PARABOLIC REFLECTOR is one device for concentrating sunlight onto an array of solar cells in order to cut down the number of cells. Such a reflector has the disadvantage that in order for it to work well it must have an automatic driving mechanism to keep it pointed at sun.

process the molten silicon rises by capillary action through a slot in a die made of graphite. A film of liquid silicon at the top of the die feeds the growing crystal, whose cross section is defined by the shape of the top of the die. The ribbon of silicon, .15 centimeter thick, can be grown at rates of more than 2.5 centimeters per minute; it is then simply cut with a minimum of wasted material into rectangular wafers 2.5 centimeters wide and 10 centimeters long. The crystals grown by the EFG process approach perfection far less closely than those grown by the Czochralski process, but they can nonetheless be made into solar cells with an efficiency of between 10 and 12 percent.

The EFG process is not the only way to produce cheap silicon crystals that are good enough for solar cells, and as I have mentioned silicon is not the only possible material for such cells. For example, solar cells with an efficiency of 18 percent have been made out of gallium arsenide. Moreover, thin films of other semiconductors that can be deposited from a vapor onto a substrate may eventually compete with silicon wafers in terms of dollars per kilowatt. The EFG process has not yet been automated to the point where it can produce cheap cells. Close analysis indicates, however, that it has the potential to provide crystals for solar cells that will cost less than the magic figure of $500 per kilowatt.

Exactly how would such cells be employed to convert into electric power the intermittent and somewhat unpredictable energy of sunlight? In the first place it should be kept in mind that the cost per kilowatt of the cells will not immediately drop to an economic level, because low cost cannot be achieved until there is production on a very large scale. There are, however, smaller-scale demands that might assist the transition. Pumping water for irrigation or powering television sets (with battery storage) in remote and sunny parts of the world are applications for which even quite expensive solar cells might be the cheapest power supply. The market could be large enough to support the early years of growth needed for a new solar-cell industry.

Such an intermediate stage, however, is relatively insignificant in terms of the national energy program of the U.S. The more important questions are whether photovoltaic devices can provide enough energy to reduce the amount of oil the U.S. imports or the number of nuclear power stations it builds, and if they can, when it might come about.

There are three quite distinct possibilities. The first is that very large solar power stations could be built, presumably in the deserts of the U.S. Southwest, where sunshine is plentiful. At an efficiency of 10 percent a collecting area of 10 square kilometers (a square 3.16 kilometers on a side) could generate a peak power of 1,000 megawatts, the typical output of a large fuel-burning power station. Although the output from solar cells is in the form of direct current, it could readily be converted into alternating current and fed into the distribution network of an electric utility.

Electricity supplied in this way could not, however, replace other means of generating electricity because it could not be relied on to provide power at all times when it is needed. Solar power stations could nonetheless reduce the demand on fuel-burning stations, and hence conserve fuel, by taking on part of the demand when the sun is shining. Such a function is sometimes called negative storage.

Alternatively, solar power stations could generate electricity when the sun is shining and store it for use when sunlight is unavailable. The energy could be stored by pumping water into elevated basins; the energy would be retrieved when the water was allowed to return to its original level through turbogenerators. It may eventually be possible to store solar energy by manufacturing hydrogen through the electrolysis of water. The hydrogen could be stored and subsequently burned instead of natural gas, or it could be fed into fuel cells to generate electricity.

Another way in which photovoltaic devices could supplement more conventional sources of power would be to deploy large panels of solar cells on space stations in stationary orbit around the earth. Energy gathered from the sun could be transmitted to terrestrial receiving stations by microwave beam. This scheme would have the advantage of providing energy continuously (if more than one space station were used), so that there would be no need for storage. Moreover, the solar cells would collect solar energy in space, where it is twice as intense as it is at the earth's surface. There is no doubt that such a power-generating space station could be built, but it would require an enormous investment. There might also be objections on the basis that the microwaves could be hazardous to life on the ground, and that the system would be vulnerable to an enemy.

The electric-utility industry has evolved in an era where the economics of power generation have demanded highly centralized power production. This requirement has resulted from the fact that it is highly inefficient to generate electricity in one-kilowatt amounts by burning fuel. The cheapest way to generate electricity by present methods is with power stations that operate at the level of 1,000 megawatts. Such considerations do not, however, hold for solar cells. For them the economy of scale applies to the production of the cells but not to their utilization, since they have the same efficiency whether the amount of electricity is large or small.

It can be argued that it might be cheaper for the electricity needed in the home to be generated by solar cells on the roof and stored in batteries. The solar cells could be integrated into a system that would provide heating, cooling and hot water by utilizing part of the 88 to 90 percent of the solar energy the cells are unable to convert into electricity. Furthermore, if the electricity were generated at the site where it is needed, the cost to the consumer for distribution would be eliminated. It is not likely that such a photovoltaic system could supply all the power needed in the home. Nevertheless, if the system supplied two-thirds of the power at a reasonable cost, the need for new power stations would be reduced, and the community would be less dependent on supplies of oil, natural gas, coal and uranium.

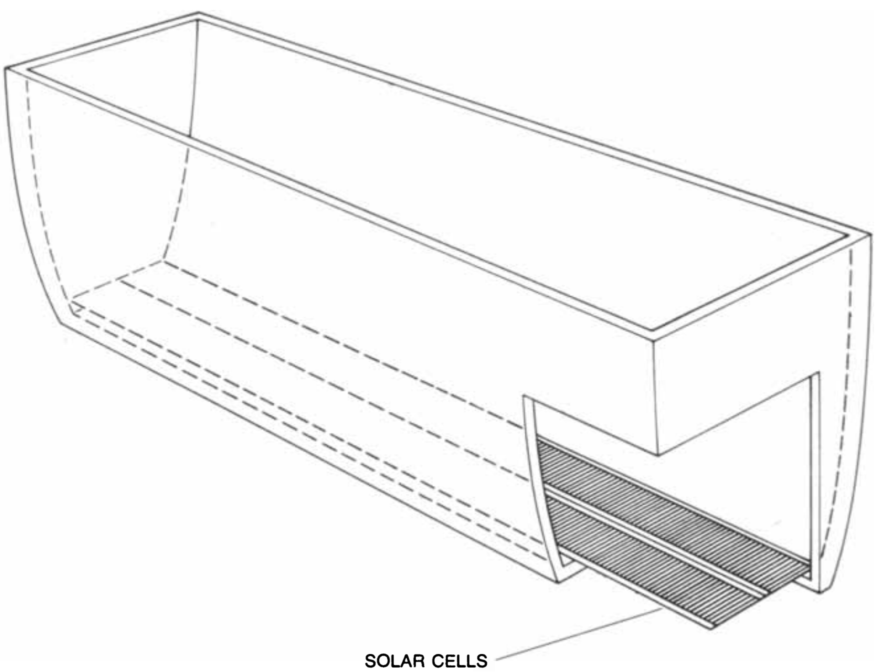
How soon might we hope to reap the benefits of this new technology? It depends partly on how quickly devices that work in the laboratory can be manufactured on an industrial scale, and that is a question of how fast the necessary engineering development can be achieved. No new scientific breakthrough is needed, only a great deal of meticulous engineering to put proved ideas into practice. Even if we had today a process complete in all details for making solar cells at a cost of $500 per kilowatt of peak output, however, it would still take five years to build up an industry that could produce the solar cells and the associated equipment to generate electricity on the megawatt scale. And it might take another five years beyond that before a significant fraction of new electric-generating capacity could be photovoltaic.

ERDA's goal is several 10-megawatt photovoltaic demonstration systems by the early 1980's and a significant contribution, perhaps exceeding 1 percent, to our total energy consumption (equivalent to nearly a million barrels of oil a day) by the turn of the century. A much faster buildup of photovoltaic generating capacity would be possible, but it would require much more capital and much more engineering effort than are likely to be deployed. Demonstrating the feasibility of generating electricity on a large scale with solar cells will not in itself, however, guarantee their widespread adoption; that will depend on who can sell the product of his electric-generating system at the lowest price.

The capital cost of $500 per peak kilowatt for photovoltaic systems is considered a realistic aim, but how valid a criterion is it for whether such systems will be economically competitive? At first $500 per peak kilowatt looks attractive, but it must be remembered that a conventional power station (also costing $500 per peak kilowatt) can generate electricity and therefore also generate a return on its capital investment, day and night, rain or shine. Official statistics show that on the average the power stations of the U.S. are "on line" for about 50 percent of the time. That means they generate about 4,400 kilowatt-hours per installed kilowatt per year. For a solar generator, on the other hand, the amount of electricity generated per kilowatt of peak output would be about 2,100 kilowatt-hours per year in Texas and about 1,400 kilowatt-hours per year in the northeastern states. Therefore if a solar installation is to provide as much energy as a 1,000-megawatt fossil-fuel or nuclear power station, it would need to have a peak output of between 2,000 and 3,000 megawatts. The future price of fuel will determine whether a factor of two or three in capital cost will be outweighed by the zero fuel cost of a solar installation.

The economic factors are considerably more favorable to solar generators in those places where the peak demand for electricity comes because air conditioners are going full blast. At such times, of course, solar energy is at its most abundant. Having solar generators as standby equipment for service under those conditions, however, could never make more than a small contribution to the overall demand for electric power. Still another set of economic considerations must be applied to the solar generation of electric power at the point of use. Since a substantial part of the cost of electricity to the consumer is represented by the distribution, as distinct from the generation, of electricity, at $500 per kilowatt of peak output the cost per kilowatt-hour to the homeowner would be well below what the electric utilities charge.

At the present time a photovoltaic power station costing $500 per kilowatt of peak output (if it could be built) would not be competitive with a station running on fossil fuel or uranium. As the price of such fuels increases, however, and as environmental restrictions further increase the cost of generating electricity with them, solar energy will become more attractive. The photovoltaic generation of electricity might find its place in the sun sooner than some now expect.



WINSTON COLLECTOR, invented by Roland Winston of the University of Chicago, is in the shape of a long trough. It concentrates sunlight on a strip of solar cells and has the advantage that it does not need to be rotated in order to follow sun across sky in course of the day.

Neutron-scattering Studies of the Ribosome

The organelle of the living cell where proteins are formed consists of 58 large molecules. How the molecules are arranged in space can be determined by irradiating the ribosome with a beam of neutrons

by Donald M. Engelman and Peter B. Moore

An understanding of biological structure can be approached in two ways. A method of long standing consists in analyzing living materials at an ever finer scale, examining first tissues, then cells and finally the component parts of cells. The principal tool of these investigations has been the microscope, including both optical instruments and in recent years the electron microscope. The second method proceeds in the opposite direction, from smaller structures to larger ones, by defining the arrangement of atoms in individual molecules of increasing complexity. In this endeavor the most important technique is the study of X-ray-diffraction patterns; over the past 20 years such studies have revealed the spatial arrangement of the atoms in a few dozen proteins, a remarkable achievement.

Between these two realms there is a gap: some features of cells are too small for their structures to be determined by electron microscopy but too large to be reconstructed in atomic detail from X-ray-diffraction studies. An example is the ribosome, an organelle found in all cells (and commonly found in large numbers) on which the genetic code is translated into its final form: the sequence of amino acids in a protein molecule. The ribosome is large enough to be visible in electron micrographs, but only the outline of its form can be discerned; the amount of structural information that can be extracted from these micrographs is limited. On the other hand, the ribosome is much larger than the structures that have so far been found suitable for analysis by X-ray diffraction. The ribosome is made up of 55 separate protein molecules (each a complex entity in itself) and three molecules of RNA, all formidably large.

Structural information at this intermediate scale is essential for an understanding of how biological macromolecules such as proteins are organized into functional units. In the case of the ribosome one can hardly expect to understand how the organelle performs its task in protein synthesis without first knowing how its component parts are assembled.

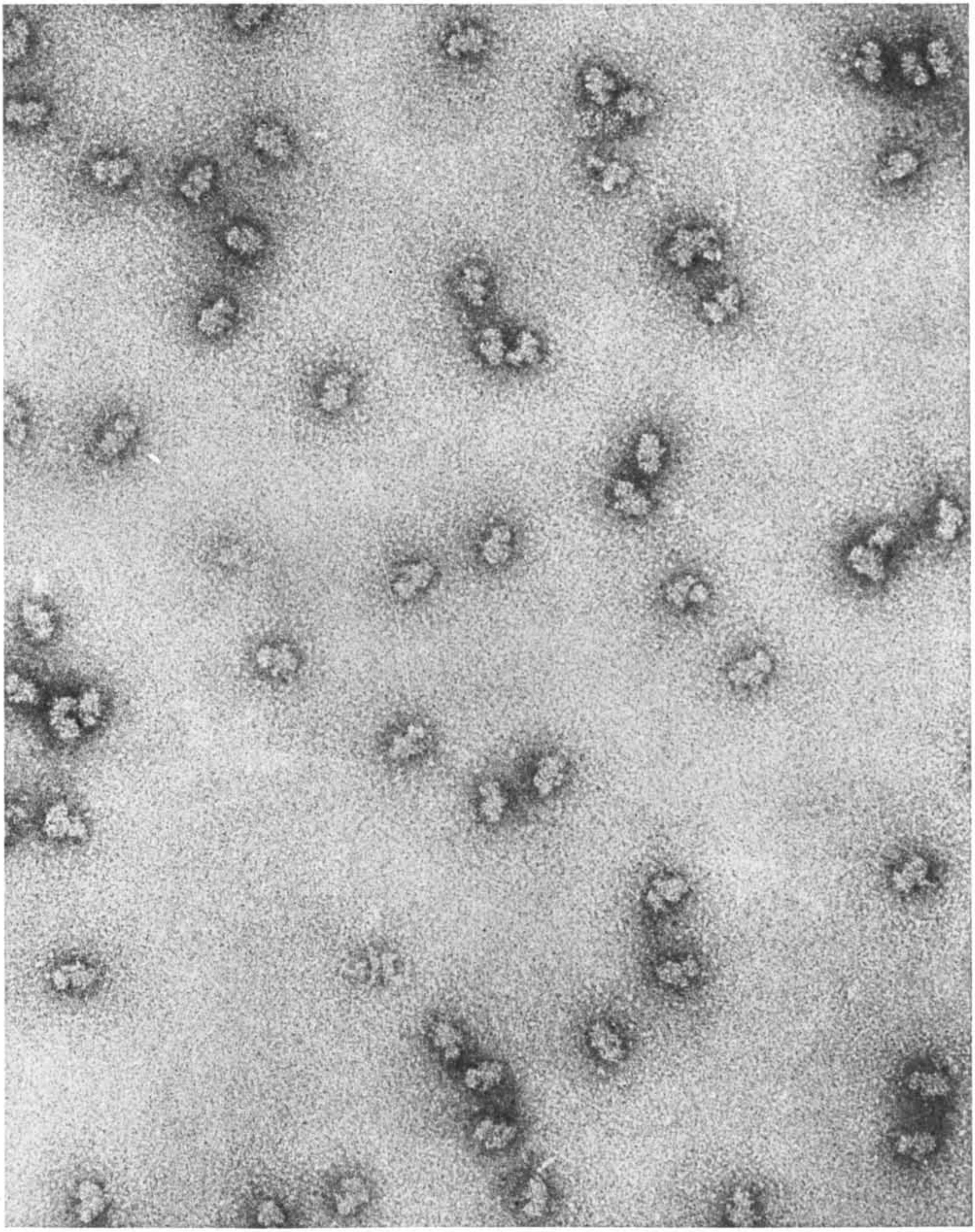
Over the past four years we have developed a method for determining the structure of macromolecular complexes of this size. In the past year or so we have begun to apply the method to the study of one subunit in the structure of the ribosome. The basis of our method is the scattering of a beam of neutrons by specially prepared ribosome subunits. The end result of our investigation, if it is completely successful, will be a map showing the arrangement in space of all the proteins in the subunit.

The "central dogma" of molecular biology states that information in the cell flows from DNA to RNA to proteins. The genetic code represented by the sequence of bases in a molecule of DNA is first copied to produce a molecule of RNA, which is exported from the nucleus into the cytoplasm as messenger RNA. The messenger RNA then serves as the template for the synthesis of a protein, but that synthesis can be accomplished only in the presence of ribosomes. The ribosomes coordinate the translation of the code: they interact with the messenger RNA and bring together the other molecules required for the orderly elongation of the protein chain. Principal among these are the component units of the growing protein: the amino acids, each of which is borne by a particular species of another kind of RNA, transfer RNA. Several factors that serve a catalytic function are also required, and the energy expended in forming the bonds between the amino acids is supplied by the energy-rich phosphate bond in the small molecule called guanosine triphosphate.

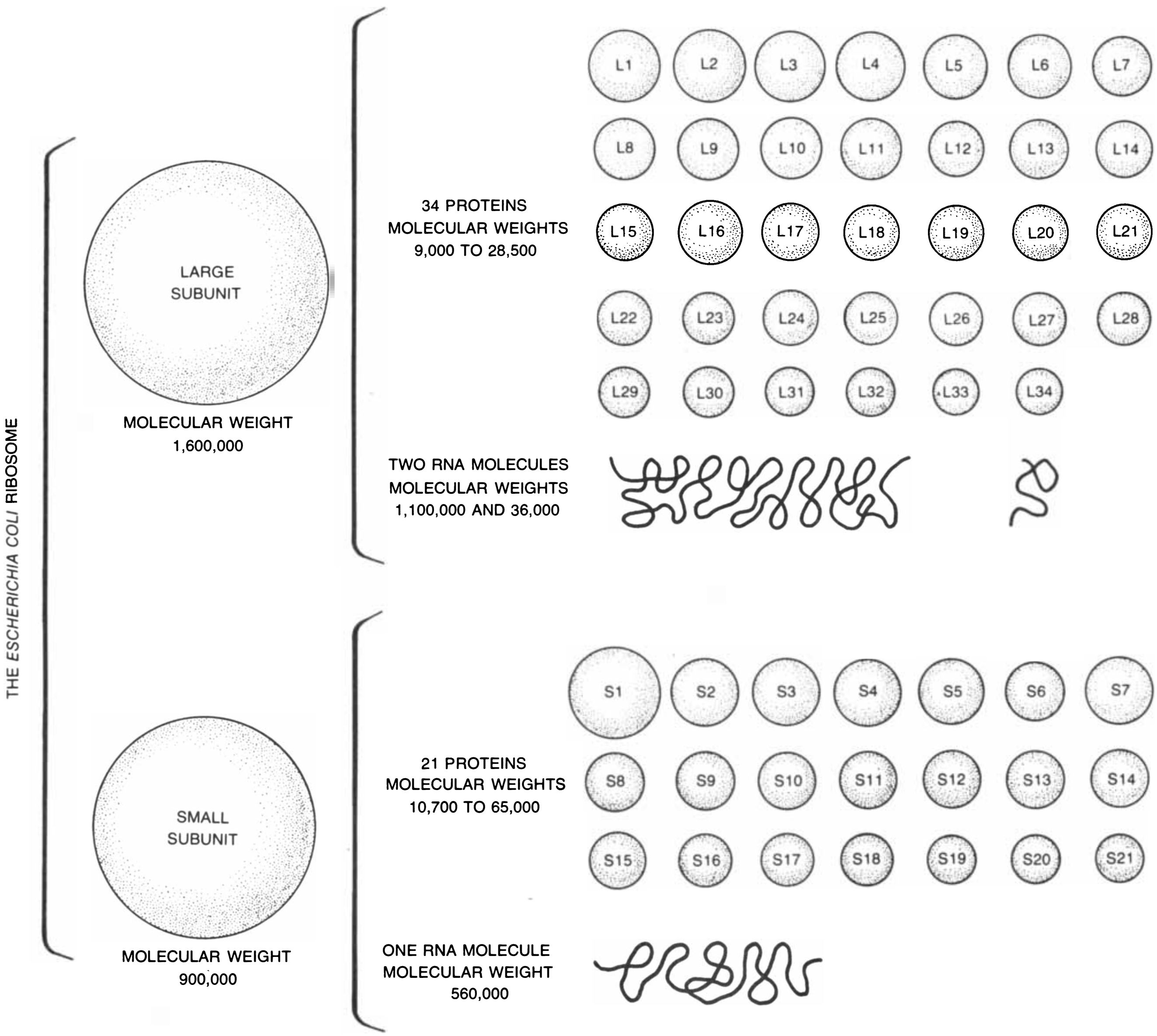
The ribosome obviously occupies a central position in the flow of information from the gene to proteins, but we have only a rudimentary idea of how it works. It is known to consist of two subunits, designated large and small, which bind to the messenger RNA separately during the initiation of protein synthesis. The large subunit consists of one copy each of 34 different protein molecules and two molecules of ribosomal RNA (one much larger than the other) and has a total molecular weight of about 1.6 million. The small subunit is made up of 21 proteins and one RNA molecule; its molecular weight is somewhat less than one million. It is the small subunit with which our studies have been concerned.

In principle the ribosome subunits or the entire ribosome could be studied by X-ray diffraction. Such studies, however, require that the specimen be in the form of a crystal; so far it has not been possible to grow crystals of ribosomes or their subunits. Even if the crystal could be grown, the analysis of the X-ray-diffraction pattern generated by such a complex structure would present formidable difficulties. The protein components of the subunits, which should be easier to crystallize, remain candidates for X-ray structural analysis.

In lieu of X-ray crystallography several approaches to a structural analysis have been developed in addition to our own. The first to yield substantial results employs chemical reagents that have two reactant groups to link together adjacent molecules in the structure. By disrupting the ribosomes and determining which molecules have been cross-linked it is possible to compile at least a partial list of proteins that are nearest neighbors. This method has been employed in several laboratories, notably the laboratory of Robert R. Traut at the University of California at Davis and that of Charles G. Kurland at the University of Uppsala in Sweden. Georg Stöffler of



RIBOSOMAL PARTICLES are the irregularly shaped objects in a transmission electron micrograph made by James A. Lake of the New York University School of Medicine. Each particle is a component, called the small subunit, of a ribosome from the bacterium *Escherichia coli.* An analysis of electron micrographs can provide an idea of the overall shape of the subunit, but there is little information in such images about internal structure. On the other hand, the subunits are too large to be investigated conveniently by methods, such as the study of X-ray-diffraction patterns, that build up a map of the structure atom by atom. Neutron-scattering studies provide information at an intermediate scale, and they have now revealed several features in the structure of the small subunit of the *E. coli* ribosome.



COMPONENT MOLECULES of the *E. coli* ribosome include 55 proteins and three molecules of RNA. These molecules are organized to form two subunits; so far only the small subunit has been analyzed by neutron scattering. The small subunit consists of 21 proteins, with molecular weights ranging from 10,700 to 65,000, and a single molecule of RNA with a molecular weight of 560,000. The object of the analysis is to determine how these molecules are arranged in space, and for now only the proteins are being considered. The proteins and the subunits are shown as spheres whose volumes are proportional to their molecular weights, but many of the proteins are not spherical.

the Max-Planck Institute for Molecular Genetics in Berlin and James A. Lake of the New York University School of Medicine have devised an immunological technique. Antibodies specific for particular ribosomal proteins are bound to the ribosomes, forming complexes that can be examined with the electron microscope. The position of the antibody can be determined, and it indicates the position of the corresponding protein in the structure. In another kind of investigation, pursued by Charles R. Cantor and his colleagues at Columbia University, two selected proteins are labeled with fluorescent dye molecules sensitive to different wavelengths. One of the dyes is then stimulated by illuminating it with radiation of the appropriate wavelength; the efficiency with which its energy of excitation is transferred to the other dye is a measure of the distance between the two proteins. Still other studies have involved measuring the scattering of X rays under conditions that do not require a crystallized specimen. Each of these methods can contribute to knowledge of the ribosome's structure.

Our method is a purely physical one; it does not depend directly on the biological or chemical properties of the ribosome. Physical experiments are usually straightforward in design, although, as will be seen, they are not always simple in execution. In our experiment a solution of ribosome subunits is placed in a beam of neutrons. The neutrons interact with the specimen, and the direction of flight of some of them is changed as a result. Those changes in direction represent information about the structure of the specimen; the challenge is to detect the changes and to interpret their information.

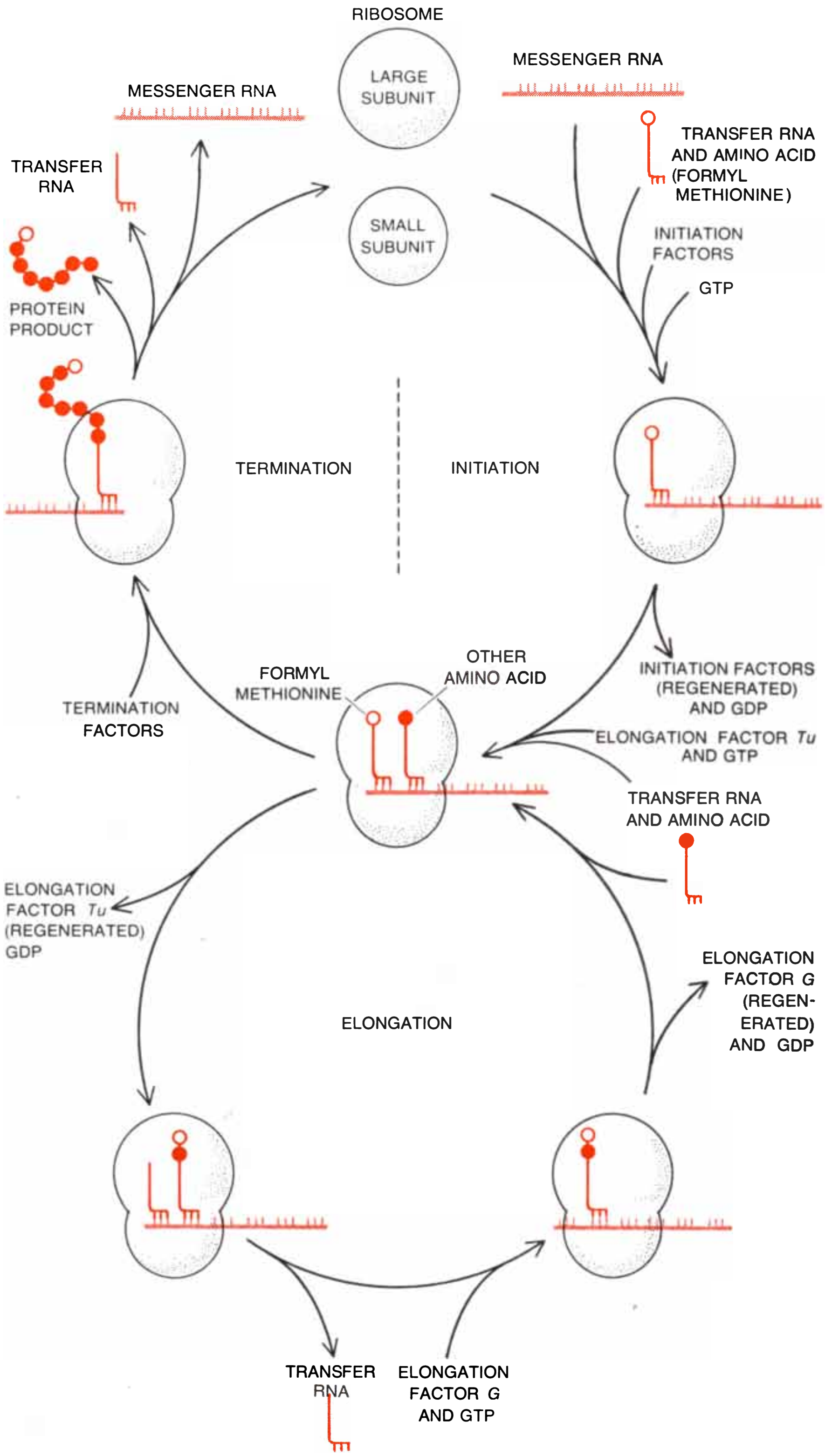
Virtually all investigations of biological structure have employed visible light, electrons or X rays. For light and for beams of electrons it is possible to construct lenses and therefore to form an image of the structure being inspected. X rays, on the other hand, cannot be focused; it is not possible to build an X-ray microscope. Instead of an image X rays yield only a diffraction pattern: a record of the directions and intensities of the waves deflected by their interaction with the specimen. Like X rays, neutrons cannot be focused, and only their scattering angles and intensities can be recorded.

Neutrons differ in an important way from all other kinds of radiation employed in biological investigations. Light, electrons and X rays interact almost exclusively with the electrons in a molecule. Neutrons are unaffected by the electrons but interact with the atomic nuclei. It follows that neutrons can be scattered differently by atoms that represent different isotopes of the same element. Such atoms have the same number of electrons and are therefore virtually indistinguishable to an electron beam or to electromagnetic radiation; in their nuclei they have the same number of protons but different numbers of neutrons. Among the isotopes of a given element there are only negligible differences in chemical properties, but there can be large differences in the interactions of the nuclei with neutrons of the appropriate energy; the isotopes are said to have different neutron-scattering lengths.

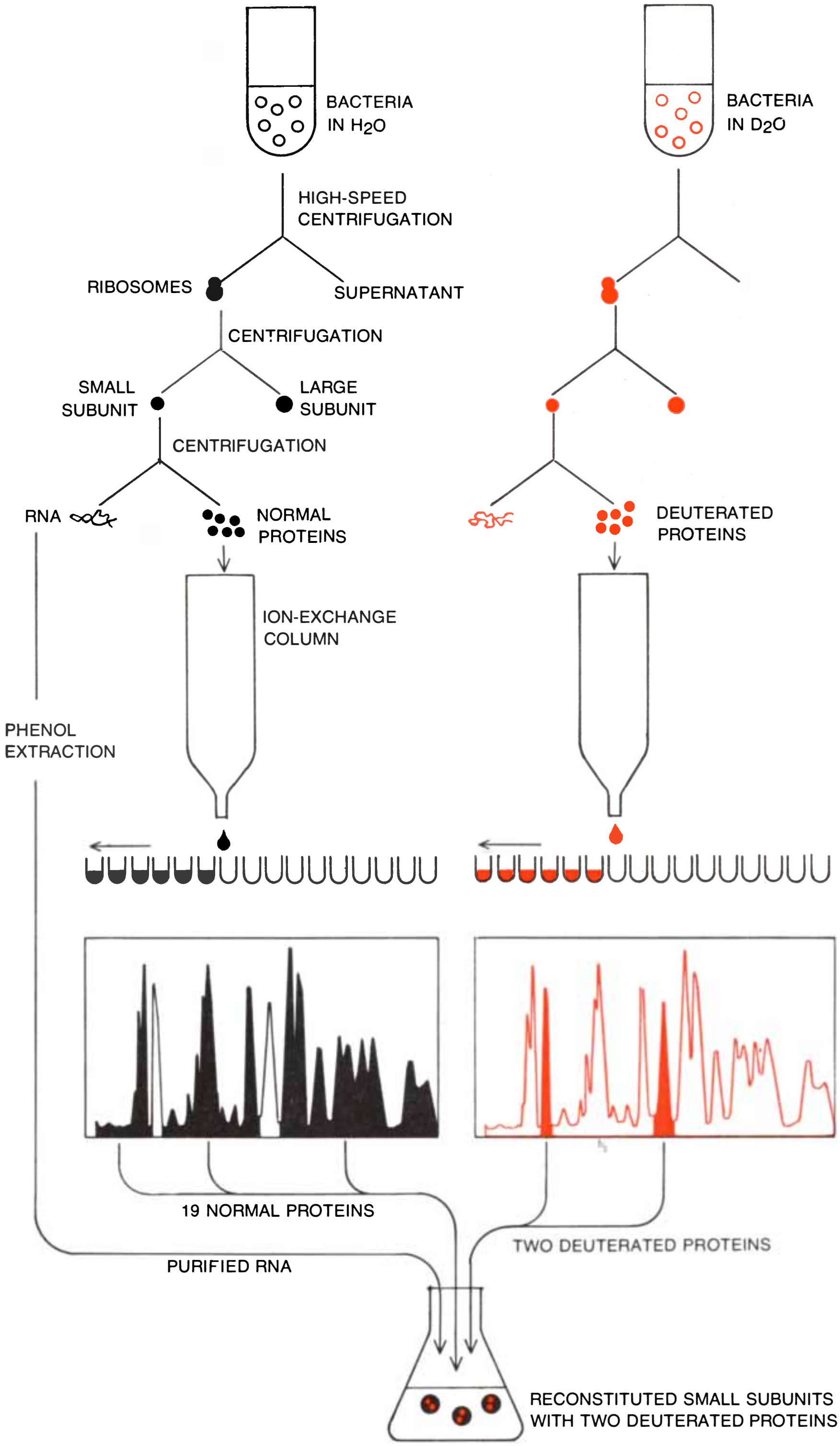
One of the largest differences in neutron-scattering length is the one between hydrogen and deuterium, two isotopes whose nuclei contain respectively a solitary proton and a proton and a neutron. Since biological molecules contain large numbers of hydrogen atoms, there is a large difference in the neutron-scattering behavior of a protein that contains only hydrogen and one in which deuterium has been substituted for most of the hydrogen atoms. In their chemical properties, on the other hand, two such molecules differ very little. Deuterium therefore provides a convenient stain for biological neutron-scattering experiments. By replacing all the hydrogen atoms in one protein molecule with deuterium atoms the protein becomes conspicuous to the neutron beam, although its biochemical properties are little altered.

In practice we prepare ribosome small subunits in which two of the 21 proteins are heavily deuterated. The subunits, in solution, are then placed in the neutron beam and the number of neutrons deflected at various angles is counted. The most significant information derived from this procedure is simply a measurement of the distance between the two deuterated proteins.

In the 1920's Louis de Broglie suggested that matter and in particular atoms and their constituents could be considered as consisting either of particles or of waves. The choice of which convention is adopted is often a matter of convenience. Our experiment is one where it is far more convenient to treat neutrons as waves than as particles.



BIOLOGICAL FUNCTION of the ribosome is to mediate the translation of the genetic code into the sequence of amino acids in a protein. The code itself is represented by the sequence of nucleotide bases in a molecule of messenger RNA. At initiation the ribosome recognizes the starting point in a segment of messenger RNA and binds a molecule of transfer RNA bearing a single amino acid. In all bacterial proteins this first amino acid is formyl methionine. In elongation a second amino acid is linked to the first one. The ribosome then shifts its position on the messenger-RNA molecule and the elongation cycle is repeated. When the end of the genetic message is reached, the chain of amino acids folds spontaneously to form a protein. Subsequently the ribosome splits into its two subunits, which rejoin before a new segment of messenger RNA is translated. Protein synthesis is facilitated by a number of catalytic proteins (initiation, elongation and termination factors) and by guanosine triphosphate (GTP), a small molecule that releases energy when it is converted into guanosine diphosphate (GDP).



PREPARATION OF RIBOSOME SMALL SUBUNITS for structural analysis by neutron scattering involves "staining" two of the 21 component proteins with deuterium, the isotope of hydrogen with an atomic weight of 2. Bacteria are grown both in normal water and in deuterium oxide, or heavy water (D_2O). In the proteins derived from bacteria grown in heavy water virtually all the positions normally occupied by hydrogen are occupied by deuterium. The ribosomes are removed from the bacteria, the subunits of the ribosomes are separated and the ribosomal RNA is extracted from the small subunits, leaving a set of 21 proteins. These proteins, from both the normal and the deuterated ribosomes, are then separated and purified by the technique called ion-exchange chromatography. From the set of normal proteins two are discarded; the corresponding two are selected from the set of deuterated proteins. The 19 normal proteins and the two deuterated proteins are then mixed, along with normal ribosomal RNA, and small subunits spontaneously reassemble. The subunits contain two proteins stained with deuterium, and these differ markedly from normal proteins in their ability to scatter neutrons.

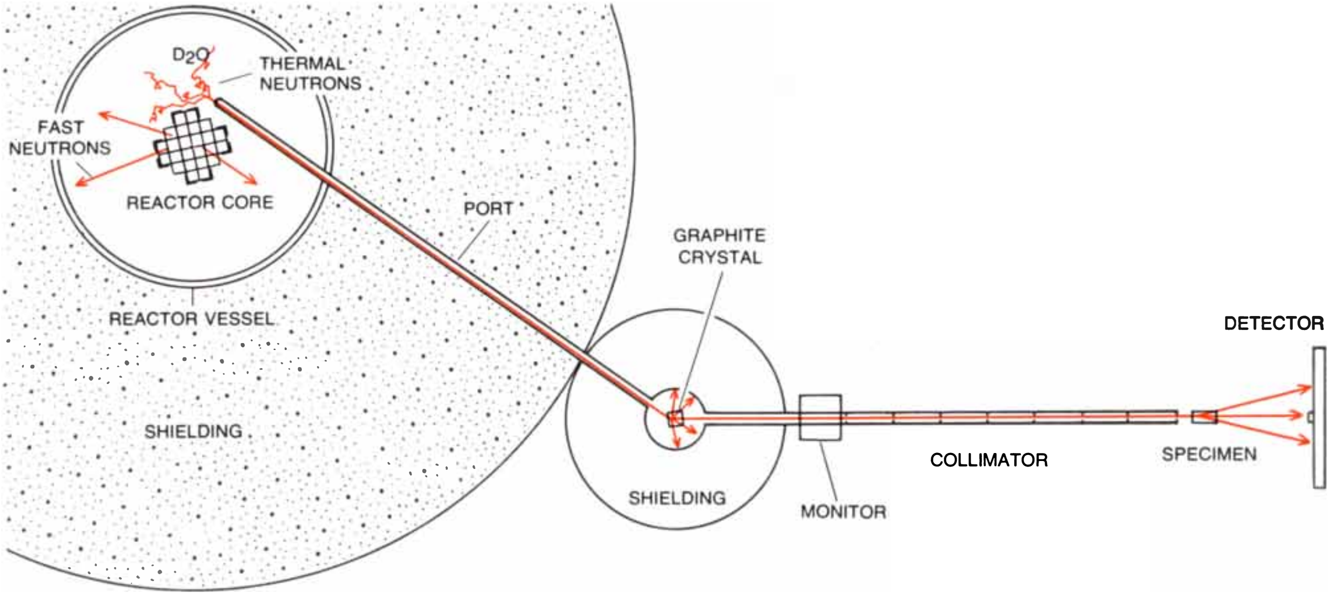
That is because the phenomena we are observing are the diffraction and interference of the neutrons, and those effects are generally associated with wavelike radiation.

The neutrons employed in our experiment have a wavelength, determined by their energy and hence by their velocity, of a few angstroms. When these neutron waves pass near the deuterated proteins, they are diffracted, or deflected; moreover, because each ribosome subunit contains two deuterated proteins the waves diffracted by each pair subsequently interfere.

When light waves emitted by a single source pass through a pair of parallel slits in an opaque screen, the waves generate an interference pattern that consists of several fringes, or parallel stripes of dark areas and illuminated areas. The effect observed in our neutron-scattering experiments is analogous. The two deuterated proteins are equivalent to the two slits, and the pattern generated consists of concentric rings in which the flux of neutrons is alternately greater than and less than the background level. In the optical experiment the distance between two fringes, for any given wavelength of light, is determined by the distance between the slits. In the neutron-scattering experiment the distance between the rings is related to the distance between the deuterated proteins.

There are differences, however, between our experiment and the analogous optical one. In our experiment there is not one pair of slits but many pairs, corresponding to the large number of ribosomes in the sample, and those pairs are randomly oriented with respect to the neutron beam because the ribosomes are in solution. The theory underlying the interpretation of such a diffraction pattern was worked out in 1915 by Peter J. W. Debye. Debye was concerned with describing the scattering of X rays by a diatomic gas, but the scattering of neutrons in our experiment is quite similar: in both cases waves are diffracted by randomly oriented objects that have two scattering centers separated by a fixed distance.

Debye's theory predicts that a graph of the intensity of the scattered radiation with distance from the beam axis will include a contribution from a sinusoidally oscillating curve that is damped at higher angles of deflection. That is, interference fringes will be observed, they will be prominent at small angles and they will fade out at large angles. What is most important for our purposes, the distance between the interference rings is related, inversely, to the distance between the scattering centers. If the scattering centers—in our case the deuterated proteins—are close together, the interference curve has slow oscillations,

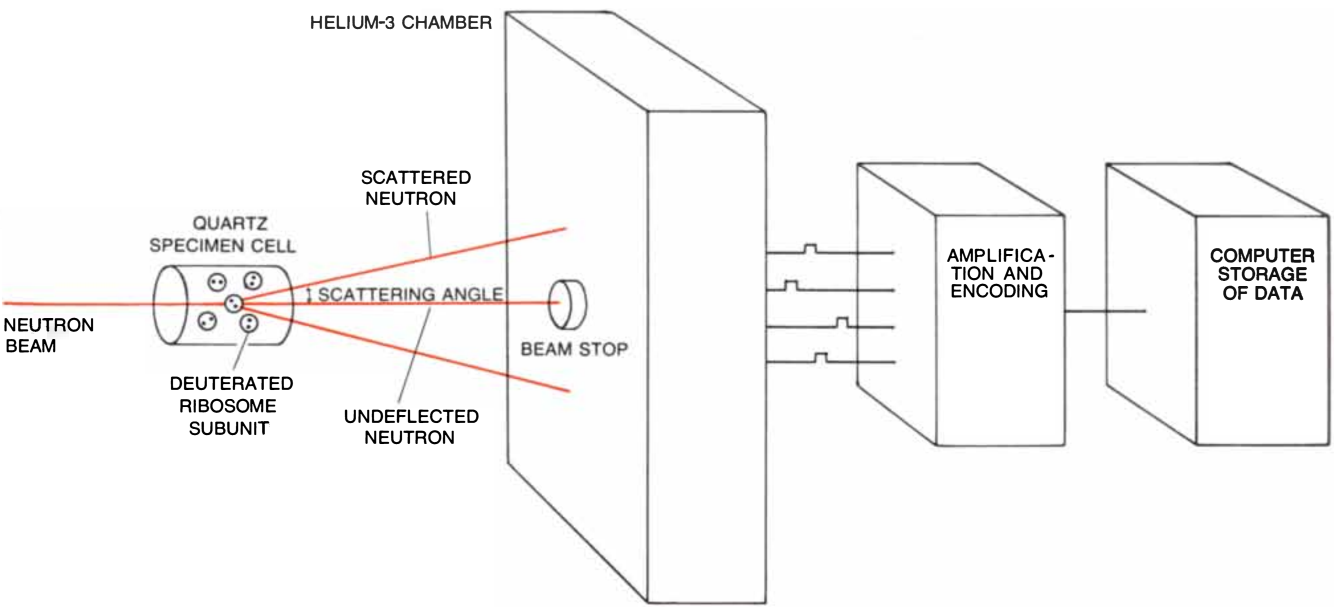


NEUTRON BEAM employed in studies of the ribosome is generated by the High Flux Beam Reactor at the Brookhaven National Laboratory. The neutrons are emitted by fissioning nuclei of uranium 235 in the reactor core. They are slowed by multiple collisions with nuclei in the blanket of heavy water that surrounds the core and pass out of the reactor vessel through a port. The distribution of energies among these slowed neutrons resembles that of a gas in thermal equilibrium, and they are called thermal neutrons. The beam of thermal neutrons is spread out according to their energies by reflection from the planes of atoms in a crystal of graphite, and a beam in which all the neutrons have the same energy is selected by a collimator. Finally the collimated beam strikes the specimen and the neutrons strike the detector.

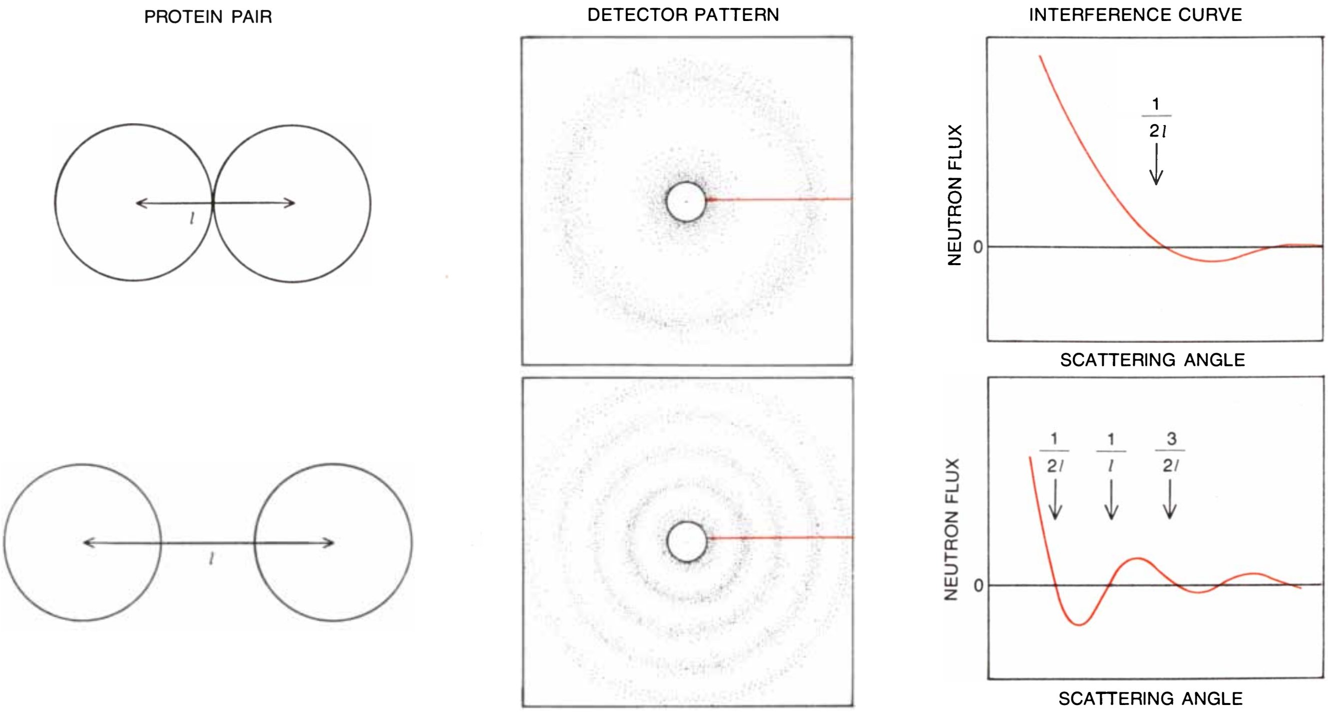
or widely spaced peaks and troughs; if the scattering centers are more distant, the interference curve oscillates more rapidly, or in other words the fringes are more closely spaced.

Preparing ribosome small subunits in which two chosen proteins contain high concentrations of deuterium is not a simple procedure. It might not be possible at all if it were not for an extraordinary property of the subunit itself. In 1968 Peter Traub and Masayasu Nomura of the University of Wisconsin discovered that under certain specified conditions the small subunit will spontaneously assemble itself from its component parts. Therefore if we could prepare and isolate the necessary component molecules, there was a good chance we could build the complete subunit to our specifications. It was because a procedure for the reassembly was known that we decided to work with the small subunit.

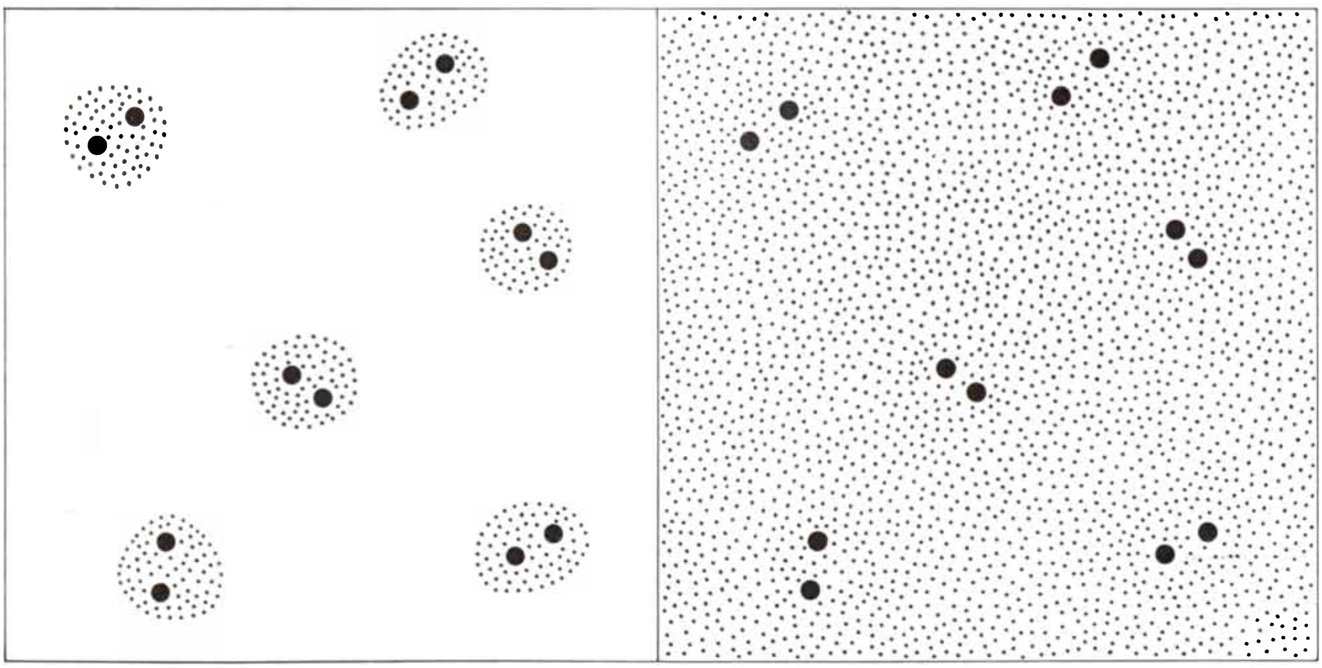
The ribosomes we have employed are those of the bacterium *Escherichia coli,* and our first task was to grow large quantities (kilograms) of the bacteria containing high concentrations of deu-



NEUTRON DETECTOR records the intensity of the neutron flux at each of more than 8,000 positions on its surface, and thereby measures the angular distribution of the neutrons scattered by the deuterated ribosome subunits. The detector is a flat chamber filled with helium 3, an isotope of helium that emits a proton when it is struck by a neutron. The electric charge of the recoiling proton induces a momentary ionization in its immediate vicinity, and the ionization is registered by a system of electrodes distributed throughout the plane of the detector. The signals detected by the electrodes are amplified, encoded to denote the position of the incident neutron and recorded on magnetic tape for later processing with the aid of a computer. The beam of undeflected neutrons is intercepted by the beam stop.



INTERFERENCE PATTERN is generated by neutrons scattered during their passage through ribosome subunits containing two deuterated proteins. The basis of the neutron-scattering studies of the ribosome is the analysis of such patterns to determine the center-to-center distance between a pair of proteins. The interference pattern appears as a set of concentric rings in which the neutron flux is alternately higher and lower than the background level; it can be represented by a graph of the neutron flux as a function of the angular deflection of the beam. In the simplest case, where the two proteins are considered to be perfect spheres, the graph is a damped sinusoidal curve. The distance between successive interference rings is inversely related to the distance between the proteins. If the proteins are regarded as spheres in contact, the rings are widely spaced (*top*); increasing the separation of the proteins crowds the interference rings closer together (*bottom*). In practice the distance is measured from the center of the pattern to the edge of the first ring, that is, to where the interference curve first crosses zero. The ripple in the interference curve has a small amplitude, a few percent of the total neutron flux.



CONTRAST-MATCHING is one technique employed to enhance the interference signal in the neutron-scattering data. The interference results from the effective contrast (for thermal neutrons) between the deuterated proteins (*heavy dots*) and the normal proteins (*lighter dots*) in each ribosome subunit. If the subunits are suspended in pure water, however, the neutron beam also detects contrast between the subunit as a whole and the medium, and a spurious component is introduced into the interference signal. This source of noise can be eliminated by adjusting the neutron-scattering properties of the medium so that they match those of the subunits. Thus at the left the most conspicuous objects are the aggregations of dots, each representing a ribosome subunit, whereas at the right the pairs of heavy dots, each representing a pair of deuterated proteins, are more conspicuous, and the subunits themselves have virtually disappeared.

terium. We found that a satisfactory procedure was to grow the bacteria in deuterium oxide, or heavy water, but to nourish them with normal, undeuterated sugars and mineral salts. Proteins from these bacteria contained deuterium at more than 95 percent of the sites normally occupied by hydrogen. The fact that the bacteria grew well in the heavy water suggested that the substitution of deuterium for hydrogen caused no more than a slight perturbation in the functioning of the cells, and therefore reassured us that the substitution would not alter the structure of the ribosomes. Indeed, when we compared the ability of normal ribosomes and deuterated ribosomes to mediate protein synthesis in a cell-free system, we found they were indistinguishable.

When enough normal and deuterated cells had been prepared, we extracted the ribosomes from each and isolated the small subunits. The subunits were then chemically disassembled, and the RNA was isolated and purified separately. The 21 proteins of the small subunit were separated from one another by the technique of ion-exchange chromatography. In this procedure the proteins are adsorbed on a substrate of modified cellulose and are distinguished by the differing rates at which they are washed through the substrate by a salt solution. Under the proper conditions all 21 proteins can be isolated. The normal and the deuterated proteins separate in exactly the same way, again suggesting that any disturbance caused by the deuteration is not profound.

The proteins purified in this way are designated *S*1 through *S*21. (The *S* is for "small subunit"; the proteins of the large subunit are labeled *L*1 through *L*34.) Each of the 21 can be identified by its relative rate of migration through the ion-exchange column and by other characteristics as well. For any given experiment two proteins are selected; for example, in one of our experiments we measured the separation between protein *S*5 and *S*8. These two fractions were removed from the set of 21 normal, hydrogenated proteins, and the remaining 19 normal proteins were mixed in solution. The *S*5 and *S*8 fractions from the set of deuterated proteins were then added to the solution, and finally the purified RNA from the normal small subunits was also added. The solvents employed in the separation and purification were then replaced by another buffer solution and the mixture was stirred under mild heat for two hours. Under these conditions the 22 component molecules fold and assemble themselves into reconstituted ribosome subunits. The process is remarkable, and even though we have seen it many times it has not ceased to amaze us. The only information stated explicitly in the genetic code is that which specifies the linear sequence of amino acids in a protein. It is generally accepted that this sequence also controls the folding of the amino acid chain into its proper three-dimensional conformation. The spontaneous reconstitution of ribosome subunits implies that the genetic code also contains the information needed to assemble a large complex of protein molecules and RNA.

The reassembled subunits contain the two deuterated proteins whose chemical identity is known and whose positions are to be determined. The remainder of the structure contains only hydrogen. Tests of the reconstituted subunits have shown that they are active in protein synthesis, that they are similar to normal ribosomes in size and shape, and that the deuterated proteins are actually incorporated into the structure.

One more crucial adjustment must be made to the sample before the neutron-scattering curve is measured. If the deuterated subunits were suspended in a medium of pure water, the neutron beam would detect the contrast between the proteins containing deuterium and those containing hydrogen, but it would also detect substantial contrast between the subunits as a whole and the solvent in which they are suspended. The interference signal is of small amplitude and under the best conditions is difficult to extract; it would be completely obscured if the scattering contribution of the whole subunits were not eliminated. That problem is solved by a technique known as contrast-matching. By adding heavy water to the solvent one can adjust its scattering characteristics so that on the average the number of neutrons scattered per unit volume is the same for the ribosomes and for the surrounding buffer. In this way the contrast between the solution and the ribosome as a whole is reduced to virtually zero, and if the ribosome were a completely uniform object, it would vanish: the neutron beam would not detect its presence. The fact is, however, that the ribosome is not uniform; it contains two heavily deuterated proteins, which now become the most prominent objects in the solution.

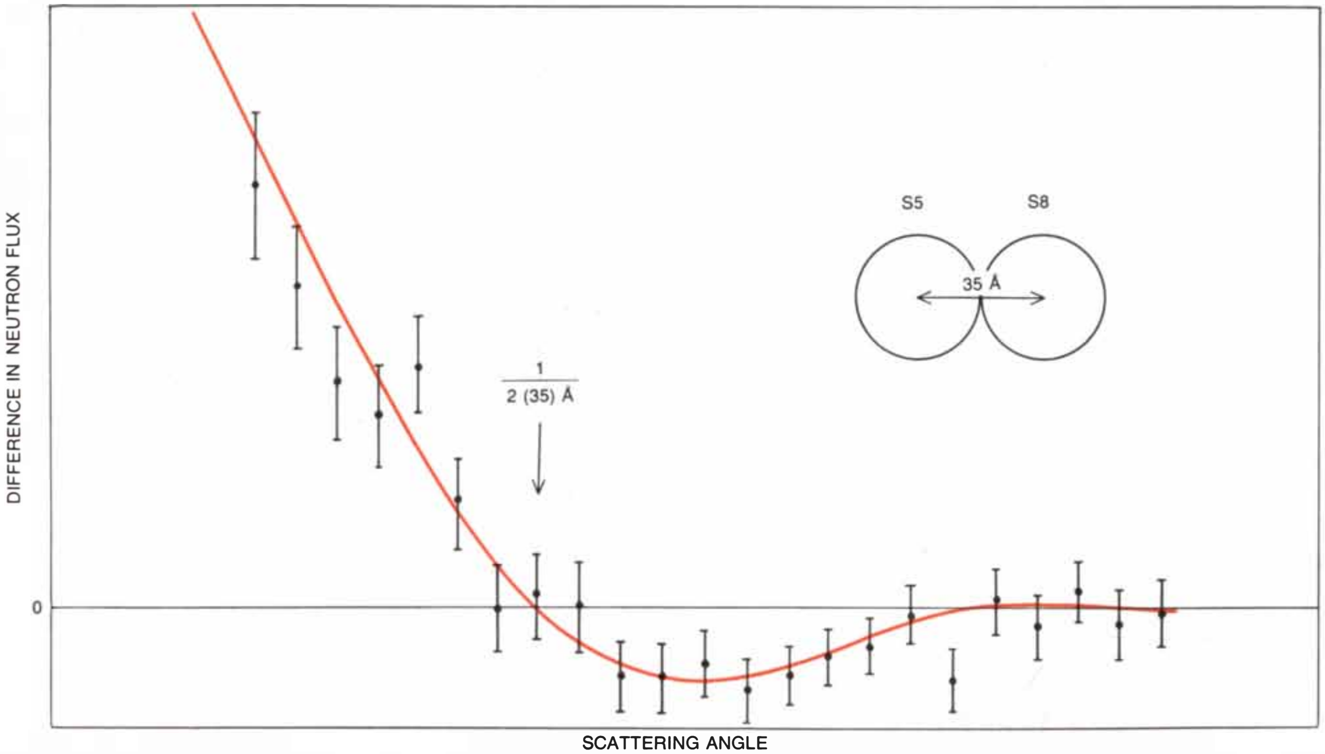
Our measurements of neutron scattering have been made with the High Flux Beam Reactor of the Brookhaven National Laboratory. The neutrons released by the fissioning of uranium nuclei in the core of this reactor are slowed by multiple collisions with deuterium nuclei in the cooling blanket of heavy water that surrounds the core. For our experiments the neutrons are slowed to thermal velocities, that is, they are slowed until the distribution of energies among the neutrons is like that of the particles in a gas in thermal equilibrium. Neutrons with these relatively low energies are useful in a variety of studies at atomic resolution, including those of biological materials.

The thermal neutrons leave the reactor through a port in the reactor vessel and are conducted into the experimental area through a long tube set into the radiation shielding. As the neutrons leave the reactor they have a broad distribution of energies, or wavelengths, but our measurements require a beam of neutrons all of which have the same wavelength. Such a beam is selected by the effect called Bragg reflection, after Sir William Bragg, who developed many of the techniques now employed in studies of X-ray diffraction and neutron diffraction. The neutrons are directed into a crystal of graphite, where they are reflected by the numerous internal planes of atoms. The angle of reflection depends on the neutron wavelength, and hence the neutron beam is spread out into a spectrum, from which a monochromatic beam made up of neutrons with a particular wavelength is selected. The neutron spectrum is directly analogous to the optical spectrum, and the planes of atoms responsible for Bragg reflection are analogous to the rulings on an optical diffraction grating.

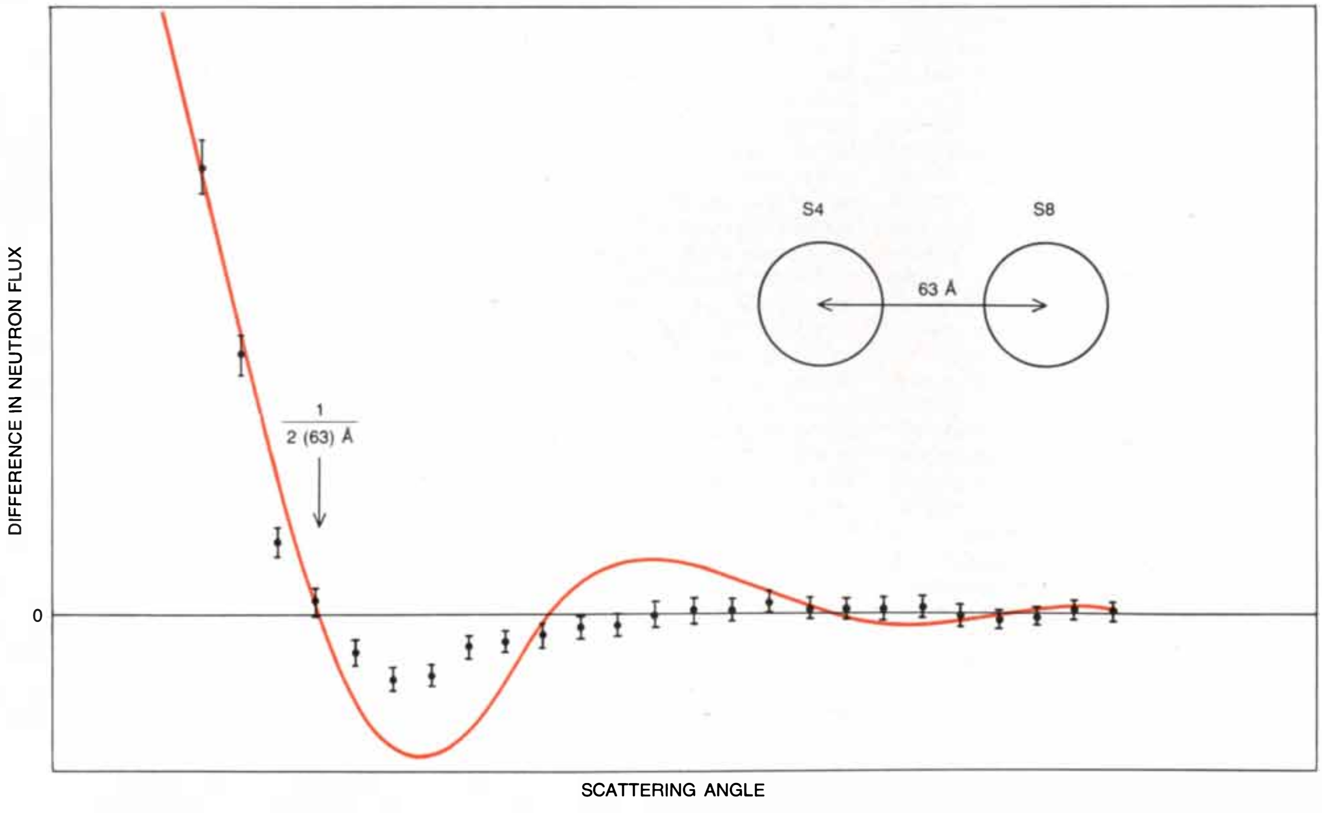
For our experiments the orientation of the graphite crystal is adjusted so that it selects neutrons with a wavelength of 2.37 angstroms. The selected beam is directed into a tube about two meters long, down the length of which there are several narrow apertures. The beam is thereby collimated and directed into the quartz cell that holds the sample of ribosome subunits.

Without the expert and enthusiastic support of the staff at Brookhaven, success in these experiments would not have been possible. In particular Benno P. Schoenborn, who was the first investigator to employ neutron diffraction for the elucidation of biological structures, has contributed a great deal to the development of needed techniques. One of his essential contributions was the development of a detector for neutrons capable of simultaneously measuring the neutron flux at many points on a plane, that is, a position-sensitive detector with good spatial resolution.

Because neutrons do not interact with electrons, they cannot be detected in the same ways that electromagnetic radiation or an electron beam is detected. Photographic film, for example, is not efficiently exposed by neutrons. The detector we have employed at Brookhaven owes its sensitivity to the special properties of the gas helium 3, the isotope of helium with two protons and one neutron. When a nucleus of helium 3 is struck by a neutron, the products of the collision are a nucleus of tritium (the isotope of hydrogen with one proton and two neutrons) and a proton; in effect



DISTANCE MEASUREMENT between two components of the small subunit, the proteins designated *S5* and *S8*, gives a center-to-center separation of 35 angstroms. The measured interference curve is given by the black dots and the vertical bars representing the possible experimental error. The solid curve is a theoretical one calculated on the assumption that the two proteins are both spherical, that they are both 35 angstroms in diameter and that they are touching each other. Since the theoretical curve fits the measured data quite well these assumptions would seem to be justified. The distance measurement is derived from the point where the curve first crosses zero.



PROTEINS S4 AND S8 were paired for a second distance measurement. Again the dots and error bars are measured data and the solid curve is based on a theoretical calculation. In this case, however, the theoretical curve is not in good agreement with the data; in particular the measured curve seems to be more strongly damped than the calculated one. The discrepancy probably indicates that one of the proteins is not well described by a sphere, and since *S8* was found to be compact in the measurement of *S5* and *S8*, it seems likely that *S4* has an extended shape. The deviation attributed to the shape of *S4* does not substantially alter the point at which the curve first crosses zero, so that the center-to-center distance between the two proteins can still be measured; the distance is estimated to be 63 angstroms.

the nucleus of helium 3 absorbs a neutron and emits a proton. The proton, of course, carries an electric charge, and it momentarily ionizes some of the helium atoms nearby. The ionization can be detected by conventional means.

The detector is a large, flat chamber filled with helium 3. Electrodes are arranged within the chamber so that the ionization of any small region of the gas is detected as a pulse of electric current. From the nature of the signal produced the position of the ionization event can be determined. The pulses are counted, encoded to denote the position of the incident neutron and recorded on magnetic tape.

The amplitude of the interference signal generated by the deuterated proteins is only a few percent of the total neutron flux incident on the detector. As a result the stability and reliability of the experimental apparatus are critical. With the largest samples of ribosomes that we can readily produce, collecting enough data to define one interference curve calls for about three days of continuous measurement.

Once the data have been collected, they are further processed with the aid of a computer. First the measurements at all locations in the detector are summed in rings that have as their center the position of the undeflected beam. Each ring therefore represents a measurement of the neutron flux at a particular scattering angle. The series of rings gives a graph of neutron flux v. scattering angle.

The object of the measurement, of course, is not the total neutron flux but the rippling interference curve. That curve can be isolated by subtracting from the total measured flux all components that do not arise from scattering by the paired proteins. Measurements are made of two solutions. In one of them half of the ribosomes contain the selected pair of deuterated proteins (such as *S*5 and *S*8); the rest of the ribosomes in this solution contain only hydrogenated proteins. In the second solution half of the ribosomes have one of the pair of deuterated proteins (such as *S*5) and the other half of the ribosomes have the other deuterated protein (*S*8). If the concentration of the two solutions is equal, then both contain the same components: the same number of ribosomes and the same number of deuterated proteins of each species. They differ only in the arrangement of these components. If the scattering-intensity curve of one solution is subtracted from that of the other, the difference is the contribution of the geometric arrangement of the proteins in the doubly deuterated ribosomes. This is the interference curve, the equivalent of the fringes observed in the double-slit optical experiment.

In 1974, when we obtained our first measurements of the distance separating a pair of proteins, we were dismayed to find that the rippling interference curve we had expected had only a single fluctuation; in other words, the neutron-flux intensity went from a high positive value to a negative value and then returned to zero, with no further oscillations. On returning to our theoretical model of the curve we were somewhat reassured: the simplest model assumed that the proteins are spherical or at least compact, and that is not necessarily the case. If the proteins had a more extended shape, a strong damping of the later peaks in the curve could result. The distance from the origin to the point where the curve first crosses zero, however, is little affected by protein shape, so that the distance between the molecules could still be measured. To make certain that the signal we had detected was a real one, we repeated the experiment, following all the steps in the procedure but employing only hydrogenated proteins; no interference signal was detected. Other control procedures confirmed that the signal we had observed is generated by the deuterated proteins and that the ribosome subunits survive under the experimental conditions. Moreover, we were able to repeat a measurement with reasonably consistent results.

In a first approximation the distance between the proteins can be obtained by a simple procedure. We have seen that the spacing between the interference fringes is inversely related to the spacing between the proteins, and this relation is readily made a quantitative one. First the angle between the undeflected beam and the edge of the innermost interference fringe (where the interference curve first crosses zero) is measured. The distance between the proteins is then equal to the neutron wavelength divided by twice the sine of that angle.

Applying this technique to the interference curve for the proteins *S*5 and *S*8 yields a center-to-center distance of about 35 angstroms. For this pair a very simple model for the shape of the proteins is entirely adequate. By assuming that both proteins are spheres with volumes appropriate for their molecular weights and by assuming that they are in contact an interference curve can be calculated that fits the data well. We therefore conclude that *S*5 and *S*8 are compact when viewed along the line connecting their centers of mass, and that they are about 35 angstroms apart. This is about as close as they could be packed together, given their molecular weights.

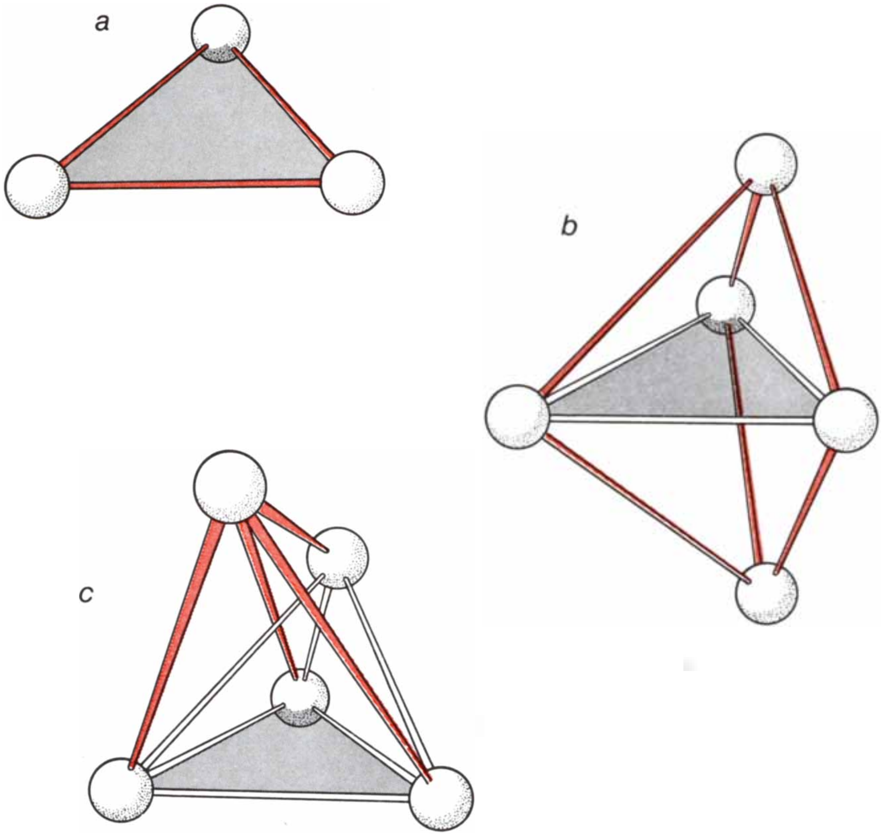
Another protein pair we have measured is *S*4 and *S*8. Here the location of the first zero intercept gives an interprotein distance of 63 angstroms, but a calculated curve for two spherical proteins of the appropriate size does not fit the data well. It therefore appears that the effects of protein shape have altered the form of the curve. The mathematical model suggests that an interference curve of the right kind would be generated if one protein were compact but the other were elongated. Since *S*8 appeared to be compact in the *S*5–*S*8 measurement, it is possible that *S*4 is elongated.

Given a few measurements of distances between pairs of proteins, it is possible to begin the construction of a map of the ribosome small subunit. The map begins with three center-to-center measurements relating three protein molecules; these three necessarily define a triangle and a plane in the subunit. A fourth protein can be located by measuring the distance between it and the original three; unless it happens to lie in the same plane, it defines a vertex of a tetrahedron.

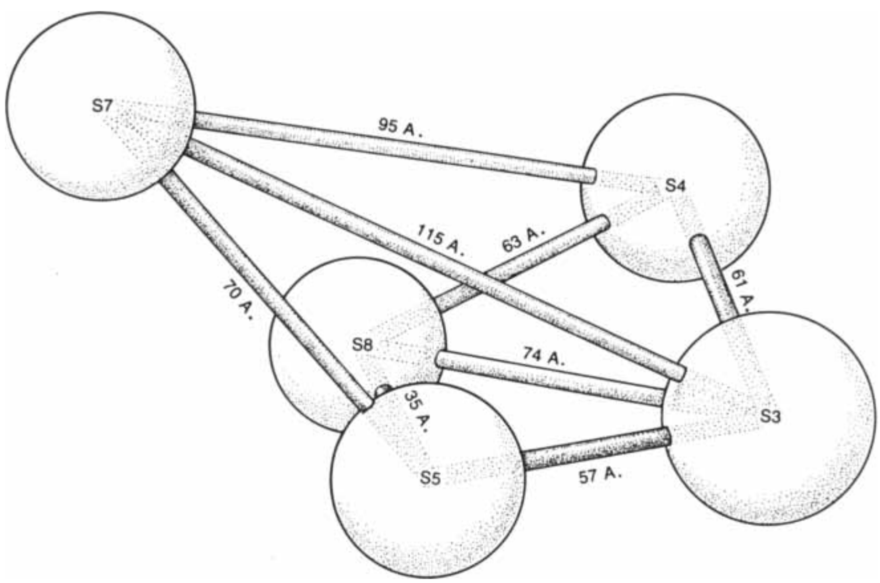
The placement of the fourth protein in the map entails a certain irreducible ambiguity. The distance measurements from the first three proteins actually define two possible positions for the fourth, one on each side of the plane, and the neutron-scattering data contain no information about which position is the correct one. If a choice is made arbitrarily, then all subsequent proteins added to the map can be placed without ambiguity, but the fact remains that the map could represent either the actual structure or its mirror image. This ambiguity concerning the handedness of the structure cannot be resolved by any system of measurements between pairs of proteins, and so the ambiguity will remain even when the map is complete. If we should reach the point, however, where the handedness of the ribosome is the only uncertainty in our knowledge of its structure, then we shall have done well indeed.

Once the positions of four proteins, defining a tetrahedron, have been established, additional proteins can be added to the map by making a set of four measurements to the vertexes of this tetrahedron or to any other four molecules whose positions are completely defined. It is not necessary to measure all the interprotein distances to map the structure completely; for the 21 proteins of the small subunit there are 210 protein pairs, but 74 measurements suffice to define the structure. Additional measurements, of course, would increase confidence in the reliability of the map.

Most of the measurements we have obtained so far we made in collaboration with Jerome Langer, a postgraduate student at Yale University. The map of the small subunit we have assembled includes five proteins: *S*3, *S*4, *S*5, *S*7 and *S*8. Four triangles have been defined among them by eight measurements, but the spatial relations of the triangles themselves are not yet definitely known. One additional but tentative measurement suggests that the triangles form a relatively compact, folded aggregation, but even if that measurement proves to



THREE-DIMENSIONAL STRUCTURE of the small subunit is obtained by triangulation from measurements of the distances between pairs of proteins. The first three measurements define a triangle and, necessarily, a plane through the subunit (*a*). Three more measurements (*b*) give the position of a fourth protein and define a tetrahedron. In placing this fourth protein, however, there is an ambiguity that cannot be resolved by any system of pairwise measurements: the protein could occupy a position on either side of the plane defined by the original triangle. If an arbitrary choice is made, additional proteins can be added to the structure by making measurements to any four proteins whose positions are already defined (*c*). Nevertheless, the handedness of the structure remains indeterminate, that is, the map could represent the actual structure or its mirror image. A map of the subunit requires at least 74 measurements.



TENTATIVE MAP of a portion of the ribosome small subunit includes five proteins, related by eight distance measurements. No complete tetrahedron is included in the map, and as a result the structure shown is not a rigid one; rather, it can flex along several axes, changing the orientation of the proteins without changing the measured distances between them. In addition the measurement between *S*5 and *S*7 is only a tentative one. The proteins are shown as spheres in order to indicate their approximate volumes, although some of them are not spherical.

be accurate, it will not rigidly determine the spatial relations of the five proteins. At least one more measurement is needed to define a tetrahedron.

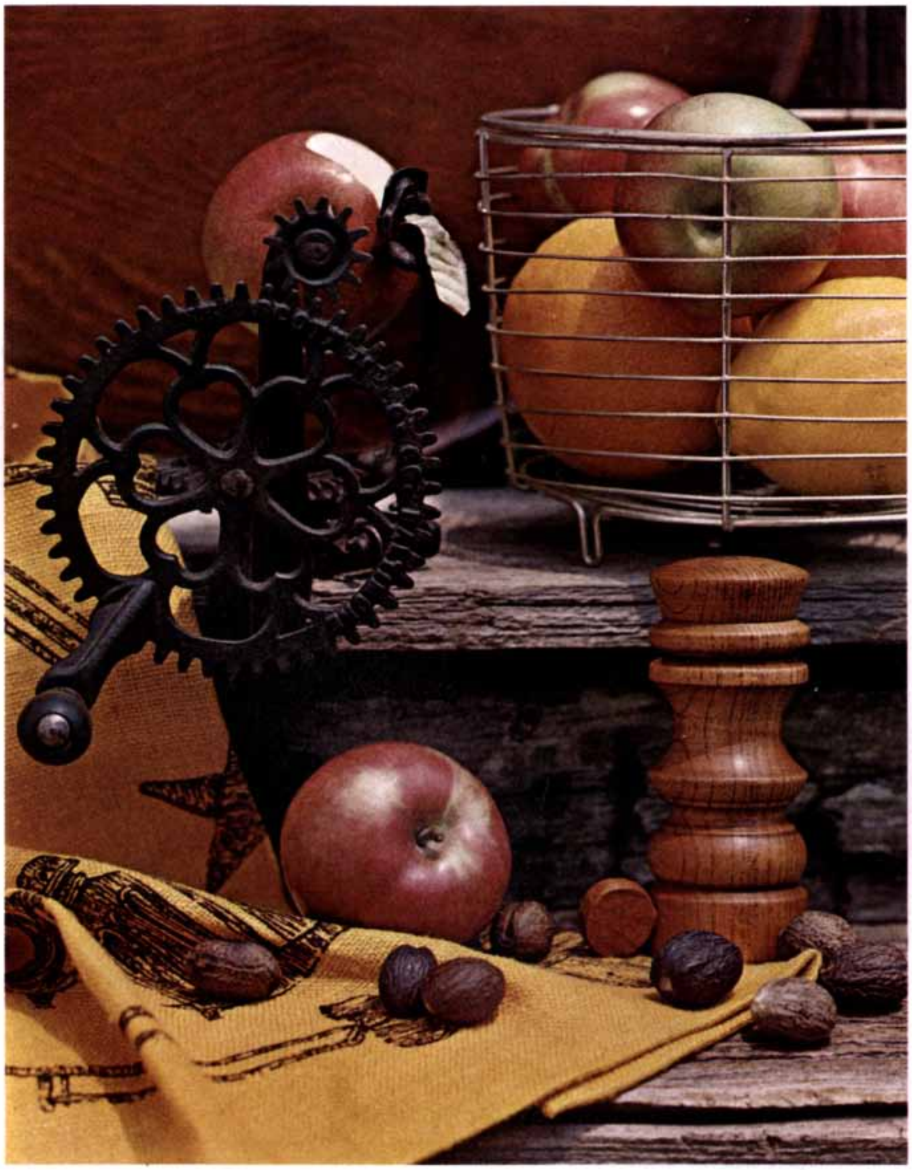
It is clear that, with only eight interprotein distances established and a minimum of 74 required, the structural analysis of the ribosome small subunit has just begun. A year has been spent acquiring the information we now have, and a few more years will be needed to complete the map. Our progress may seem slow, but it is similar to that encountered in many other kinds of scientific undertaking; moreover, because of the importance of the ribosome in the metabolism of the cell the potential dividends in new knowledge are large.

A map showing the relative positions of 21 proteins is not the only product to be expected from the study of neutron-interference curves. As we have suggested, the shape of the curve may provide additional information about the general shape and orientation of each protein. In principle the ribosomal RNA could be included in the neutron-diffraction map as well. The application of the technique to RNA would require a method for cutting the RNA at known points along its length, then substituting deuterated segments in the reassembled molecule.

Several proteins and other molecules, including messenger RNA and transfer RNA, interact with the ribosome during protein synthesis. If any of these interactions are stable, or if they can be halted in some stable intermediate form, the locus of interaction could be determined by relating the molecules to proteins in the map. It might also be possible to measure interprotein distances within the ribosome during various stages of protein synthesis, again assuming that stable intermediate states are formed. The conformation of the ribosome may change during protein synthesis, and the nature of such changes would be revealed by the neutron-diffraction measurements. Information from studies of this kind should bring us much closer to an understanding of one of the focal processes of life: the translation of the genetic code into a gene product.

Finally, the ribosome is not the only biological structure that can profitably be explored by neutron scattering. One particularly interesting assemblage of molecules, which is expected to be studied soon, is DNA-dependent RNA polymerase, which mediates the process of transcription: the copying out of the DNA code into messenger RNA. Other candidates for study by the same methods are chromatin, the complex of molecules, including DNA, that makes up the genetic material of eukaryotic organisms, and various biological membranes. We thus have before us the exciting prospect of understanding all these structures at the molecular level.

How do you like the way our new color slide film reproduces oranges, yellows, reds, wood tones?



Which is a rather meaningless question when you consider that at the moment you are not looking at a projected transparency in a darkened room but at four standard printing inks deposited on a certain grade of paper by a high-speed printing press.

Autumn can be appreciated without the aid of a spectrophotometer. Biogenic pigments differ greatly in molecular structure from photographic dyes and printing inks with which one may try to suggest them. That you can relive a visual experience and consider it a good match is a miracle, for which we are most grateful.

A few months ago when our new instant photography was announced, outsiders said "better color," whatever that may mean to the individual who, to our vast delight, says it.* Though the dyes in color prints made from Kodak instant film are a new departure, differing from the previous technology in being released by the oxidized developer from a physically and chemically bound state instead of being formed by reaction of the oxidized developer with couplers, a projected slide can still come closer than a print to simulating the original perception.

Now, four new Ektachrome films. We hope you will *like* them for your slides. They are all designed for the new, simplified Process E-6.

The New Film

Now at photo supply houses for 35 mm slides and larger transparencies:

KODAK EKTACHROME 64 Professional Film (Daylight)
KODAK EKTACHROME 50 Professional Film (Tungsten)
KODAK EKTACHROME 200 Professional Film (Daylight)
KODAK EKTACHROME 160 Professional Film (Tungsten)

Figures signify average film speed. Actual speed is given on enclosed package directions. That 200 figure is a new high for Kodak color films without "push processing." If you need fine grain and sharpness more than you need speed, use 64 or 50 film. Nevertheless, the long-run trend of constant improvement in speed:graininess ratio continues with all four of these. They represent advances in rendition of orange, yellow, red, and neutral subjects and in being designed for a quicker, more efficient process. Ask about Kodak processing service when buying the film. Before long, another group of films of the same new family will be available with "Professional" omitted from the name. The term warns that for proper color balance, storage before exposure and processing must be at 55°F (13°C) and not beyond the date stamped on the box. (For film stored at 0°F, time almost stands still.) People not in the business of taking pictures generally don't cool their film and often let quite a bit of time go by between the first shot and the processing. They will probably get better results by waiting for the new Ektachrome film not labeled "Professional."

Kodak

*To see what a complex subject "better color" is, ask Dept. 55W, Kodak, Rochester, N.Y. 14650 for "Objectives in Colour Reproduction," by R. W. G. Hunt.

1977 Ford Pinto 3-door Runabout. Exciting new front and rear styling.

A high spirited new Pinto has been born for 1977. With a fresh new look that will still be exciting and distinctive years from now.

Pinto is highly spirited with new options, too. Like our new glass third door, a flip-up/removable open-air roof, and new 4-way driver's seat (it easily adjusts up and down, as well as fore and aft) for more convenience, more comfort.

And for '77, our spirited new Pinto is as nimble and sure-footed as ever. But for

One of our beautiful new interiors. Proof that the practical Pinto can also be luxurious.

Our electro-dip priming process: part of our Improved Corrosion Protection Plan for '77.

even more spirit, there's Pinto's optional 2.8 liter V-6 engine with automatic transmission. Or equip your Pinto with the sports rallye package with a new fun-to-drive suspension, for the sportiest handling Pinto has ever had.

But the beauty of the '77 Pinto doesn't stop with its new look and spirit. Critical areas around all front lights are now made of a new resilient material that's dent, chip and crack resistant under normal use . . . *and absolutely rustproof.*

And to make sure your high spirited Pinto lasts, there's Pinto's built-in durability. Durability that starts with the electro-dip prime coating of Pinto's entire unit-body construction.

Find out how easy it is to corral the spirited '77 Pinto. See your local Ford Dealer soon for a test drive.

Our most economically priced Ford Pinto, the Pinto Pony. Shown here with optional white sidewall tires and deluxe wheel covers.

FORD

When America needs a better idea, Ford puts it on wheels.

SCIENCE AND THE CITIZEN

On Limited Nuclear War

From time to time since the nuclear era began the idea has been advanced among military planners and thinkers that national-defense policy should aim not only at preventing a nuclear war but also at fighting, surviving and winning one. Such a war is generally called "limited," since it is difficult to define survival in an unlimited ("all-out") nuclear war, with casualties on each side in the range of 100 million. The possibilities of limited nuclear war were raised in Herman Kahn's *On Thermonuclear War* in 1961, and the concept was implicit in the brief effort, early in the 1960's, to induce Americans to undertake a serious civil defense effort. It is implicit once again in the counterforce strategy that has been advocated since 1974 by the Department of Defense, and once again it is associated with a new infusion of energy into civil defense.

A counterforce strategy calls for attacks on military objectives rather than on urban-industrial centers and requires in particular the ability to execute "surgical" strikes against hardened concrete missile silos. To that end the Department of Defense is asking for a $1-billion development program for weapons with improved accuracy and yield. The hope is that both sides would limit their attacks to military targets and that, given an effective civil defense system, a counterforce war could be fought with an acceptable level of civilian casualties. To that end, in turn, the Defense Civil Preparedness Agency has been reorienting its planning to give high priority to evacuating civilians from high-risk areas, during the few days of "rising international tension" before the first attack, to presumably safe small-town and rural "host areas."

Opponents of the proposed strategy combining counterforce with civilian defense doubt that a limited war would remain limited. They question, as the independent Center for Defense Information puts it, whether the Russians "have the same incentives as the U.S. to limit nuclear attacks to purely military targets" or the required "sophisticated command and control machinery." They challenge the Defense Department estimates of casualties as being much too low, in part because they see no likelihood of efficient evacuation. And they argue that limited-war planning actually increases the probability of war. For one thing, political leaders may actually believe they can keep a nuclear exchange limited and may therefore initiate one more readily. More particularly, fear of a preemptive counterforce strike by the U.S. could induce Russian leaders to strike first or to "launch on warning"—and vice versa.

The faith of the Defense Department in the possibility of limiting a nuclear war has also been challenged in Congress, and it is apparently not shared by Democratic Presidential candidate Jimmy Carter. In an interview with Hearst Newspapers editors in July he said: "My own belief is that limited nuclear war would be unlikely." The "presumption that massive strategic attacks on population centers would not follow" a limited first strike and retaliation is, he said, "a possibility, but I think a doubtful one." Carter has emphasized the "overwhelming capability" of each side to wreak havoc on the other as the major deterrent, and he believes the "most important strategic element in the entire defense mechanism ... is nuclear-powered submarines. They are almost completely invulnerable to missile attack and their deterrent value is superb." Carter has opposed production of the B-1 bomber and has urged a five-year comprehensive ban on all nuclear tests, and SALT negotiations aimed at controlling qualitative as well as quantitative aspects of the arms race.

Four Colors Do It

For 125 years mathematicians have been baffled by a cartographer's problem: How many colors are needed to make a map in which no two neighboring countries are the same color? (Regions meeting only at a point are not considered neighbors.) No one has ever drawn a map requiring five colors or more, but it was not until this past summer that it was mathematically proved that four colors are enough to color any flat map. To crack the celebrated four-color-map problem Wolfgang Haken and Kenneth Appel of the University of Illinois at Urbana-Champaign employed large computers and a novel mathematical aesthetic.

By the classical standards of mathematics their proof is highly inelegant. An "elegant" proof makes use of powerful theorems to arrive at a result with exemplary efficiency. Haken and Appel invoked few theorems and examined some 10 billion individual cases. "Without the computers," Appel said, "the proof would have been impossible."

To prove the four-color theorem the Illinois mathematicians applied the method of mathematical induction to the number of regions on a map. It was obvious that the theorem is true for a map with the smallest possible number of regions (one). It remained to be proved that if the theorem is true for maps of n regions (where n is any number), it is also true for maps of $n + 1$ regions. Once this was established the four-colorability of maps with any number of regions would be assured and the theorem would be proved.

It had already been shown that certain maps, called reducible configurations, were not incorporated in maps of $n + 1$ regions that required five colors or more. Therefore if Haken and Appel could find a set of reducible configurations that had a member in every flat map with $n + 1$ regions, the theorem would be proved.

This method of proof presented several problems. Existing computer programs for the method called for several hours of computer time to prove that a single configuration was reducible. Since Haken and Appel were looking for a very large set of reducible configurations, the computer time that would be needed appeared to be prohibitive. This requirement was eased by John Koch, then a graduate student at Illinois, who wrote an efficient program for checking reducibility. The major difficulty, however, was applying the "theory of unavoidable sets," that is, developing a set of reducible configurations that would have a member in every flat map and still be of a manageable size.

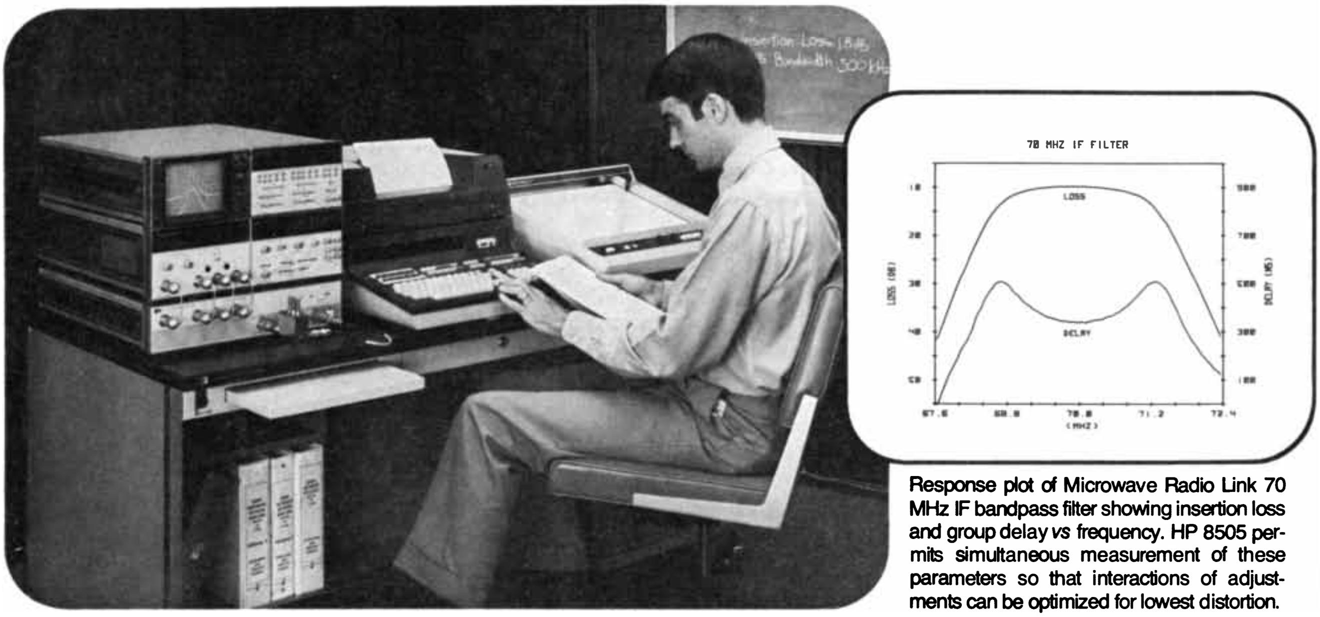
Once the Illinois mathematicians had succeeded in applying the theory of unavoidable sets, they checked more than 10,000 map diagrams in order to select some 2,000 to submit to the computer program. As many as 200,000 possible colorings were checked for each configuration, a process that required 1,200 hours of computer time before a suitable set of reducible configurations was established.

Haken and Appel are not troubled by the unwieldiness of their proof. Indeed, they believe it is intrinsic to problems such as the four-color theorem. Said Appel: "While we are excited to solve a problem as well known as the four-color theorem, we think that the real significance of our work lies in the area of mathematical proof."

The Illinois mathematicians believe a concise, elegant proof of the four-color theorem does not exist. They think they have found the first example of a type of problem that cannot be solved by the usual mathematical techniques. "We believe there must be many true statements in mathematics that can only be proved by such methods," said Haken. "We hope our work on this problem will encourage other mathematicians to investigate these techniques."

Mother Lode

East Africa, from Tanzania northward to Ethiopia, continues to be a gold mine of fossils illuminating the evolu-



Response plot of Microwave Radio Link 70 MHz IF bandpass filter showing insertion loss and group delay *vs* frequency. HP 8505 permits simultaneous measurement of these parameters so that interactions of adjustments can be optimized for lowest distortion.

New network analyzer brings powerful measurement capabilities—including automation—to design and manufacture of high-frequency products.

Measurements in the medium-to-high-frequency range (500 kHz to 1.3 GHz) are significantly expanded, speeded, and simplified with the new HP 8505 Network Analyzer. In combination with the HP Interface Bus and a controller such as the HP 9830 desk-top computer, it is a powerful, extremely accurate automatic system. The measurements it makes let designers produce higher-performance equipment to help cope with the dynamic growth in the communications field.

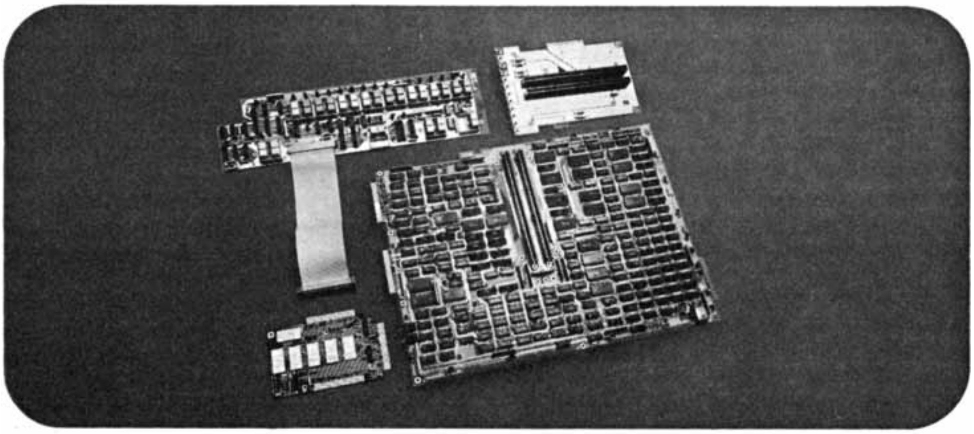
Characterizing circuits to predict their performance is a fundamental activity in both the design and test of communications components and systems. Over the past 10 years there has been a dramatic growth in the use of network analyzers to perform this task. By providing plots of gain or loss, phase shift, and complex impedance *vs* frequency, network analyzers have given electronic engineers practical insights into circuit and component behavior.

The new HP 8505 makes these network measurements with a greater frequency range (500 kHz to 1.3 GHz), greater dynamic range (100 dB), better resolution, and higher accuracy than its predecessors, and with considerably more convenience. It also makes direct measurements of electrical line length, deviations from linear phase, and group delay. Heretofore, these measurements were extremely difficult, tedious, and often impractical to make. The HP 8505 makes them simply, directly, and accurately. Its capabilities facilitate more precise designs, which improve system performance—for example, reducing distortion in communications channels, with results you can appreciate when you pick up the phone to talk overseas.

The 8505 is also the most highly programmable network analyzer presently available. With an HP 9830 as controller, and the Hewlett-Packard Interface Bus (HP's implementation of IEEE standard 488-1975), the analyzer can perform automatically. HP cassette-recorded programs, including accuracy enhancement, diagnostics, and performance verification, simplify system start-up. Key advantages of automating the network analyzer are extreme measurement accuracies by virtue of the system's ability to measure, store, and then subtract vector errors; ability to make many measurements quickly; and ability to manipulate data and put the answers in the desired format. In other words, to deliver "computed measurements": results, not just undigested data.

A singular contribution of the automatic system is its "Learn" mode, whereby measurements can be automated without programming. A single keystroke command on the desk-top controller will cause it to "read" and store the analyzer's front-panel switch settings and registers. The user can set up test sequences with the front-panel controls, store them in the controller, and recall them when needed. By recording these data on tape cassette, an entire test procedure can be created without writing a single program line.

A factory-assembled automatic network analyzer that includes the 8505, a 9830B desk-top computer with printer, an s-parameter test set, cassette programs, and appropriate cables and fixtures is available as Model 8507A for $45,490*. The 8505 alone costs $22,500*.



Now HP makes available microprogrammable computer power in component form.

To system designers who would like to take advantage of economical microprocessors but need a higher level of performance, HP offers the new 21MX-K *minicomputer on a board*. Designed to fill the need for high-power processing in a low-cost form, it integrates easily into OEM systems and is fully compatible with the entire 21MX family of computers, software, and peripherals.

The 21MX-K is essentially the powerful processing board of HP's popular 21MX computer, packaged as a component for use by system designers.

An extraordinarily fast 24-bit processor, the 21MX-K can be used as a high-performance alternative to microprocessors and microcomputers. It executes a register-to-register *add* in a 325-nanosecond cycle and, through instruction pipelining, it can combine and execute up to six instructions in a single machine cycle.

An optional instruction set ROM transforms the 21MX-K into a full microprogrammable minicomputer, with 128 instructions including floating point, multiply and divide, integer arithmetic and bit/byte manipulation. In this configuration, designers will find that it is relatively easy to tailor the 21MX-K to their specific applications, using the same high-level microprogramming tools that are available with all 21MX computers.

Incorporating true minicomputer features and performance, the 21MX-K is fully capable of serving the needs of high-performance instrumentation and data systems such as spectrometers, numerical control units, *smart* terminals, graphic display systems, medical diagnostic instruments—even intelligent remote satellites in a distributed computing network.

The 21MX-K processor is hardware- and software-compatible with the entire family of HP 21MX minicomputers. It supports connections to a broad range of peripherals from teletype to high-speed disc; and it is supported by the complete library of 21MX software including all languages, operating systems and diagnostics.

In addition to the processor and instruction ROM, other 21MX-K components include a front panel control assembly and cages for I/O and memory cards. Accessories include memory systems, user control store and writable control store.

Prices for the processor board (OEM quantity 100) are \$975* without and \$1205* with the instruction ROM.

For more information on these products, write to us. Hewlett-Packard, 1505 Page Mill Road, Palo Alto, California 94304.

*Domestic USA prices only.

HEWLETT hp PACKARD

Sales and service from 172 offices in 65 countries.

For assistance call: Washington (301) 948-6370, Chicago (312) 677-0400, Atlanta (404) 434-4000, Los Angeles (213) 877-1282, Toronto (416) 678-9430

Mail to: Hewlett-Packard, 1505 Page Mill Road, Palo Alto, CA 94304.
Please send me further information on

( ) HP 8505/8507A Network Analyzer
( ) HP 21MX-K processor/components

Name ____________________
Company ____________________
Address ____________________
City ____________ State ________ Zip ________

00648

tion of man. The latest strike, reported in *Nature* by Mary D. Leakey and her colleagues, is a collection of teeth and jaw fragments, most of them surface finds, in a fossil-rich formation of volcanic rocks not far from Olduvai Gorge in northern Tanzania. The fossils, mostly collected in 1975, evidently represent no fewer than 13 individual hominids, that is, members of the Hominidae, the primate family that includes man.

Three of the finds, a mandible containing both milk teeth and permanent teeth, and two isolated groups of milk and permanent teeth, were actually embedded in volcanic tuffs (compacted volcanic ash) rather than lying loose on the ground. Samples of tuffs stratigraphically identical with those that contained the fossils yield potassium-argon isotope ratios indicating that the specimens are more than 3.5 million years old. The geographic and stratigraphic locations of the individuals' remains do not, however, suggest that they once formed a social group.

Preliminary study indicates that some of the fossils can be assigned to *Australopithecus africanus* (the gracile, or relatively light-boned, species of "man ape") and others to the lineage, now proposed for membership in the genus *Homo*, whose remains have been found in the Lake Rudolf area of Kenya by Richard E. Leakey and in the Afar region of Ethiopia by Donald C. Johanson. So far Johanson's finds include the largest social group: three to five adults and two children. The Tanzania hominids, however, may be more than half a million years older than those from either Kenya or Ethiopia.

Applied Biochemistry

An important constituent of the ability of the world's agriculture to keep pace with rising demand is the steady evolution of new biochemical strategies to combat the problems of farmers. Two recent innovations illustrate this evolution: a chemical that increases the amount of nitrogen available to a growing crop after fertilizer has been applied to the soil and a herbicide that clears both grasses and broadleaf weeds from a field before the crop emerges and without affecting the crop itself.

The nitrogen stabilizer is N-Serve, manufactured by the Dow Chemical Company. It slows the conversion of ammonium nitrogen (the form in which nitrogen reaches the soil in commercial fertilizers) to nitrates by soil bacteria. Plants can utilize nitrates, but much of the nitrogen converted from ammonium nitrate is lost by being leached out of the soil or escaping into the air as a gas.

The conversion proceeds in two stages: first *Nitrosomonas* bacteria transform the ammonium nitrogen to nitrite, and then *Nitrobacter* organisms convert the nitrite into nitrate. Neither nitrites nor nitrates are adsorbed on soil particles, as ammonium nitrogen is, and so they tend to move away from the roots of the plants. N-Serve conserves fertilizer nitrogen by acting against *Nitrosomonas* bacteria so that the conversion of ammonium nitrogen into nitrite is retarded. According to the manufacturer, the chemical (a pyridine that is applied to the soil at the same time as the fertilizer) does not inhibit the conversion of nitrites into nitrates and has no significant effect on other microorganisms in the soil.

The herbicide is Roundup, manufactured by the Monsanto Company. It is a water-soluble formulation of the isopropylamine salt of glyphosate and is applied as a spray (at four pounds per gallon of water) when the weeds have reached a vigorous stage of growth but before the crop has emerged (or after it has been harvested). What is unusual about the herbicide is that it attacks the weed only by traveling through the leaves and stem to the roots, where it acts to kill the plant. When the herbicide comes in contact with the soil, it is inactivated. Thus, according to the manufacturer, it is nonpersistent and can do no harm to crop plants that emerge later. It is effective against both annual and perennial weeds.

Chat at the Highest Level

The first digital computers were programmed in machine language: the string of binary digits actually processed by the machine. Later programming was done in assembly languages, in which the basic semantic units are the individual instructions executed by the machine. Today communication with computers is carried out mainly in higher-level languages, in which complex statements are interpreted by the machine and automatically compiled into programs. The higher-level languages, with names such as FORTRAN, ALGOL and BASIC, consist of specially defined terms and marks of punctuation, combined according to the rules of an explicit grammar.

There remains a set of languages at a still higher level: the languages employed by people in everyday discourse. Indeed, these are the highest-level languages man has devised. In the past decade several projects have attempted to teach computers to converse in natural languages, and particularly in English. Some of these projects are discussed in a recent series of articles in the *IBM Journal of Research and Development*.

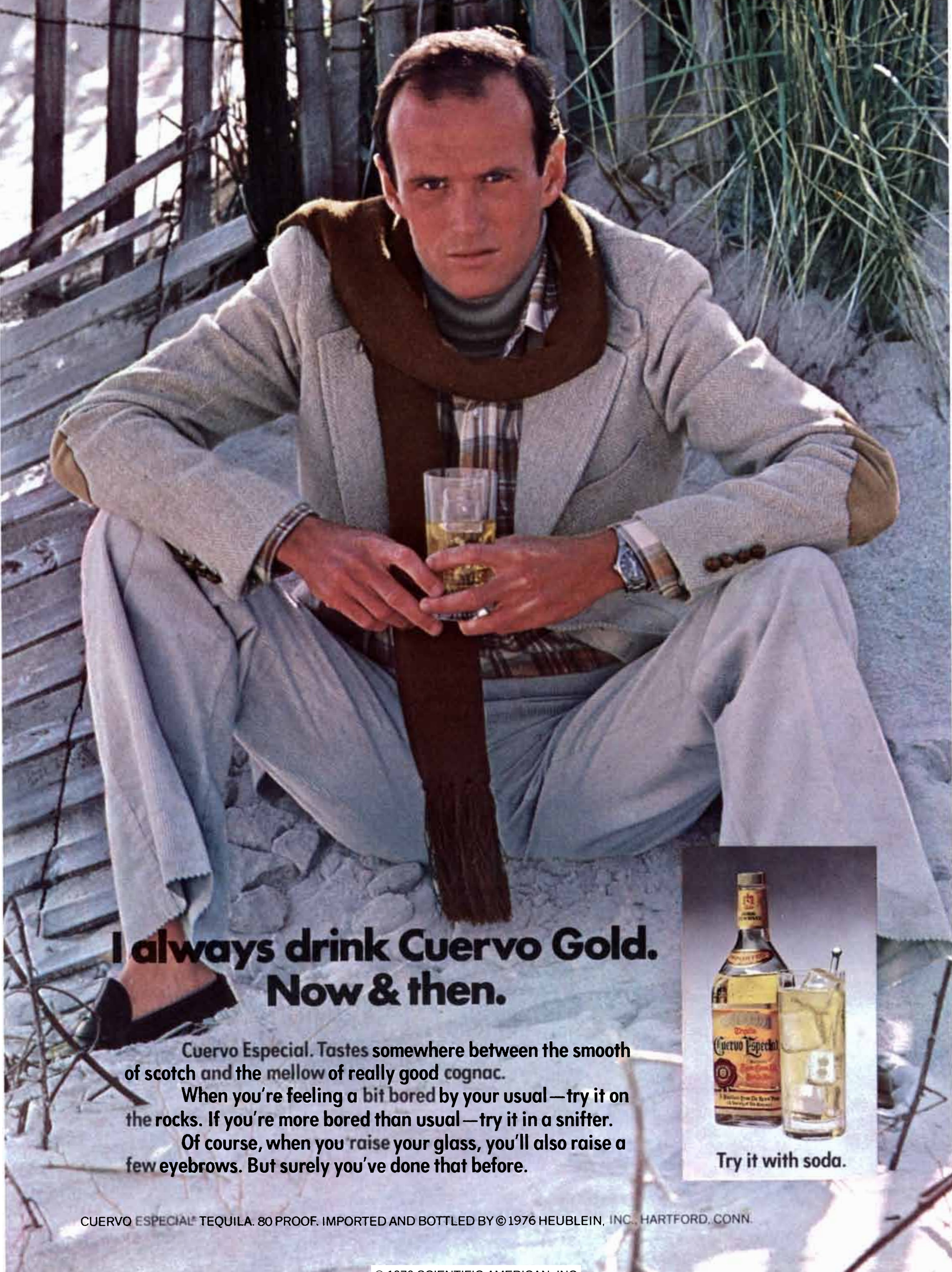
Programming languages, including even the higher-level ones, differ fundamentally from natural languages. The programming languages have a smaller vocabulary and a simpler grammar. More important, they are formal languages: every grammatical statement has one and only one meaning, and that meaning can be abstracted by a simple procedure from the structure of the statement. In a natural language, on the other hand, ambiguity is commonplace, and even statements that are grammatically flawless can often be understood only by inference. If the computer is to comprehend such a language, it must be supplied with explicit rules of inference.

A simplifying strategy adopted in all natural-language programs developed so far has been to sharply restrict the range of topics the computer can discuss. One program is concerned entirely with the manipulation of toy blocks and similar objects; another answers questions about the samples of lunar material collected during the Apollo program. By establishing such a limited universe of discourse it is possible to demonstrate a capability for conversation even though the program describes only a small subset of the English language.

A program developed by George E. Heidorn of the International Business Machines Corporation is concerned with problems in queuing theory. A program for simulating such a problem is written by the computer (in a conventional higher-level language) after the problem has been defined through an English-language conversation with the user. The machine asks questions until all necessary information has been provided. For example, Heidorn reports the following dialogue in a problem dealing with the queuing of cars in a gas station. The computer asked, "Where are the vehicles serviced?" and it was told, "At the pump." The machine responded by asking, "How long are the vehicles serviced at the pump in the station?" A certain capacity for inference is suggested by the exchange: the computer was able to determine that "the pump" must be "in the station."

Heidorn's program is based on about 300 "records," each containing information on words and concepts related to queuing problems. In addition the program incorporates about 300 rules for interpreting English sentences in terms of these records and about 500 rules for generating sentences from the records. The basis of the rules is a phrase-structure grammar. A sentence is analyzed by breaking it down into component phrases, and conclusions about the meaning of the sentence are derived from the syntactic relations of the phrases. The method is similar in some respects to the diagramming of sentences in the teaching of grammar.

Phrase-structure grammars are employed in most of the natural-language programs; an exception is a program called REQUEST, written by Warren J. Plath and his colleagues at IBM. REQUEST is a system for retrieving information from a "data base," in this case a file of information on *Fortune's* list of the 500 largest American companies. Queries submitted in English are inter-


I always drink Cuervo Gold.
Now & then.
Cuervo Especial. Tastes somewhere between the smooth of scotch and the mellow of really good cognac.
When you're feeling a bit bored by your usual—try it on the rocks. If you're more bored than usual—try it in a snifter.
Of course, when you raise your glass, you'll also raise a few eyebrows. But surely you've done that before.
Try it with soda.
CUERVO ESPECIAL® TEQUILA. 80 PROOF. IMPORTED AND BOTTLED BY © 1976 HEUBLEIN, INC., HARTFORD, CONN.


MARTINI
PORSCHE
MARTINI
5
PORSCH


TURBO
Turbolader

preted according to the rules of a transformational grammar.

In a transformational grammar the meaning of a sentence is still determined by analyzing the relations of its component parts, but this analysis is not attempted until an underlying structure in the sentence has been identified. The underlying structures represent irreducible "kernel sentences," all of which (in principle) can be interpreted unambiguously by simple laws. The diversity of expression possible in a natural language results from transformations operating on the kernel sentences. For example, all the kernel sentences are affirmative, declarative statements in the present tense and in the active voice. Negative statements, questions and sentences in other tenses or in the passive voice are all generated by transforming the kernel sentences according to well-defined rules.

A transformational grammar has the potential for describing a large body of language with a relatively small number of rules; in the REQUEST system there are about 100. REQUEST is, of course, confined to answering questions that pertain to its data base, but within that domain it allows considerable lexical and syntactic variation. In an analysis of the system Stanley R. Petrick of IBM considers the computer program's analysis of a particular query: "Were GE's earnings greater than IBM's in 1973?" Recognition of the various elements in the sentence is relatively straightforward. The nouns "GE," "IBM" and "earnings" are included in the machine's lexicon, and so is the phrase "greater than," which calls up a computational function the machine is programmed to perform. Nevertheless, the interpretation of the sentence is not a trivial procedure. First the machine must correctly identify the number "1973" as representing a year. Then it must determine that the year qualifies both "GE's earnings" and "IBM's." Finally, it must understand that "IBM's" refers to "IBM's earnings" and supply the deleted word. The REQUEST system correctly answered the question, and it could have done so when presented with several other possible phrasings of it.

The goal of computer communication in natural languages is formally distinct from investigations of artificial intelligence. It is closely related, however, to a well-known proposal for a test of artificial intelligence. The British mathematician A. M. Turing suggested in the 1940's that a machine could be considered intelligent if someone conversing with it could not tell whether he was talking with the machine or with another person. None of the natural-language programs developed so far could pass the Turing test, but some of them chat quite plausibly.

Constant Constants

Astronomers are still trying to establish unequivocally whether or not quasars, which on photographic plates look like stars, are the remotest and most luminous objects in the universe. If the red shifts exhibited by their spectra are due to the cosmological expansion of the universe, the same phenomenon that gives rise to red shifts in the spectra of galaxies, one must conclude that the starlike quasars are up to 100 times brighter than ordinary galaxies and many times farther away.

The red shifts of cosmic objects are usually determined by an analysis of emission lines in the optical spectrum: the lines of characteristic wavelength emitted by elements heated to incandescence. For all distant objects the emission lines are found to be shifted from their normal position toward the red end of the spectrum. The red shift, z, is obtained by dividing the displacement in wavelength by the normal wavelength. For example, when an object is receding at six-tenths the velocity of light, the emitted wavelength is stretched to exactly twice its initial value, giving z a value of 1. A few quasars have a red shift of more than 3, indicating that they are receding at about 90 percent of the velocity of light and that their radiation has been traveling toward us for 90 percent of the age of the universe.

If quasar red shifts are indeed cosmological (which some astronomers continue to doubt), they offer a unique opportunity for examining the constancy of fundamental physical values over cosmic time periods. A recent observation by astronomers at the National Radio Astronomy Observatory in Green Bank, W.Va., tends both to support the cosmological origin of quasar red shifts and to provide strong evidence for the constancy of such values. The observation was made by Morton S. Roberts, Robert L. Brown, W. D. Brundage and A. H. Rots of Green Bank in collaboration with M. P. Haynes of Indiana University and A. M. Wolfe of the University of Pittsburgh.

The quasar in question, AO 0235 + 164, resembles a number of others in that its spectrum is devoid of emission lines. Some of its radiation, however, is absorbed by intervening gases, so that it is possible to establish a red shift for certain absorption lines. One such set of lines yields a z of .851; another yields a z of .524. Since the quasar must lie behind the absorbing material, its z must be greater than the higher value. In many instances where the red shift of the quasar itself can be determined the red shift in the absorption spectrum is similar enough to the red shift of the quasar to suggest that the absorbing gases may well represent material ejected from the quasar. In the case of AO 0235 + 164 the absorption at a red shift of .851 could therefore be due to a shell of gas near the quasar. On the other hand, the material producing the absorption at a red shift of .524 would have to be so much closer to us, if z represents its cosmological recession velocity, that its ejection from AO 0235 + 164 seems quite improbable. Nonbelievers in the cosmological interpretation of quasar red shifts would argue, however, that the z of .524 simply represents a subtraction of two velocities: the velocity of the quasar minus the velocity of an expanding shell or jet.

The Green Bank astronomers directed the observatory's 300-foot radio telescope toward AO 0235 + 164 to see if any of its radiation might be absorbed in the radio region of the spectrum, specifically by neutral, or un-ionized, hydrogen, which normally radiates (and absorbs) at a frequency of 1,420 megahertz, also known as the 21-centimeter line. Brown and Roberts had previously detected hydrogen absorption at a z of .692 in the direction of one other quasar, 3C 286, but in that instance there was no absorption counterpart in the optical part of the spectrum. Not only did they and their co-workers detect hydrogen absorption in the direction of AO 0235 + 164 but also they found that the red shift was almost precisely the same as one of the two values obtained at optical wavelengths: .52385 ± .00001, radio, compared with .52393 ± .00010, optical. There could be no doubt that whatever cloud of gas had given rise to the two absorptions was part of a single mass.

A striking feature of the absorbing cloud is that it produces such sharp absorption lines in both the optical and the radio parts of the spectrum. If the cloud had been ejected from the quasar, one would expect it to consist of material with an assortment of velocities, which would broaden the absorption lines. Actually the absorbing material resembles the cold, tenuous gas between the stars in a normal galaxy. At the distance represented by a red shift of about .5 the starlight from the galaxy itself would be invisible. (The largest measured red shift for any galaxy is about .3.) It is therefore plausible to assume that the light from AO 0235 + 164 has on its way to us passed through the gas of an invisible galaxy, thus supporting the cosmological hypothesis of quasar red shifts.

The close agreement between the radio and optical values for the red shift of AO 0235 + 164 (to within one part in 10,000) also places a new and strong constraint on any hypothesis that attempts to show that large red shifts are not due to the expansion of the universe but are produced in some fashion by a slow change in atomic constants over a long period of time. An analysis of the optical- and radio-absorption measurements, which represent conditions as they were when the universe was between half and two-thirds its present age, shows that certain fundamental atomic values, such as the masses of the electron and the proton, cannot have changed more than two parts in 10^{14} per year, or at most one part in 2,500 over the age of the universe. For a greater variation to have occurred exactly compensatory changes would be required in two or more values.

Fire in the Lake

Lake Kivu, a large lake in central Africa, has a remarkable feature that is not usually associated with bodies of water: it is charged with fossil fuel. The waters of the lake hold in solution some 60 billion cubic meters of the natural gas methane, equivalent in energy content to roughly 60 million tons of crude oil. Harald Steinart, writing in the newsletter of the German Research Service, describes the efforts of German geologists to find the source of this curious "deposit" and to determine how it can best be utilized by the developing economies of the countries on the shores of Lake Kivu, Rwanda and Zaïre.

Some 10 years ago an experimental

pipeline was laid from Gisenyi, on the north side of the lake, down to 300 meters below the surface. When water was pumped out, the reduced pressure at the mouth of the pipe caused the gas to bubble out of solution. Because of its high carbon dioxide content, however, the gas was incombustible, and it had to be filtered to remove part of the carbon dioxide before being utilized to heat the boilers of a nearby brewery. Nevertheless, by 1973, 60 million cubic meters of gas with a methane content of about 14 million cubic meters had been extracted from Lake Kivu to brew beer.

Compared with current plans for exploitation the Gisenyi pipeline was a modest effort. The German development-aid staff believes the "lake gas" could provide energy for the industrialization of the district around the lake or for the production of nitrogen fertilizers. The first step, however, is to explore the deposits themselves and to investigate means of tapping them on a large scale. According to the German geologists, the most efficient method of tapping would be to pump the water to the surface and degas it there. The degassed water would then be transferred to shallow basins around the edge of the lake and allowed to return slowly to the natural circulation.

The question of how the gas deposit—the only known deposit of its kind—came into existence is of more than academic interest. The answer will reveal whether the gas content of the lake is continually being renewed or whether after a few years of intensive tapping the resource will be exhausted. There are two main hypotheses for the presence of the gas in Lake Kivu. One is that the methane is of biological origin: it is "marsh gas" of the kind that forms in the bottom sediment of warm lakes rich in organic matter but poor in oxygen. This hypothesis is contradicted by the high concentration of carbon dioxide, which can form only in the presence of oxygen.

The alternative hypothesis is that the gas is of volcanic origin. Central Africa is intensely volcanic, and it is possible that underwater eruptions in Lake Kivu have released both the carbon dioxide and the methane. In normal volcanic gases, however, there is not nearly enough methane to account for the enormous amounts of the gas in Lake Kivu. The German geologists are leaning toward a compromise between the two hypotheses: that the methane comes from the decay of organic matter in bottom sediments and the carbon dioxide is of volcanic origin. Should this prove to be the case, the countries in the area will have cause for optimism about the future of Lake Kivu as an energy resource: since large quantities of dead vegetation are continually being fed into the lake and converted into marsh gas at the bottom, the lake's methane may not be exhausted for a long time.

Cosmic Gamma-Ray Bursts

Unknown until four years ago, they are detected at the rate of about one a month by instruments in satellites. They "outshine" all other sources of comparable radiation, and their origin remains a mystery

by Ian B. Strong and Ray W. Klebesadel

The first description of the phenomenon now known as a cosmic gamma-ray burst was published more than three years ago. Although a considerable body of observational data has now been assembled on more than 50 individual events, there is still no generally accepted explanation of how the gamma-ray bursts are generated or of where in the universe they originate. We know they come from outside the solar system, but whether they come from within our galaxy or outside it no one is yet sure.

If cosmic gamma-ray bursts were a subtle effect whose detection required highly sensitive instruments, if they lasted so briefly that they could be confused with a burst of electronic noise or if they were simply very rare, the continuing mystery of their origin would be less surprising. Such, however, is not the case. Cosmic gamma-ray bursts are so intense that they temporarily dominate all other celestial sources of gamma rays by several orders of magnitude. Many of the bursts last for more than 10 seconds, making them easily visible even in data with poor time resolution, provided by detectors that accumulate photons for large fractions of a second. In fact, all but a few of the known bursts were detected by instruments not suited to the purpose. The newer detection systems record an average of a dozen bursts a year. If a comparable phenomenon had happened to fall within the visible region of the spectrum, it would have been known to early man and would have been vividly described in the oldest astronomical writings.

Such a dramatic and unexpected discovery naturally stimulated astrophysicists to fairly uninhibited speculation. Depending on how one classifies the various hypotheses, at least 30 distinct models have now been advanced, each purporting to explain where the gamma-ray bursts come from and how they are generated. As we shall see, the data are more than adequate to stimulate the imagination and quite inadequate to limit it.

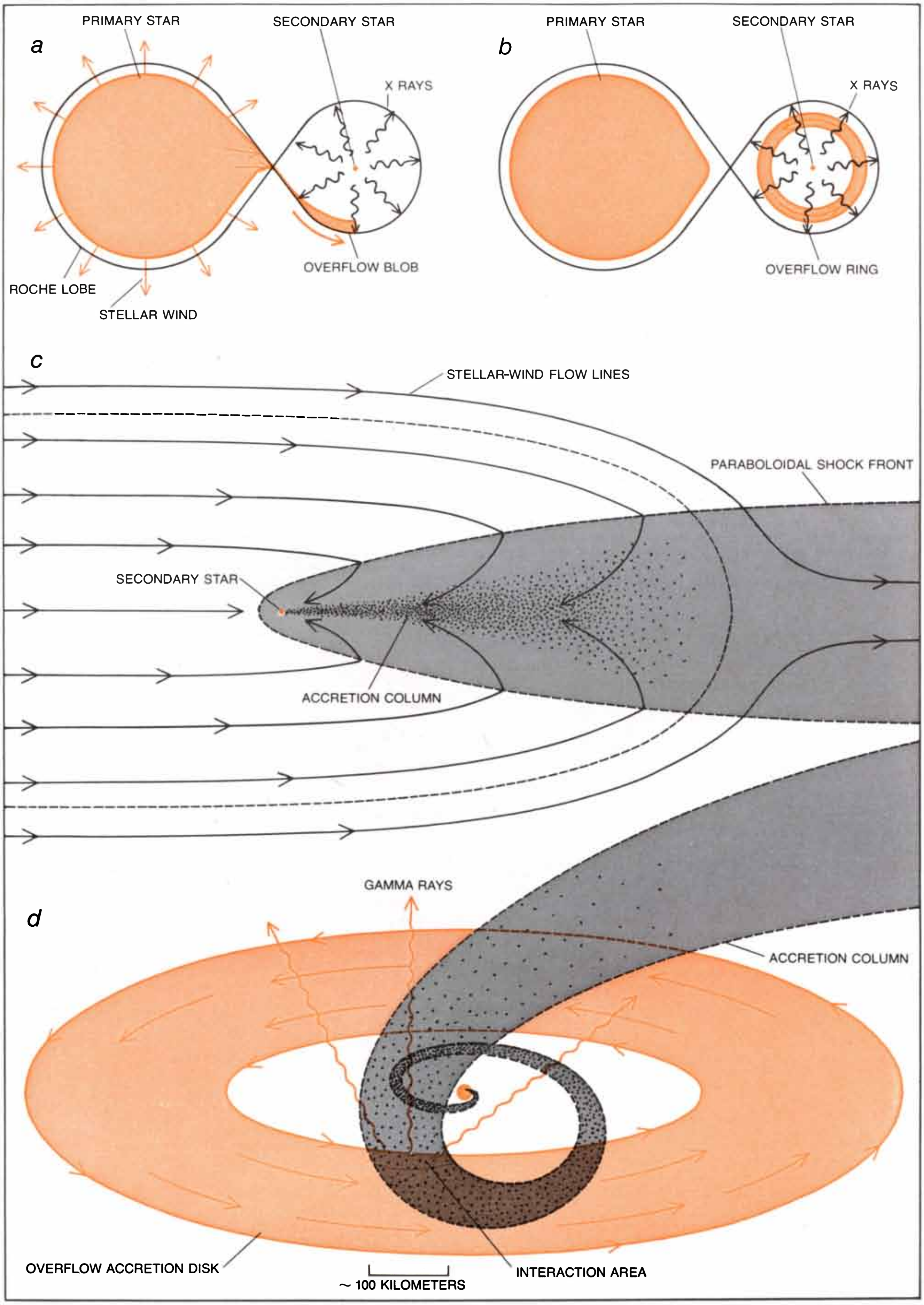
The story begins 13 years ago, when the U.S. launched the first of a series of Vela satellites for the purpose of monitoring the nuclear test-ban treaty of 1963, which forbids the signatories from exploding nuclear devices in the atmosphere or in outer space. The satellites, the instrumentation for which was designed at the Los Alamos Scientific Laboratory, were equipped to detect radiations of the kind released by nuclear explosions: the gamma rays emitted at the moment of fission or fusion, thermal radiation (most of it in the X-ray region of the spectrum) and energetic neutrons. The Vela satellites are paired: two identical vehicles circle the earth on opposite sides of a circular orbit 250,000 kilometers in diameter, so that no part of space is shielded by the earth from observation. Over most of the past 13 years the Vela system has consisted of several satellites, thus providing redundancy and ensuring that the detected radiation has not simply originated in the immediate vicinity of one satellite.

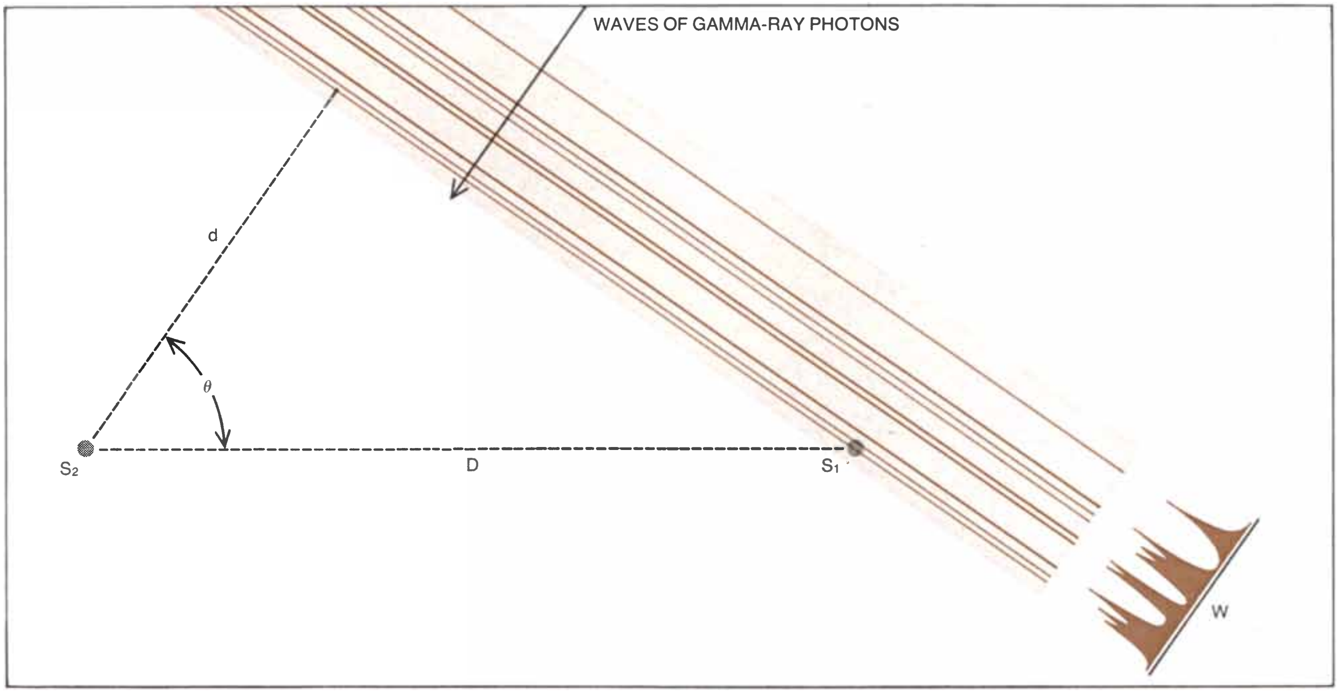
Successive Vela satellites were improved by the addition of new detectors and by better processing of the signals to allow a more detailed study of the natural radiation environment, for its possible effect on the instrumentation as well as for the possibility that nature might mimic the signals from nuclear weapons. The new data would also be of great scientific interest.

How the Bursts Were Detected

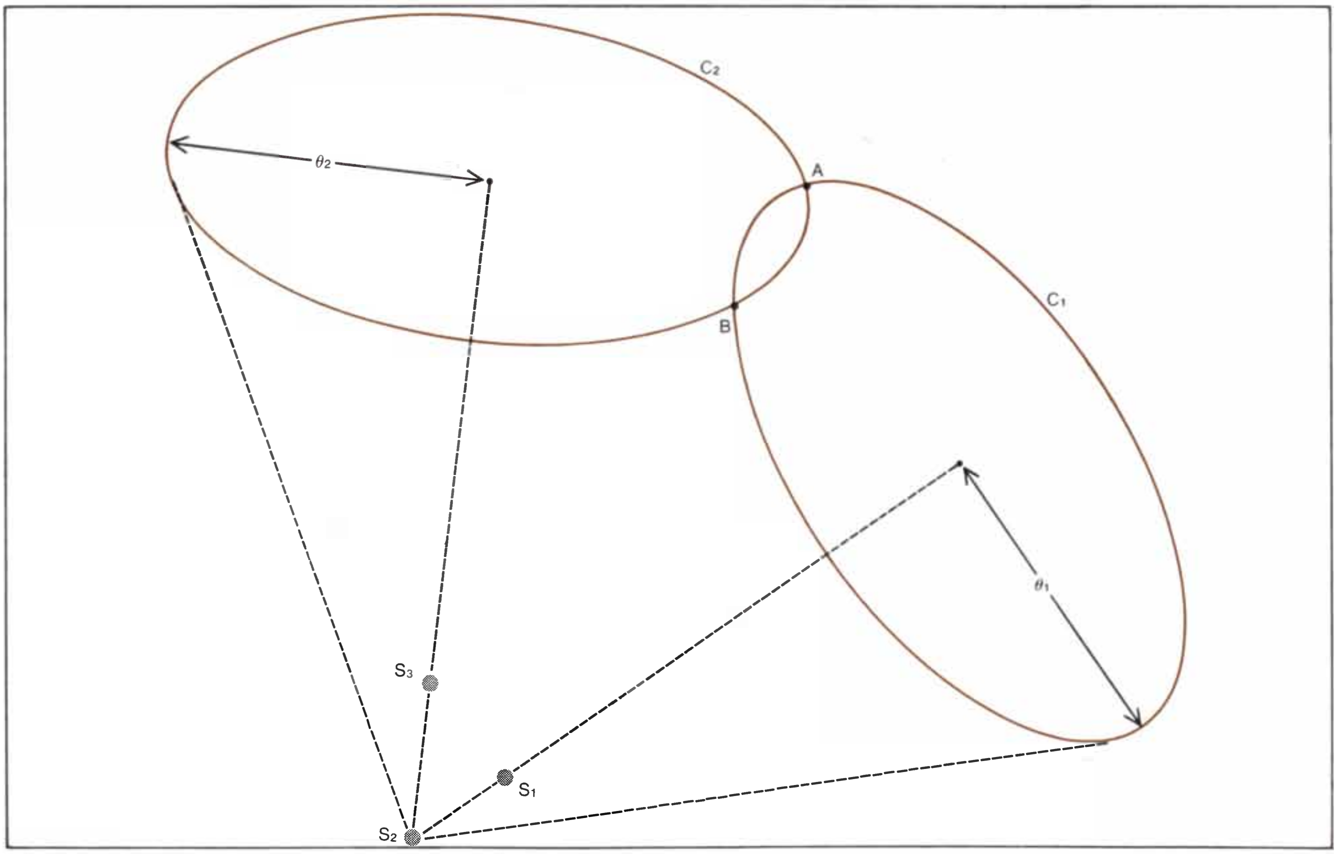
The Vela satellites have all performed exceptionally well, exceeding their expected lifetimes (which were based on the average lifetimes of other satellites of the period). This performance had the disadvantage, however, of generating considerably more data than had been expected, so that the processing of the more routine data fell far behind. Such a fate befell the data from a gamma-ray detection system added to the fourth series of Vela satellites. The system was designed to study the details of

PLASMA-INTERACTION MODEL has been proposed by the authors as one conceivable source for the energetic photons detected near the earth as gamma-ray bursts. The model is suggested by the possibility that the X-ray source known as Cygnus X-1 may also give rise to some of the bursts. Cygnus X-1 is evidently a binary system consisting of a large primary star and a compact secondary, which may be either a neutron star or, as some think, a black hole. A black hole is the hypothetical object that would be created if a spent star collapsed to infinite density, producing a gravitational field so powerful that nothing, not even light, could escape. Whether or not the secondary object is a black hole, a stellar "wind" of charged particles blows outward from the primary star, enveloping the compact secondary (*a*). A portion of the stellar wind forms an accretion column in the wake of the secondary and continuously spirals inward toward it (*c*). The usual X-ray output of Cygnus X-1 arises from the deceleration of plasma particles being transferred in this way from the primary to the secondary. Periodically, however, blobs of plasma from the primary star may spill beyond the Roche lobe, or Roche limit, and thus might be directly captured by the gravitational field of the secondary. The blobs first form a thin ring, which becomes a slowly contracting thin disk with a hole in it (*b*). Although the disk lies strictly in the plane of the orbit of the binary system, because of fluctuations in the stellar wind the plane of the accretion column can wander somewhat above or below that plane. If the two planes are not the same, the overflow plasma in the contracting disk will eventually collide violently with the plasma flowing in the accretion column, releasing millions of electron volts of energy per colliding particle (*d*). The result is a burst of gamma radiation, probably augmented by high-energy photons from decaying pions formed in proton-proton collisions. The sudden production of radiation tends to separate the two plasmas, interrupting the release of radiation. This could account for the fine structure seen in gamma-ray bursts. The gamma-ray burst occurs only while the strengths of the two plasma flows are comparable. Eventually the overflow disk will dominate, and the usual X-radiation will be altered in its characteristics until the supply of plasma is exhausted, which may take several months. Such changes have been observed in X-ray output of Cygnus X-1, and several may coincide with gamma-ray bursts.


a
PRIMARY STAR
SECONDARY STAR
X RAYS
OVERFLOW BLOB
ROCHE LOBE
STELLAR WIND
b
PRIMARY STAR
SECONDARY STAR
X RAYS
OVERFLOW RING
c
STELLAR-WIND FLOW LINES
PARABOLOIDAL SHOCK FRONT
SECONDARY STAR
ACCRETION COLUMN
d
GAMMA RAYS
ACCRETION COLUMN
OVERFLOW ACCRETION DISK
~ 100 KILOMETERS
INTERACTION AREA



DIRECTION ANGLE OF GAMMA-RAY EVENT can be determined if the arrival times of photons from the burst are accurately measured by two satellites whose location in space and separation, *D*, are also accurately known. The difference in arrival times yields the distance *d*. Knowing *d* and *D*, one can obtain angle θ, the angle the incoming plane wave makes with a line drawn between the two satellites. Note, however, that angle θ can be rotated 360 degrees around line *D*. As a result data from only two satellites can do no more than specify a circle on the celestial sphere. The photons may have arisen from a point anywhere on the perimeter of the circle. If the recorded gamma-ray event has a structure in time (*W*), common features can be used in determining the arrival times at the two satellites.



FURTHER RESTRICTION IN LOCATION of a source can be accomplished if an event is recorded by three satellites. The difference in arrival times at one pair of satellites, say S_1 and S_2, yields angle θ_1, which determines the size of one circle in space, C_1, whose center is on a line joining the two satellites. The difference in arrival times at another pair of satellites, S_2 and S_3, yields angle θ_2, which determines the size of a second circle, C_2, whose center is on a line joining the second pair of satellites. Since the source must lie on both circles, it must lie at one of the two intersecting points, either *A* or *B*. One can be eliminated if the event is recorded by a fourth satellite not in the plane of S_1, S_2 and S_3, or if one of the two points can be shown to have been hidden from another satellite that also detected the burst.

bursts of gamma rays that might last for seconds or even minutes before dying away, assuming that such bursts existed.

It was 1969 before such a burst was found, embedded in the data from 1967. Both Vela 4 satellites detected more or less identical signals, showing the source to be roughly the same distance from each satellite. The source could be close to the earth, but it could equally well be close to the sun. If we knew the arrival times accurately enough, that could be checked, since both satellites are always the same distance from the earth but are usually at different distances from the sun.

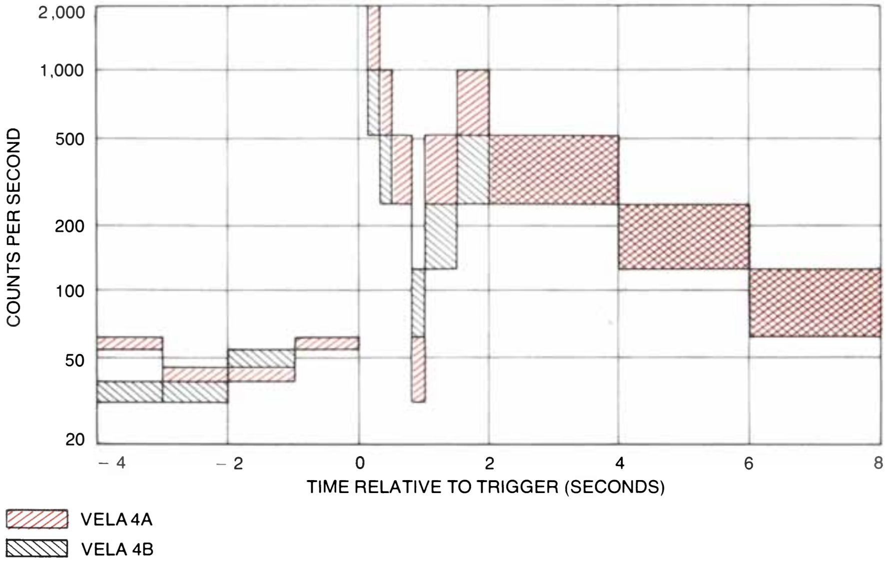
Later satellites were modified so that they could record the time of such events to within .016 second. By 1972 a careful search for random events that had occurred almost simultaneously in the data from two, three or four satellites revealed a surprising number of events that not only were almost simultaneous but also were of the same order of intensity and the same distribution of intensity with time. If the sources were outside the orbit of the satellites, they would need to be quite far away; otherwise the nearest of the detectors would record a signal stronger than that recorded by the detectors farther away. Moreover, the timing of events at two or more satellites yielded directional information showing that the gamma rays could not be coming from the earth, the moon or the sun or even from any of the planets.

We could now be sure that we were dealing with a hitherto unobserved phenomenon, presumably originating outside the solar system and clearly involving enormous quantities of energy at the source. The first account of this puzzling phenomenon was published in 1973. A total of 55 bursts have now been recorded, not only from Vela satellites but also from many other satellites and from balloons.

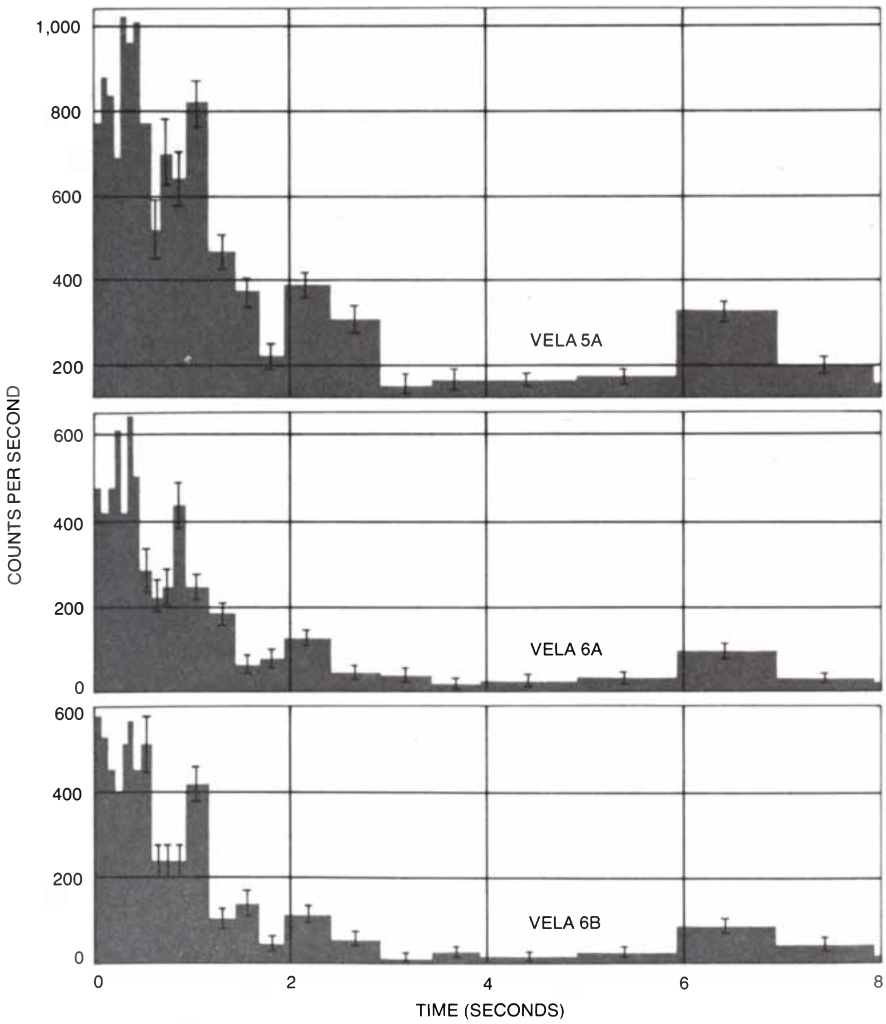
The Direction of the Sources

The gamma-ray detectors aboard the Vela satellites consist of six cesium-iodide crystals that emit visible light when gamma rays penetrate them. The light is converted into electronic pulses and recorded. Working in combination, the six crystals are equally sensitive to radiation from any direction. Because the intensity of the radiation being examined might be expected to decrease more or less exponentially with time, the detectors were designed to store their counts in a sequence of expanding time periods, beginning with 1/64 second and continuing for a total of 15 minutes. The detectors begin storing data only if the count rate increases by a large amount in any given quarter second.

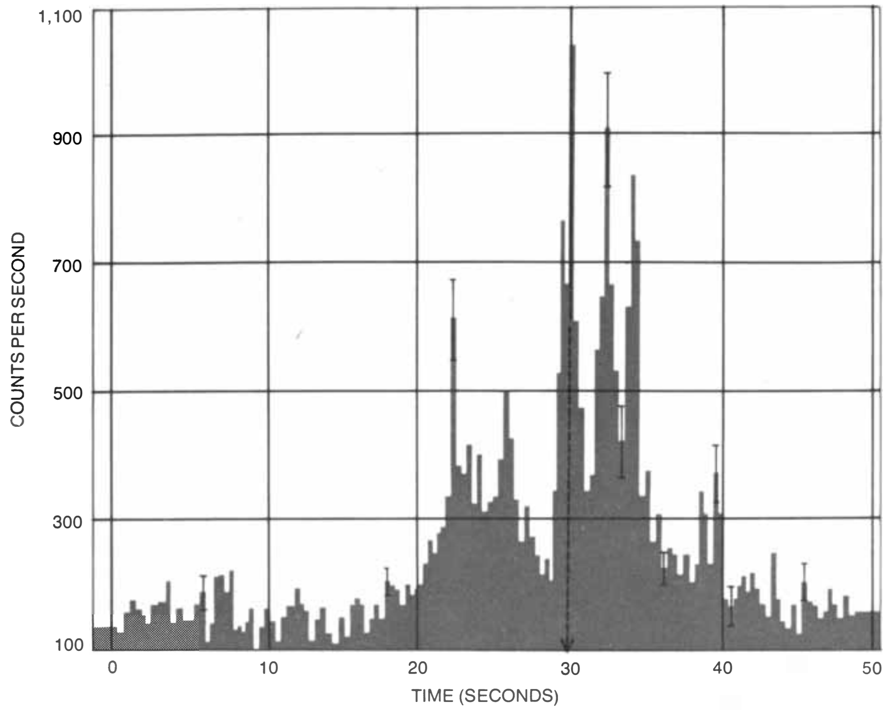
It is evident that if two satellites detect a pulse of gamma rays simultaneously, the source of the radiation must be equi-



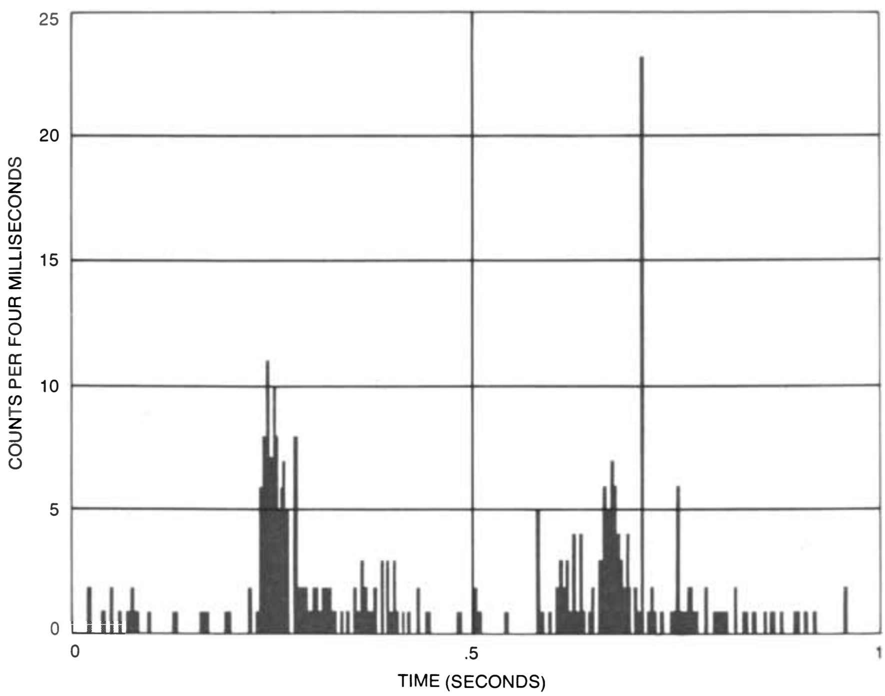
FIRST GAMMA-RAY BURST was observed on July 2, 1967, by two Vela satellites, which carried instruments designed to detect radiations of the type that would be produced by a nuclear explosion in space. Responding almost simultaneously, detectors aboard the satellites, traveling in orbit about 250,000 kilometers apart, gave rise to nearly identical responses. Detectors respond to photons in energy range between .2 MeV (million electron volts) and 1.5 MeV.



TRIPLE RECORDING of a gamma-ray burst was made by three Vela satellites on August 22, 1970. The cesium-iodide gamma-ray detectors begin storing data only if the count rate rises steeply in any given quarter second. The counts are then stored in a sequence of time slots that systematically double in length. It can be seen that the fine structure of the burst, the seventh to be observed, is recorded in the same way by the detectors on all three of the satellites.



FORTUITOUS RECORDING of a gamma-ray burst that barely triggered the Vela detectors was made by instruments aboard *Apollo 16* on April 27, 1972, just before the manned spacecraft reentered the earth's atmosphere after its trip to the moon. The *Apollo 16* detectors, larger and more sensitive than those on the Vela satellites, provided the most detailed picture of a gamma-ray burst available until the first event recorded by the West German satellite *Helios-2* (*see illustration below*). After reaching a peak of more than 1,000 photons per second, the count fell to zero for a period of 18 milliseconds (*broken line in color*), then returned almost to its previous level. On this scale the brief period at zero cannot be accurately depicted.



FIRST GAMMA-RAY BURST OF 1976, detected on January 23, was also the first to be recorded by an instrument specifically built for the purpose. The instrument was designed by the Goddard Space Flight Center and carried aboard *Helios-2*. The event recorded by *Helios-2* is typical both in its general structure and in the tendency for the fine spikes to arrive in clumps. The detector counts gamma rays in consecutive time periods of four milliseconds each. Only the first second of the burst, which lasted for many seconds, is shown here. The highest peak consists of 23 counts with no more than a single count in the immediately adjacent time slots. Therefore the peak signal probably lasted for no more than two milliseconds, including both rise and fall. This implies source is remarkably small, no more than 300 kilometers in diameter.

distant from both of them. Indeed, it must be somewhere in a plane perpendicular to a line drawn between the two satellites. If the arrival times are not simultaneous, one can readily determine the angle between the direction of the arriving wave and the line joining the two satellites. The angle determines a circle projected onto the sky on which the source must lie.

If the gamma radiation is detected by a third satellite, the angle between the wave direction and a line joining the third satellite with one of the others determines the size and location of a second circle intersecting the first. The source will lie at one of the two points where the circles intersect. Which of the two it is cannot be determined unless a fourth satellite not in the same plane as the other three also records the event. Sometimes other information is available for distinguishing between the two possibilities. Many satellites (but not the Vela satellites) come so close to the earth that only half of the sky is visible to them at any one time. If such a satellite has a detector that responds to a gamma-ray burst in any way, and only one of the two possible sources is visible, one can choose the correct source direction. On the basis of data from the hard-X-ray detector on the OSO-6 satellite G. G. C. Palumbo, G. Pizzichini and their colleagues at the University of Bologna were able to eliminate one of the two alternative sources for a number of gamma-ray bursts, and other choices were made on the basis of responses from the *Uhuru,* SAS-2 and OGO-5 satellites.

The direction of one of the gamma-ray-burst sources was independently confirmed by a directionally sensitive X-ray telescope on the OSO-7 satellite. Another event is of particular interest because, although it was so weak that it barely triggered the Vela detectors, it was recorded in considerable detail by large, sensitive instruments aboard the manned spacecraft *Apollo 16* on its return from the moon, just before it reentered the earth's atmosphere on April 27, 1972.

The Nature of the Bursts

The bursts exhibit great variety. The rise time of the initial pulse varies from .002 second to more than a second. The initial pulse lasts from less than .01 second to 10 seconds. There can be anywhere from one pulse to five or more distinctly separate pulses, and the entire burst can last from .1 second to 100 seconds or more. As better instruments have begun to provide observations with better time resolution we see that the structure of the bursts is even more complex than we had originally believed. The amount of energy reaching the satellites has ranged from 3×10^{-6}


*PRICE MAY VARY ACCORDING TO STATE AND LOCAL TAXES
12 YEARS OLD WORLDWIDE • BLENDED SCOTCH WHISKY •
86 PROOF • GENERAL WINE & SPIRITS CO., NEW YORK, N.Y
$85000
OLD-87
$16500
$3200
$300
$190000
20331
20331
$13*
12
CHIVAS REGAL
1801
BLENDED SCOTCH
WHISKY
Why think of it as an expensive Scotch
when you can think of it as an inexpensive luxury?

Fallingwater at Mill Run, Pennsylvania. Designed in 1936 by Frank Lloyd Wright. Photographic sequence ranging from 10.4 inches (upper left) to one-tenth mile (lower right) records varied aspects of the house from its smallest detail to its overall integration with the building site. The SX-70 photographs above are reproduced actual size.

How to buy a wagon and drive home in a sportscar.

Simple.
Just buy a Rabbit.

For like a wagon, the Rabbit has a fold-down rear seat, giving you 19.4 cubic feet of cargo room. (Which, incidentally, is more room than you'll find in the trunk of a Cadillac Fleetwood.)

And like a sportscar, the Rabbit has incredible performance. With front-wheel drive that literally pulls you around corners. And acceleration that takes you from 0 to 60 faster than a Jaguar XJ-6L or an Alfetta GT.*

Few cars in the world can give you all this room and performance.

And no other car — none that we know of — can give you all this room and performance *plus* 39 mpg on the highway and 25 in the city.

(These are the impressive 1976 EPA estimates of what the Rabbit got with a standard transmission. Your mileage may vary, depending on how and where you drive, optional equipment and the condition of your car.)

No matter which you're looking for — wagon or sportscar — you owe it to yourself to at least test drive the Rabbit before making up your mind.

After all, how many other cars can you think of that give you the opportunity to drive home in both.

* Tests conducted by Motor Trend.

The Rabbit. You've got to drive it to believe it.

erg per square centimeter per second to 300 times that value, with an average of about 10^{-4} erg per square centimeter per second.

Good information on the wavelength distribution of the energy emitted is essential to the understanding of any astrophysical process involving radiation, whether it is in the visible region of the spectrum or in the X-ray and gamma-ray region. The Vela detectors provide only a minimal amount of such spectral information. The Vela 5 detectors respond to photons with energies of between 150 and 750 kiloelectron volts (KeV), whereas the Vela 6 detectors respond to photons with energies of between 300 and 1,500 KeV. Thus the two ranges overlap. The situation is analogous to the use of red-sensitive and blue-sensitive photographic plates for studying stars in the visible region. From the difference between the brightness of a star on one plate and the brightness on the other we can say that the star is reddish or bluish and to what extent, but not much more. In a similar way, by comparing the number of photons counted during a burst by a Vela 5 detector and the number counted by a Vela 6 detector we can say whether the radiation is soft or hard (lower-energy or higher-energy).

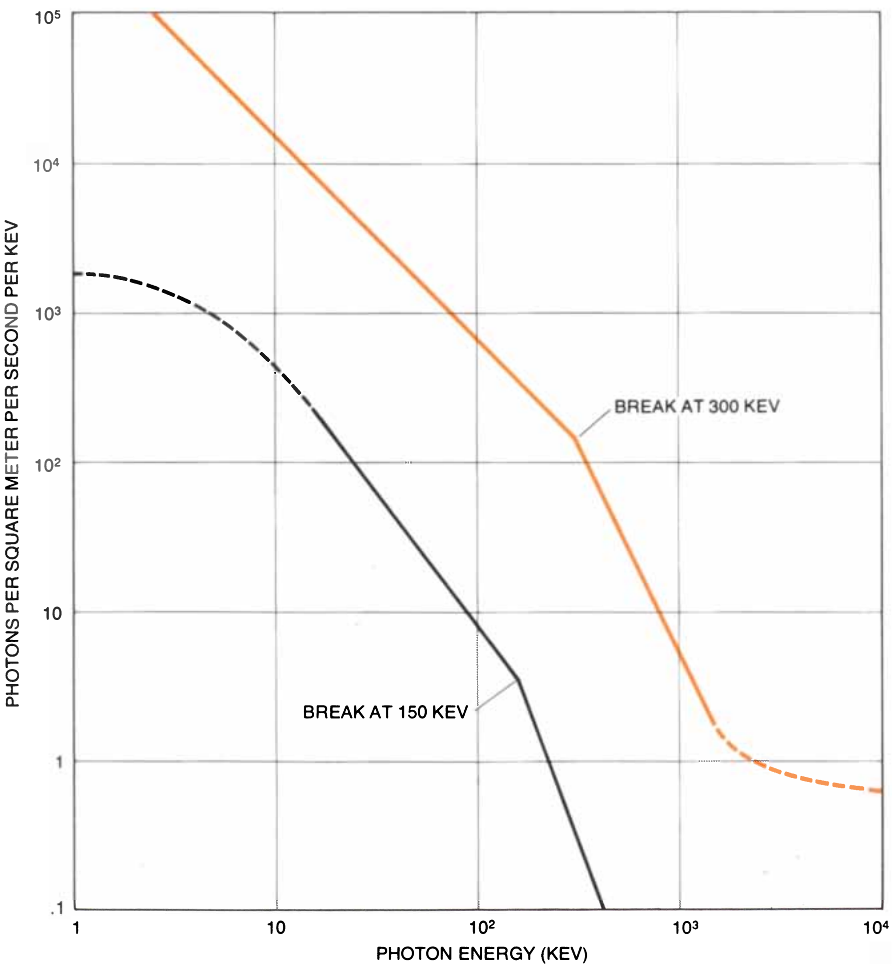
The IMP 6 and IMP 7 satellites, which measure brightness in eight or nine different energy ranges, have supplied spectra of intensity with respect to energy for about 30 of the longer gamma-ray bursts. Thomas L. Cline and U. D. Desai of the Goddard Space Flight Center have concluded from a study of these data that the spectra, when averaged over the entire burst, look much alike. The most detailed single spectral measurement, the one recorded by *Apollo 16,* has been carefully analyzed by Albert E. Metzger and his co-workers at the Jet Propulsion Laboratory of the California Institute of Technology. Although the curve shows no evidence of any emission lines, it is not completely smooth. The photons are on the average much more energetic than those typical of cosmic X-ray sources, and the number of photons falls off less steeply at higher energies. There is a suggestion of extra numbers of photons at energies of a few million electron volts, over and above what one might expect from the main part of the curve.

The *Apollo 16* burst seems to have a more complicated temporal structure than some of the other events, although it may be due to the high quality of the data rather than to the character of the burst. It is evident that the radiation can start and stop rather abruptly, say within a few milliseconds. During the highest peak of the best *Apollo 16* data the radiation stopped completely for 18 milliseconds. An important corollary of such rapid fluctuations is that the region radiating the gamma rays must be a small one. If the photon count rises or falls sharply in *t* seconds, one can infer that the effective diameter of the source cannot exceed the distance the radiation can travel in the same interval. On the basis of the *Apollo 16* data one can say that the maximum diameter of the source was 1,500 kilometers, hardly more than a tenth the diameter of the earth.

We cannot say whether or not all sources of gamma-ray bursts are roughly the same size. Whatever the case, the source of the burst recorded by the detector on *Helios-2* on January 23, 1976, sets a new record for smallness. Gamma-ray photons were counted for consecutive periods of four milliseconds. During this event the detector recorded on the average one photon every other counting period. In one period 23 photons were counted, with one photon in the preceding period and none at all in the succeeding period. It is highly unlikely that this brief "spike" of gamma rays lasted for a full four milliseconds. It would almost certainly have spilled into two counting periods. Its duration was probably two milliseconds or less, including its rise and fall and any sustained emission that might have occurred at its peak. We conclude that the rise time is one millisecond or less, in which case the maximum diameter of the source is only 300 kilometers.

The Energies of the Sources

It is evident that if we know the amount of energy reaching the detectors, we should be able to calculate the total amount emitted at the source, provided that the energy is emitted uniformly in all directions (which we have no reason to disbelieve) and provided that we know the distance to the source. There, of course, is the rub. Unless the sources are very close to the detectors, which they are not, we cannot directly determine their distance. The distance could be established only if a gamma-ray burst were unequivocally associated with a celestial object whose distance had been measured by other means. One



SPECTRUM OF GAMMA-RAY BURST (*color*) bears a resemblance to the average spectrum (*black*) of X rays emitted by Cygnus X-1, the X-ray source suspected of containing a black hole. The gamma-ray spectrum is based on measurements made by the instruments on *Apollo 16.* Both curves are plotted in photons per square meter per second per KeV (kiloelectron volts), which emphasizes the low-energy end of the spectrum. The change in slope at 300 KeV in the gamma-ray spectrum is evidently real. It may be significant that the spectrum of Cygnus X-1 shows a similar break, but at a lower energy. The dashed regions of the two curves shown here represent uncertainties in the observational data or in their interpretation.

can nonetheless estimate the energy that must be radiated, assuming that the source lies at distances typical of various celestial objects.

For even moderate source distances the energy released is enormous. On the basis of a typical gamma-ray flux at the satellite detector of 10^{-4} erg per square centimeter per second, the energy released at a source located at the distance of the nearest stars (a little more than one parsec, or 3.26 light-years) would be 10^{34} ergs per second. If the source were located 1,000 parsecs away, which would still be only a tenth of the distance from the sun to the center of the galaxy, the energy released would be 10^{40} ergs per second. At that rate the energy of the X rays and gamma rays emitted by the source would be some 2.5 million times greater than the energy radiated by the sun at all wavelengths. Considering that the source is smaller than the earth, it would have to be radiating at least 10 billion times more energy than the sun per unit of surface. If the sources are in other galaxies, which has been suggested, the foregoing values will have to be increased by at least six orders of magnitude.

The gamma-ray bursts show no directional correlation with nearby stars or with any nearby stars known to have peculiar characteristics, so that we are thrown back on statistical arguments for inferring distances. One approach is simply to study the distribution of intensities of the gamma-ray bursts so far recorded. As long as the sources are on the average distributed evenly throughout three-dimensional space, the intensity distribution will follow a simple law. The number of sources more intense than any value S is proportional to $S^{-\alpha}$, where α is 3/2. (This relation reflects the fact that whereas the number of sources goes up as the cube of the distance, their apparent brightness falls off as the distance squared.) The law holds even when directions are not known, and it is not affected by nonsystematic variations with distance in the intrinsic brightness of the collection of objects involved.

We know, however, that our galaxy is very thin compared with its diameter. Therefore if the sources of gamma-ray bursts follow the same distribution as stars of a known class, the number of sources will not increase in proportion to the volume when the distances begin to exceed the thickness of the galaxy. The value of α will in fact tend toward unity. A difficulty with this simple relation is that as the value of S, the observed intensity, is set lower some of the events become too weak to be detected. Below some value of S one will miss all the events. When one plots the value of S against the number of sources observed, the value of α will appear to drop below 3/2 and will eventually reach zero, not because of the distribution of the sources but because of our inability to detect the weak ones. Unfortunately this effect sets in with the Vela data rather quickly, and we have not been able to draw any definite conclusion from the study.

A second approach is to study the few directions we have, looking for clues to the distribution of the sources in space, from which we can then estimate the distances. When we plot the locations of the seven events for which we have unique positions, we find a reasonably random distribution. If the sources are fairly close to the sun, say within 100 parsecs, this is the kind of distribution one would expect to find. If they are 10 to 20 times farther away, one would expect to see them beginning to cluster along the galactic equator (the Milky Way), following the pattern of the stars. If they are at distances of 10,000 parsecs or more, we would be detecting the majority of the sources in the galaxy, and the asymmetry, with most of them lying toward the galactic center, would be obvious. No such asymmetry is observed.

For distances of more than 10,000 parsecs the sources would be outside the galaxy, and one would again expect to find a random distribution, unless the sources were associated with neighboring galaxies, in which case we should be able to see a correlation with those galaxies. We do have some directional information on a total of 16 sources of gamma-ray bursts, including the seven unique directions we have mentioned. None of them corresponds even roughly to the location of any of the galaxies in the local cluster, the cluster to which our galaxy belongs. We can therefore be sure that if the sources are extragalactic, they are at distances greater than four million parsecs.

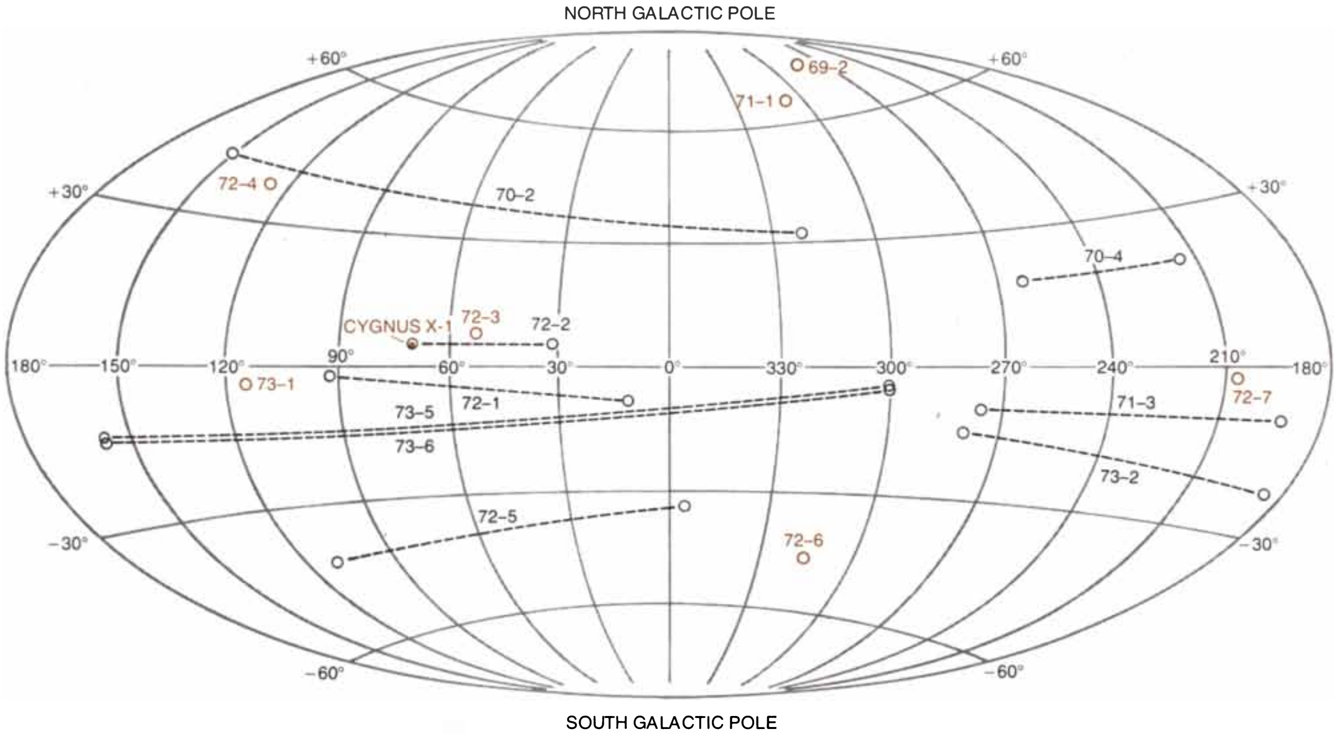
Although we have only seven unique directions, we have pairs of alternative positions for nine more sources. Because of the orientation of the orbital planes of the observing satellites the alternative points are at approximately the same galactic latitude. Thus al-

	UNITS	RANGE	TYPICAL
TOTAL ENERGY AT EARTH	ERGS PER SQUARE CENTIMETER	3×10^{-6} TO 1.5×10^{-3}	10^{-4}
NUMBER OF DISTINCT PULSES		1 TO MORE THAN 5	2
DURATION OF FIRST PULSE	SECONDS	< .01 TO 10	3 TO 5
TOTAL DURATION OF BURST	SECONDS	< 0.1 TO ~ 100	10
RISE TIME OF FIRST PULSE	SECONDS	.002 TO 3	.2
PEAK INTENSITY	ERGS PER SQUARE CENTIMETER PER SECOND	10^{-4} TO 5×10^{-4}	

CONSIDERABLE VARIABILITY is exhibited by the 50-odd gamma-ray bursts recorded since the first event in 1967. The table shows ranges and typical values for the major parameters. The figures given for the total energy flux at the earth assume that the measurement is made outside the earth's atmosphere and that the energy spectrum is similar to the one shown on preceding page. The narrow range in peak intensity, which varies by a factor of only about five, is probably due to the limited ability of present detectors to record the weaker bursts.

DISTANCE (PARSECS)	ENERGY PER BURST (ERGS PER SECOND)	POSSIBLE SOURCE
1	10^{34}	NEAREST STARS
10	10^{36}	X-RAY FLARE STARS (MINIMUM)
10^2	10^{38}	MAXIMUM OF GALACTIC-DISK STARS (POPULATION I STARS)
10^3	10^{40}	GALACTIC HALO STARS (POPULATION II STARS)
10^4	10^{42}	GLOBULAR CLUSTERS
10^5	10^{44}	NEAREST EXTRAGALACTIC OBJECTS
10^6	10^{46}	LOCAL GROUP OF GALAXIES
10^7	10^{48}	CLUSTERS OF GALAXIES

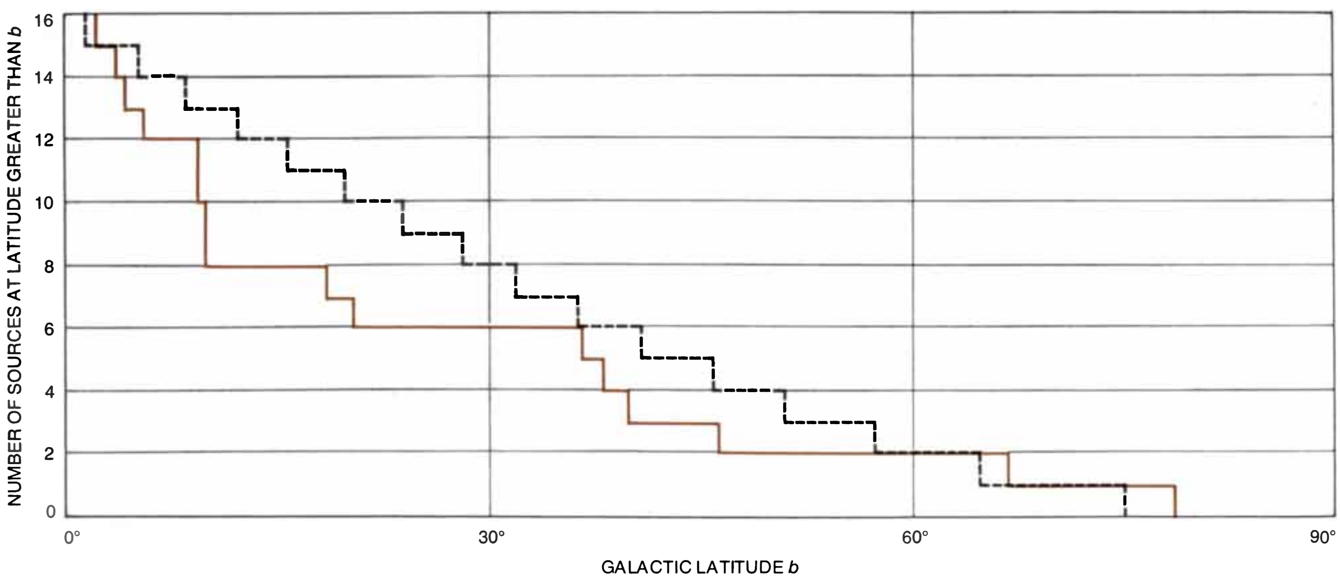
TOTAL ENERGY EMITTED at the source in a gamma-ray burst depends on its distance, which so far has not been determined for any of the puzzling events. Assuming that the flux of a gamma-ray burst measured near the earth is about 10^{-4} erg per square centimeter per second, and that the source radiates energy uniformly in all directions, the total flux at the source can vary from about 10^{34} ergs per second (if the source is at the distance of one of the nearer stars) to more than 10^{48} ergs per second (if the source lies somewhere outside our local group of galaxies). For comparison sun radiates 4×10^{33} ergs per second. A parsec is 3.26 light-years.



POSITIONS OF GAMMA-RAY SOURCES have been difficult to determine uniquely and with high accuracy. As explained in the illustrations on page 68, if a gamma-ray burst is detected by three satellites, the position of the source can be assigned to either of two points where two circles intersect. The open circles show such alternative positions for nine events, plotted in galactic coordinates. (The first two digits indicate the year of the event; the third digit denotes the order of discovery during the year.) In seven cases, marked by colored circles, unique positions are known because it has been possible to rule out one of the alternative positions since it was hidden behind the earth from a fourth near-earth satellite. One of the alternative positions for event 72–2 includes known location of Cygnus X-1.

though we do not know the galactic longitudes, we can estimate the latitudes satisfactorily by averaging the alternative values. The latitude distribution can then be compared with what one would expect for a random distribution. Such a comparison reveals a noticeable concentration toward the galactic equator, but the size of the sample is still too small for a firm conclusion to be drawn.

If we assume that the sources of gamma-ray bursts do cluster near the galactic equator, we should expect the source distances to have values ranging between a few hundred and a few thousand parsecs, depending to a large extent on how thick the disk of sources is. There is no agreement among astrophysicists on whether the present evidence is convincing or not. If one believes the points are clustering near the



SOURCES OF GAMMA-RAY BURSTS seem at first to be rather randomly scattered in the sky. One simple test for possible clustering is to ignore longitudes, for which information is sparse in any case, and to plot a curve showing how many bursts are located at more than a given number of degrees, *b*, from the galactic equator. The colored step curve shows the distribution of all 16 events, using average latitudes for the nine events with alternative positions. The black curve shows the distribution that would be expected if the events were isotropically scattered, that is, if the events exhibited no preferred direction in space. Although the sample of positions is too small to be conclusive, a comparison of the two steplike curves suggests that the gamma-ray bursts indeed cluster near galactic equator.

equator, then 1,000 parsecs is a typical distance. If one does not, then the distances are either about 100 parsecs or at least four million parsecs.

Models of the Sources

Essentially all the explanations for gamma-ray bursts are based on the limited number of facts we have presented here. A few models even manage to neglect or contradict some of the facts. In a recent article Malvin A. Ruderman of Columbia University put the problem nicely. "Most theoretical astrophysicists," he wrote, "function well in only one or two normal modes. Therefore we often twist rather strenuously to convince ourselves and others that observations of a new phenomenon fit into our chosen specialties."

It would not be possible to describe all the models of gamma-ray bursts in detail, if only because many of the authors have not published the details. We shall describe only those models where at least some of the properties of the bursts are predicted.

We have shown that extragalactic models involve source distances of at least four million parsecs, or 13 million light-years, and hence imply a minimum total radiated energy of 10^{47} ergs. Historically the first proposed explanation was a model, conceived by Stirling Colgate of the New Mexico Institute of Mining and Technology, in which the shock wave from a supernova is expanding at relativistic velocities, that is, velocities close to the velocity of light. As the core of the star collapses, initiating the supernova process, a strong spherical shock wave is set up, traveling outward at ever increasing speed as the density of the star decreases. At the surface of the star the speed is relativistic, and the X rays emitted from the hot shock front are Doppler-shifted to form a pulse of soft gamma rays. This mechanism, the original Colgate one, has stimulated much observational and theoretical work on transient stellar outbursts, but it does not seem able to explain cosmic gamma-ray bursts. The total energy emitted is probably too low and the fine structure of the bursts is unexplained.

A second model, proposed by J. Cohen of the University of Pennsylvania and Reuven Ramaty of the Goddard Space Flight Center, involves the collapse of a white-dwarf star to a neutron star. It is also an extragalactic model, because although such a collapse might occur in any galaxy, the frequency of occurrence in our own galaxy would be much too low to account for the known number of gamma-ray bursts. It is a plausible argument that there exist massive white-dwarf stars that slowly accrete matter from their surroundings. When the gravitational force on the core of the star exceeds the ability of matter to withstand it, the star rapidly collapses, giving rise to an object with the density typical of a neutron star (roughly 10^{14} grams per cubic centimeter). The surface temperature of the collapsing star might easily reach 10^9 degrees Kelvin and emit a flash of soft gamma radiation. One can only speculate on the sequence of events that would ensue. It might consist of a succession of collapses of a shell followed by some kind of rebound, but again it is difficult to see how the model could account for the nearly random peaks common in the records of gamma-ray bursts.

If the present weak evidence for the location of the sources of gamma-ray bursts within our galaxy is strengthened, both of these extragalactic models must be rejected. Moreover, Ruderman has pointed out that with certain possible exceptions it is very difficult to explain a luminosity greater than that of an extremely hot black-body radiator with a temperature related to the average gamma-ray photon energy, which would be about a billion degrees K. Only an object with a larger surface area could have a greater luminosity. If the diameter of gamma-ray sources is less than 300 kilometers, it follows that there is an upper limit to the absolute luminosity; it is about 2×10^{47} ergs per second. This puts the maximum distance of the sources at about four million parsecs. Since no known source has a location corresponding to any galaxy within four million parsecs, all extragalactic models are beginning to seem unlikely (unless, of course, the luminosity restriction can be lifted). The models are nonetheless important because they may predict radiation phenomena that have not yet been discovered.

Galactic Models

We now turn to the large class of galactic models of sources of gamma-ray bursts. Superficial similarities between gamma-ray bursts and solar flares, which can emit X-radiation, inspired a number of flare models. In one model, proposed by Floyd W. Stecker and Kenneth J. Frost of the Goddard Space Flight Center, the putative flares occur on white-dwarf stars with a strong magnetic field. In a model suggested by Kenneth Brecher and Philip Morrison of the Massachusetts Institute of Technology the gamma-ray bursts would be generated when low-energy photons were boosted in energy by extremely energetic electrons in flares from a commoner type of star. Other hypothetical sources of gamma-ray bursts include encounters between matter and antimatter and between relativistic dust particles and photons.

When we turn to neutron stars, which may have a magnetic field of a trillion gauss, it is quite easy to find several possible sources for the energy in a gamma-ray burst. It is believed the orbital-spinning pulsars, which emit radio pulses at precisely spaced intervals ranging from a fraction of a second to several seconds, are neutron stars. The magnetic-field energy of a neutron star would be much greater than the burst energy for sources located at moderate distances, although how it could be tapped is quite unclear. "Glitches," or sudden changes in the rotational speed of a neutron star, are known to occur and could perhaps account for the energy released. Franco Pacini of the European Southern Observatory and Ruderman have shown that the energies of gamma-ray bursts agree in magnitude with the energy lost by a neutron star during a glitch, but it is not easy to explain exactly how the rotational energy would be transformed into gamma radiation. In still other neutron-star models the source of the gamma rays is ascribed to "starquakes," volcanic activity or other sudden changes in size or shape. Unfortunately for admirers of this class of models, none of the known locations of gamma-ray bursts coincides with that of a known pulsar, but there are now thought to be a large number of neutron stars in addition to the couple of hundred so far identified through their emission of radio pulses.

The class of models that probably appeals most to astrophysicists, including ourselves, is one in which matter from somewhere, usually from a normal star, is attracted toward the surface of a neutron star or of a white dwarf or into a black hole. Black holes are the still-hypothetical points where a spent star has collapsed to such an enormous density that no radiation can escape but where infalling matter can emit energetic photons before disappearing from view. One hydrogen atom falling onto a white dwarf can release a million electron volts (MeV), and one falling onto a neutron star can release as much as 100 MeV; a similar amount of energy might be released near a black hole. The broad range of behavior exhibited by X-ray stars has been successfully explained by models in which the star accretes matter. The binary X-ray sources have been understood on this basis in remarkable detail [see "X-Ray-emitting Double Stars," by Herbert Gursky and Edward P. J. van den Heuvel; SCIENTIFIC AMERICAN, March, 1975].

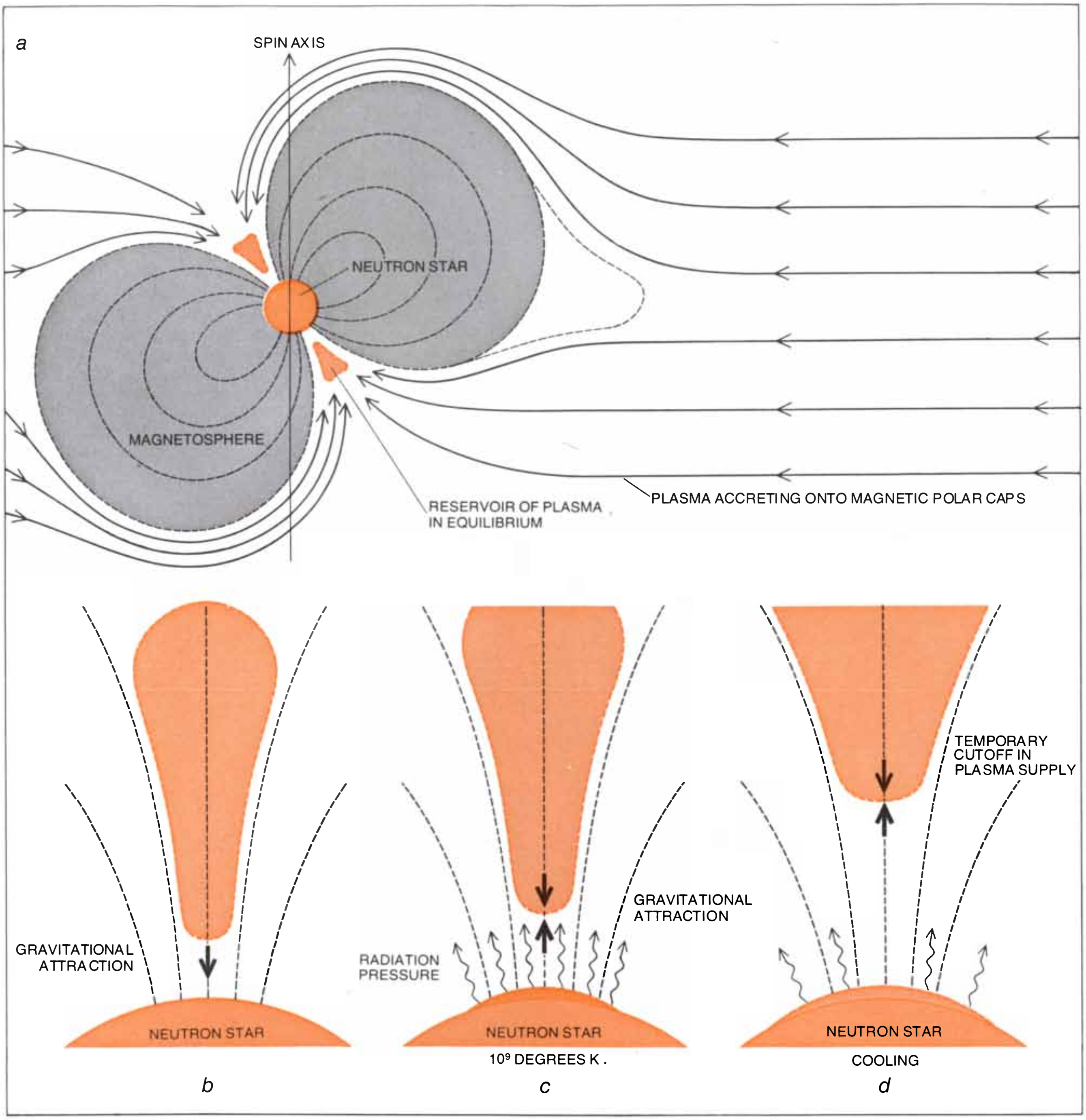
One possible way to obtain gamma-ray bursts from the intermittent accretion of matter by a neutron star has been proposed by us. Briefly, the idea is that blobs of plasma (a gas of electrons and ions) are balanced in equilibrium over the magnetic poles of the neutron star. When a glitch occurs, the blobs accelerate downward, releasing gamma rays in the process.

Cygnus X-1: A Possible Source

For some two years we have been tantalized by the possibility that certain gamma-ray bursts might originate in Cygnus X-1, a powerful X-ray source and the binary system that is believed the most likely one to include a black hole. A gamma-ray burst observed on April 12, 1972, has two possible locations in the sky, each of them within an area of about 20 square degrees. Cygnus X-1 lies in the middle of one of them. An area of 20 square degrees, we hasten to point out, is so large that the inclusion of Cygnus X-1 could be no more than a coincidence.

Another burst, with a similar energy flux and duration, on March 15, 1971, has given us many headaches. Data from two Vela satellites combined with some from OGO-5 give us a large possible area for the source that also contains Cygnus X-1, but the four Vela-satellite times are not compatible with one another. We must adjust two of them by about a second to get any direction at all from the data. If we have a temporary timing error of exactly one second on two of them, we would get two good locations, one of which again contains Cygnus X-1. The problem has not yet been resolved.

A better reason for associating Cygnus X-1 with these bursts is that its usual X-radiation drastically changed its character in the middle of March, 1971;



NEUTRON-STAR-"GLITCH" MODEL invokes sudden changes in magnetic-field strength or in rotation rate (glitches) of neutron stars to explain gamma-ray bursts. Neutron stars are thought to be supernova remnants: rapidly rotating objects as massive as the sun, less than 50 or so kilometers in diameter and possessing extremely strong magnetic fields. Presumably some plasma in the vicinity of such stars would collect at the magnetic polar caps, forming a reservoir of charged particles held in equilibrium by a combination of gravitational attraction, radiation pressure and centrifugal force (*a*). A glitch would upset the equilibrium and allow a blob of plasma to plunge onto the polar caps (*b*), releasing as much as 100 MeV per particle and creating radiation with a temperature equivalent of a billion degrees (*c*). Pressure of radiation drives back plasma, cutting off radiation (*d*). Process can repeat until the plasma reservoir is empty.

we have since found two other bursts during the rare occasions when Cygnus X-1's X-radiation changed. One might say that the evidence is sufficient to indict but not to convict.

The consequences of the association between gamma-ray bursts and Cygnus X-1, if it is proved, are fascinating. Not only does Cygnus X-1 possibly incorporate a black hole but also its distance (2,600 parsecs) is known with some accuracy and the object has been studied in great detail for years. On the basis of its known distance the bursts would have an energy output at the source of about 10^{40} ergs per second, or two million times the sun's output at all wavelengths. The gamma-ray emission would also have about 100 times the energy of the usual X-ray output of Cygnus X-1. The gamma-ray flux is also well in excess of the object's Eddington limit: the amount of steady radiation that would blow away the surface of a star if it were to persist.

It is tempting to suggest plausible mechanisms for gamma-ray bursts based on the possible Cygnus X-1 connection, and we, among others, have not succeeded in resisting that temptation. One model that we find plausible involves interactions between two infalling streams of plasma near a black hole. The interaction generates gamma rays [*see illustration on page 67*].

Some of the constraints on such models should be noted. They have to explain how the bursts are related to the two (at least) X-ray states of Cygnus X-1 and to the characteristic duration and luminosity of the bursts, and particularly why bursts occur during the transition between states. If Cygnus X-1 is indeed found to incorporate a black hole, the possible association with gamma-ray bursts raises the intriguing possibility that gamma-ray bursts are a characteristic feature of black holes, and might even be a useful way to find them.

Many of the models we have mentioned here are better characterized by the term scenario, which has recently become popular with astrophysicists. The implication is that there is so little detailed information that the proposals should not be dignified by the term model. Nevertheless, a good scenario can sometimes lead to a good model.

We now look forward to the first results from combining the data from three new satellites: *Helios-2,* which carries the first detectors specially designed to observe gamma-ray bursts, and a pair of Naval Research Laboratory satellites named SOLRAD, for which Los Alamos has supplied two gamma-ray detectors. With luck we shall be able to locate gamma-ray bursts with sufficient accuracy to identify them with visible celestial objects. When that happens, we can expect to make the kind of rapid progress that has been made toward the understanding of X-ray stars.

GM
MARK OF EXCELLENCE

A demonstration of Buick Opel's passing ability.

It's no trick at all to find a car that can pass other cars in a flash. As long as you adopt a rather devil-may-care attitude regarding gas mileage.

And finding a car that can pass gas stations right and left isn't all that difficult either. As long as you can get used to appointments and performance that are a bit on the sparse side.

Ah, but the Buick Opel is different. Because it wasn't designed to be simply a great little performer.

Nor was it designed to be strictly an economy car. It was designed to be both. And, to provide a nice comfortable passenger compartment in which to enjoy them both.

It comes standard with reclining bucket seats. Flow-through ventilation. Tinted glass. It's available with options like air conditioning. Automatic transmission. AM/FM.

It accelerates easily to highway speeds thanks to a lusty little overhead-cam hemi engine. Coupled with a short-throw four-speed manual transmission.

And while your mileage may vary depending on how and where you drive, the condition of your car and how it's equipped, that combination of engine and transmission got an estimated 36 miles per gallon in the EPA highway test. And 23 in the city test.

In California, EPA estimates are lower.

The comfortable, fun Buick Opel. A quick little car that uses gas slow.

Priced at $3,282.20.

Manufacturers Suggested Retail Price including preparation, (White Walls shown, $130.00 extra, body side stripe $31.00 extra). Tax, license, destination charges and other available equipment additional.

At your Buick-Opel Dealers.

BUICK OPEL

White-Light Holograms

Several ways have been devised for viewing holograms with ordinary incoherent light, making it unnecessary to rely on the coherent light of the laser or even the quasi-coherent light of a mercury-arc lamp

by Emmett N. Leith

The technique of holography, or photography by wave-front reconstruction, is now nearly 30 years old, having been invented by Dennis Gabor in 1947 as a possible means of improving the resolving power of the electron microscope. It was not until the early 1960's, however, that the holographic technique began to attract widespread attention. At that time Juris Upatnieks and I, working at the University of Michigan, introduced a number of innovations that made it possible to extend the Gabor principle well beyond its initial application. For example, by the use of the off-axis reference-beam method we were able to make holographic images of a far higher quality than had previously been achieved. In addition, by exploiting the intense, highly coherent light of the laser we were able for the first time to make startlingly realistic holographic images of various types of three-dimensional reflecting object [see "Photography by Laser," by Emmett N. Leith and Juris Upatnieks; SCIENTIFIC AMERICAN, June, 1965]. The subsequent development of holography has been closely linked with that of the laser, as the demand for bigger and better holograms of bigger and richer scenes has tended to place an increasingly stringent coherence requirement on the light source employed to create these fascinating images.

At the same time a countervailing desire to make holography more practical and hence more universal has spurred the search for ways to reduce the coherence requirement of the process. Excellent progress has been made in the past decade toward both goals. Indeed, recent public exhibits have included holograms that can be viewed not only with the aid of coherent light from a laser but also with incoherent, or white, light derived from the sun, say, or from an ordinary incandescent bulb. Many people viewing these holograms have expressed surprise that holograms, which had long been associated with the sophisticated and expensive technology of the laser, could be viewed with such common light sources.

In this article I shall explain why coherent light has played such an important role in holography; I shall then tell how the coherence requirement differs for different kinds of holograms, and finally I shall describe a number of advances that have contributed to the substantial reduction in the coherence requirement for both making and viewing holograms.

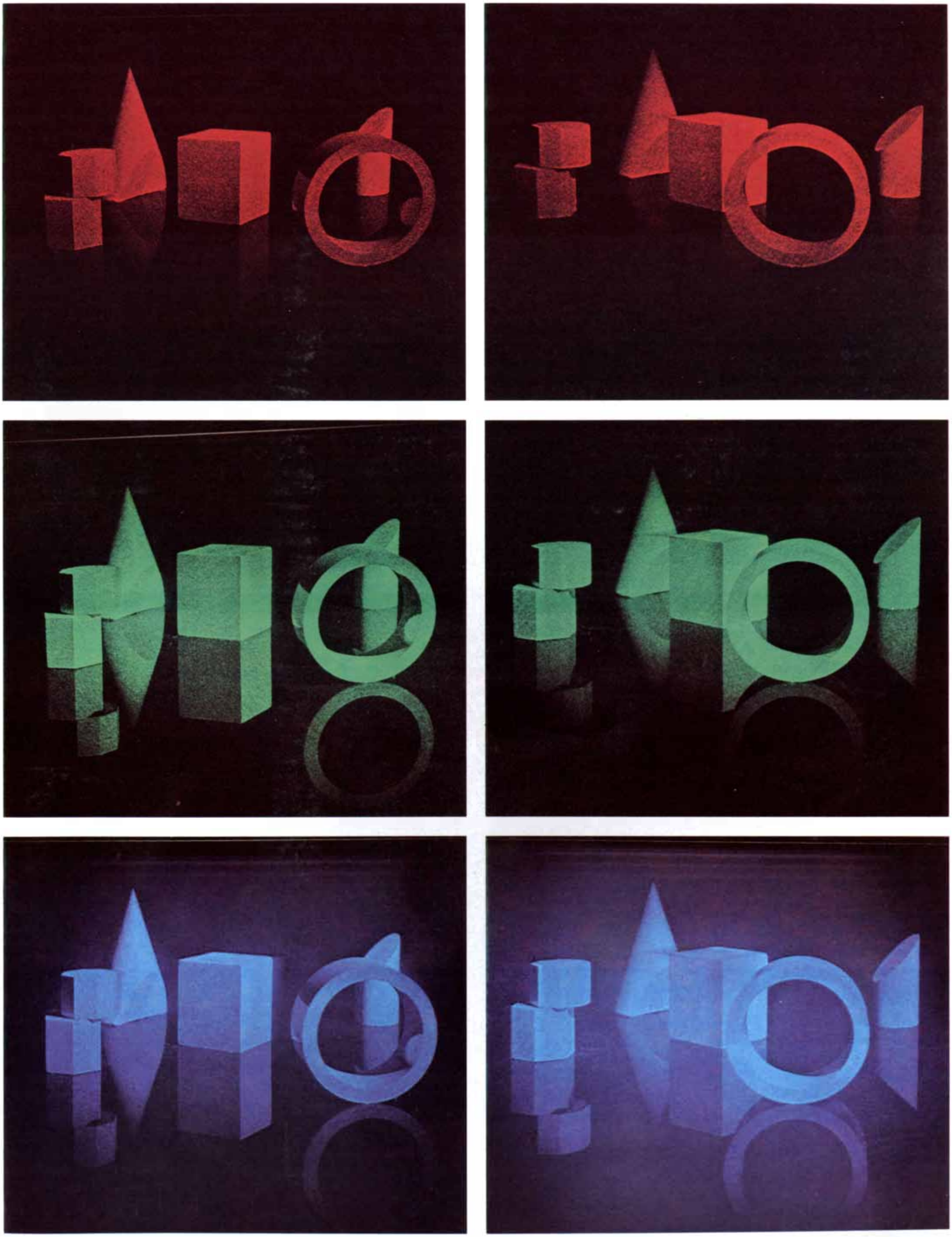
In its best-known form holography is a comparatively straightforward process. Coherent light from a laser is split into two beams. One beam is used to illuminate an object, and the portion of the beam that is reflected from the object falls on a photographic plate. The other beam is aimed directly at the plate by means of a mirror. The light from the mirror, called the reference beam, combines at the plate with the light reflected from the object to form a complex interference pattern. The developed plate, which records that interference pattern, is the hologram.

When a hologram of this type is illuminated with the reference beam alone, the light rays passing through the plate are selectively transmitted or absorbed in such a way as to create in the emerging beam a component that exactly duplicates the original light waves that were reflected from the object to the plate when the hologram was formed. An observer receiving these waves will perceive them as emanating from the original object; in effect he will "see" the object as if it were actually there [*see illustration on page 85*].

The coherence requirement arises because the hologram is a record of the interference of two waves, and in general (but not always) coherent light produces interference patterns and incoherent light does not. This requirement has definitely inhibited the development of holography, particularly for display purposes, since coherent-light sources are not only more expensive but also less convenient. Any reduction in the coherence requirement, particularly in the viewing stage, would therefore have considerable significance for holographic displays.

What is the difference between coherent and incoherent light? There are two types of coherence, temporal coherence and spatial coherence, and light is said to be coherent when it possesses both types. Temporally coherent light is light that is monochromatic: it has only one wavelength. Spatially coherent light is light that is derived from a point source or is capable of being focused to a point. Today the laser is the usual source of coherent light. Coherent light can also be obtained from incoherent sources, however, as it was routinely before the laser was invented. Spatial coherence can be achieved, for example, by placing a pinhole in front of the source, so that the light comes from a point. Temporal coherence can be achieved by placing a color filter in front of the source, so that only light in a narrow spectral band is transmitted. Each process involves throwing away by far the greater part of the light; hence an extremely intense source of incoherent light is needed to generate a very small amount of coherent light.

"RAINBOW" HOLOGRAM made by Stephen Benton of the Polaroid Corporation and Fritz Goro appears in the photographs on the opposite page. Although the hologram is shown in three different colors, it was photographed in transmitted white light obtained from a microscope illuminator with a tungsten-filament incandescent lamp. The three different colors result from the fact that the color of the hologram changes as the camera (or the eye) is moved up or down, and the hologram was photographed with the camera in three different vertical positions. When the camera is moved from side to side, a characteristic feature of all holograms can be seen. It is parallax, the apparent displacement of objects in the three-dimensional scene when the hologram is viewed from different angles. Such displacement is visible in each of these three pairs of photographs. A diagram showing how hologram was made appears on page 94.



DENISYUK WHITE-LIGHT HOLOGRAM, made by the Russian investigator Yu. N. Denisyuk, was photographed from two different vertical angles to demonstrate not only the vertical parallax achieved with this method but also the abrupt change in the color of the perceived image when it is seen from two different directions. Holograms of the Denisyuk type are viewed in reflected light, obtained from a pointlike source, in this case a small incandescent spotlight (*see lower illustration on page 93*). Object in hologram is a medallion with relief.

Before the laser the mercury-vapor arc lamp was one of the best available sources of coherent light. The arc is intense and typically only a few millimeters in diameter. Moreover, the light is emitted in several narrow spectral bands rather than in a continuous spectrum. Thus the techniques employed to obtain coherent light from a mercury-arc source result in less light loss than if the source were, say, an incandescent tungsten filament.

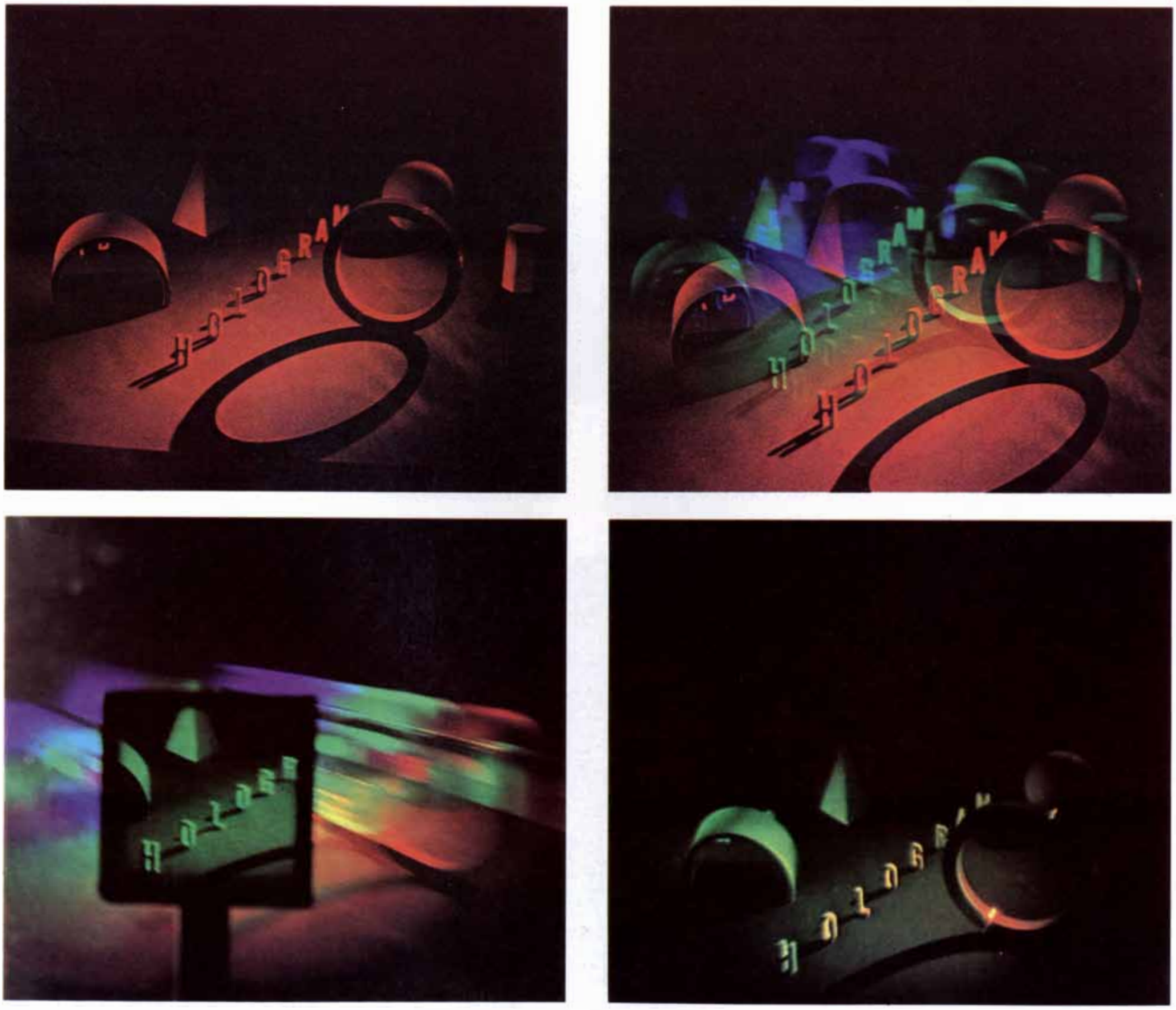
Gabor's original holograms were made and viewed with a mercury-arc source. The objects were rather simple: transparencies with opaque lettering and without a gray scale. A portion of the incident light was passed through the comparatively broad open areas and served as the reference beam. Light in the vicinity of the dark lines that formed the letters was diffracted, or scattered, and thereby spread out, so that the letters were not recognizable when they were viewed at the plane where the hologram was recorded. Such holograms do not seem impressive by present standards, but they did constitute a convincing demonstration of a new and exciting principle in optics. In 1971 Gabor was awarded the Nobel prize for physics in recognition of his invention.

The coherence requirement for Gabor's early holograms was not very great. In general the coherence requirement for both forming and viewing a hologram can be calculated from quite elementary considerations. The light scattered from each resolution element (the smallest resolvable point on the object) interferes with the unscattered reference beam to give rise to a "bull's-eye" diffraction pattern consisting of concentric bright and dark "fringes" that become finer toward the edge. The light from each resolution element is spread into such a pattern; the patterns from nearby points overlap, forming a more complex pattern. The hologram is an overlay of many such elementary patterns, one for each resolution element of the object.

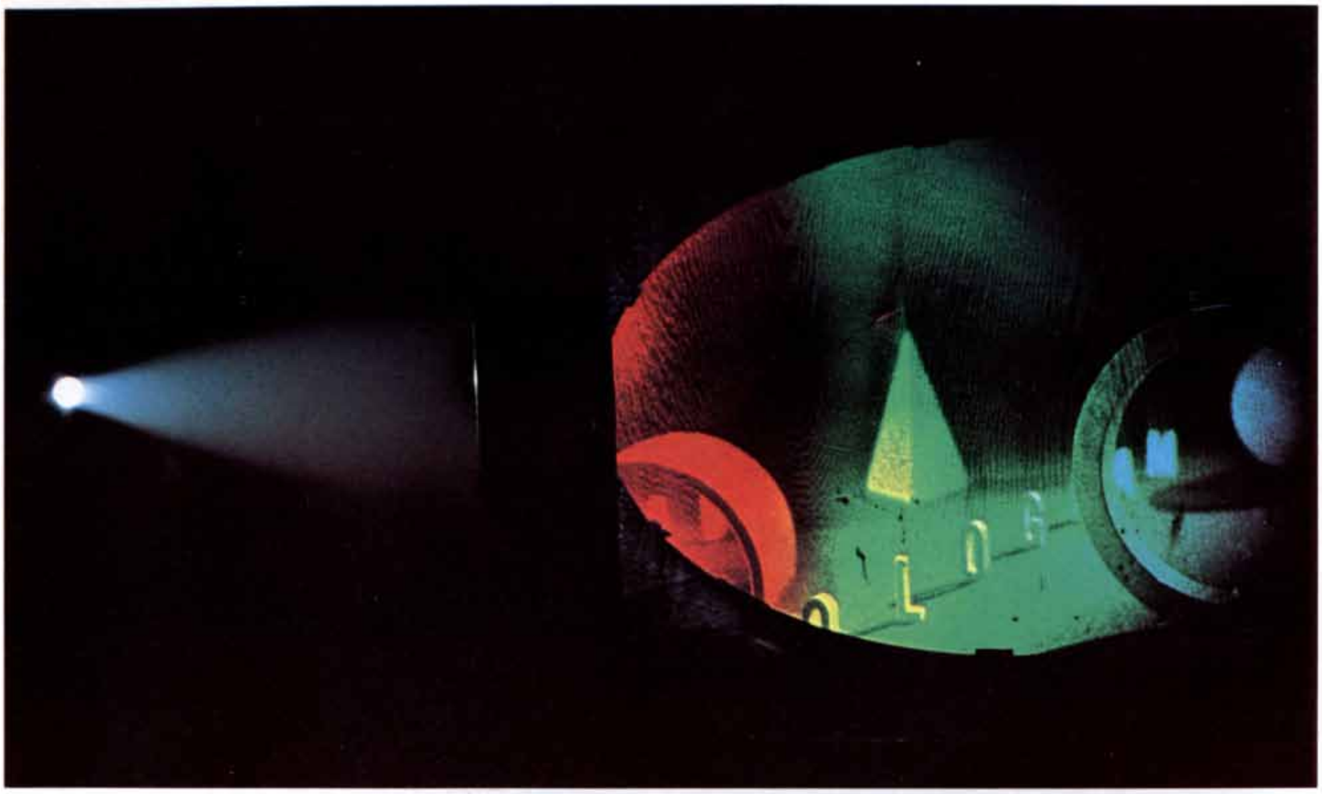
The scale of this bull's-eye pattern is proportional to the wavelength of the light. If the light is polychromatic, that is, if it has more than one wavelength, one can think of each component wavelength as separately forming its own set of diffraction patterns, one from each resolution element of the object [*see upper illustration on page 88*]. Because the patterns for the different wavelengths differ in size, a beam with a wide band of wavelengths will blur the resulting pattern, with the finest lines being blurred the most. The acceptable spectral bandwidth becomes narrower as the distance between the object and the hologram is increased or as the size of the resolution element is decreased. The ratio of the size of the diffraction pattern to the size of the resolution element is sometimes called the expansion ratio, since each resolution element is in effect spread out by that ratio to form the hologram. It can be shown that to make this kind of blurring negligible the spectral bandwidth should be less than four times the average wavelength divided by the expansion ratio.

For example, if one takes as an object a transparency whose finest detail is .01 millimeter and illuminates it with a beam of light that has a wavelength band centered at 5,000 angstroms, and one chooses an object-to-hologram separation so as to get an expansion ratio of 50, one finds that a wavelength spread of 400 angstroms is acceptable. That is an eighth of the entire visible spectrum, and light covering such a bandwidth could hardly be called coherent. A typical mercury-arc lamp radiates a strong green line one angstrom or two angstroms wide, and thus it has more than enough temporal coherence to make holograms.

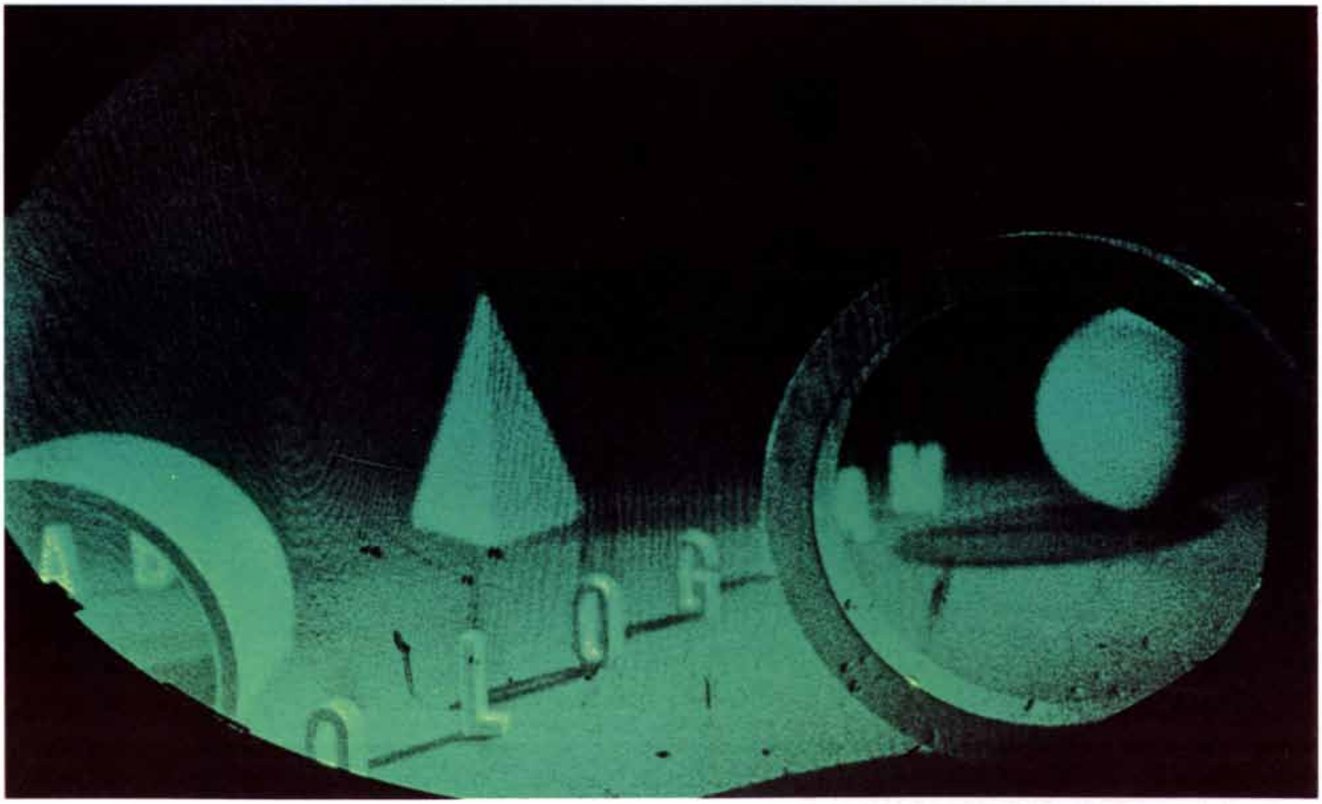
Alternatively, one can describe the temporal-coherence requirement in terms of the coherence length, a basic quantity that is essentially the distance over which the vibration frequency of a



CONVENTIONAL HOLOGRAM made by Goro in the author's laboratory at the University of Michigan was photographed in transmitted light from four different light sources in order to illustrate the importance of the coherence requirement in the traditional holographic process. The photograph at the top left shows the hologram viewed in the highly coherent red light generated by a helium-neon laser (the same light source originally used to make the hologram). The photograph at the top right shows the hologram viewed in the less coherent, multicolored light emitted at several discrete wavelengths by two different light sources; the red image is formed by light from the helium-neon laser and the blue, green and amber images are formed by light from a mercury-arc lamp. The displacement of the images corresponding to the different wavelengths arises from the fact that this particular type of hologram, called an off-axis or carrier-frequency hologram, acts like a diffraction grating, bending the light in proportion to its wavelength and hence dispersing the incident light beam into its component colors. The photograph at the bottom left is a double exposure; it was made by first masking a small portion of the hologram and then photographing the image formed by the rest of the hologram in the white light emitted by a zirconium-arc lamp. The blur of colors arises because each of the wavelengths in the continuous spectrum of the white light forms its own image, and the images are all displaced slightly with respect to one another. The masked portion of the hologram was next replaced by a narrow-band green filter, and the rest of the hologram was masked in turn. The second exposure, made with the aid of the same zirconium-arc white-light source, reveals a sharp green image of the original scene. The photograph at the bottom right shows the hologram again illuminated with the zirconium white-light source but with a special grating placed between the light source and the hologram so as to compensate for chromatic dispersion. Image is now reasonably sharp, particularly at center, where compensation is best.



RAINBOW HOLOGRAM OF SAME SCENE as the one that appears in the photographs of the conventional hologram on the preceding page was also made in the author's laboratory. When a hologram of this type is viewed in transmitted white light from a distance of about three feet, the resulting holographic image varies in color across the width of the hologram (hence the term rainbow hologram). White-light source in this case is an ordinary projector with a tungsten-filament lamp; smoke helps to make the light beam visible.



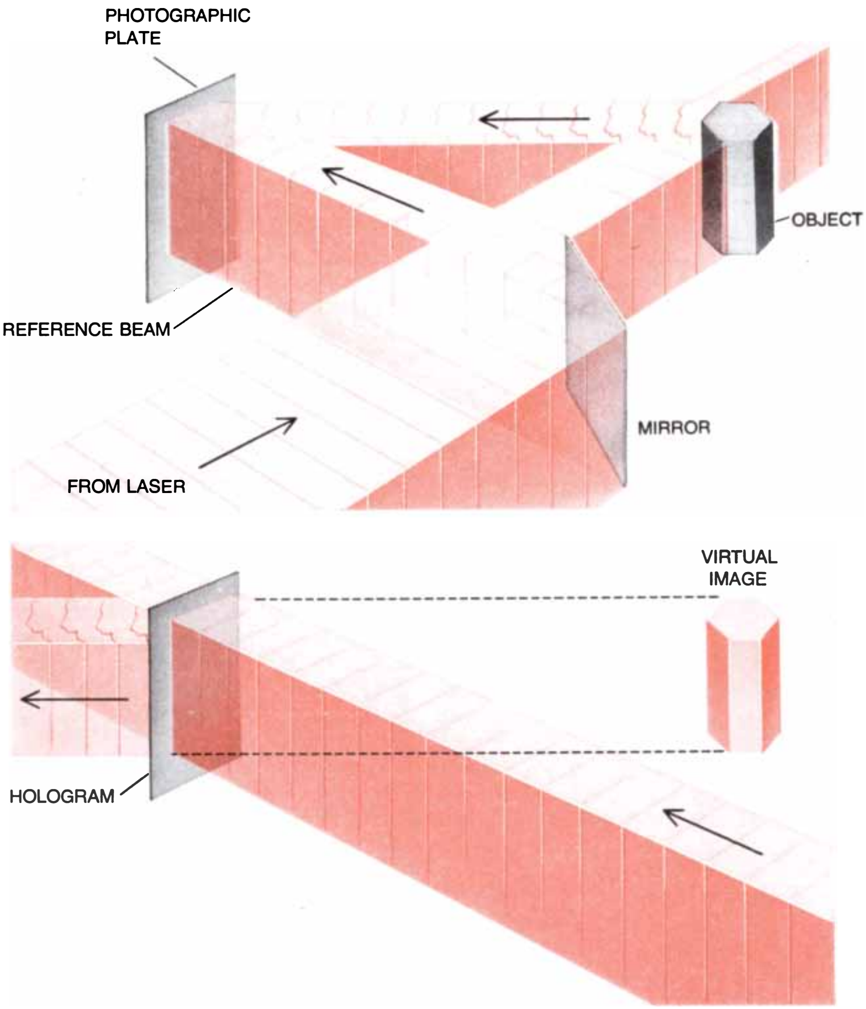
CLOSE-UP OF THE RAINBOW HOLOGRAM shown at the top of the page, photographed here with the same tungsten-filament white-light source but now from a distance of about 1½ feet, appears to be uniform in color across the entire image. If one were to shift one's viewing position laterally, however, the color would change but would remain uniform. The original scene that appears in the holograms pictured on this page and preceding one is 20 inches across; holograms in both cases are about 11 by 14 inches in size.

wave is accurately maintained; highly monochromatic waves have long coherence lengths, whereas polychromatic waves do not. Accordingly if one were to split a light beam into two parts and then recombine them downstream, interference effects would arise provided that the difference in the two optical paths did not exceed the coherence length of the light. Although this way of describing temporal coherence appears to be unrelated to the earlier one based on the misregistration of the interference fringes formed at different wavelengths, the two formulations are in fact equivalent.

The reconstruction, or viewing, stage of holography is essentially a retracing of the hologram-forming stage, and the same temporal-coherence considerations apply. When a beam of collimated, or parallel, light rays falls on the hologram, each elemental diffraction pattern appropriates a portion of the incident light, converting it into a spherical wave front that seems to emanate from the original object element; the image formed is called a virtual image. The recorded elemental diffraction pattern behaves much like a lens in that it focuses the incident light to a point image; indeed, the bull's-eye structure of the diffraction pattern is similar to that of a Fresnel-zone lens, a device often demonstrated in introductory optics courses. The size of the resolution element of the virtual image is exactly the same as the size of the resolution element of the original object, unless imperfections in the process have degraded the image. The ratio of the size of the diffraction pattern to the size of the resolution element of the image is called the compression ratio; it is of course exactly equal to the expansion ratio.

The need for monochromaticity in the viewing process arises because the curvature of the wave front, and hence the distance to the focal point, is proportional to the wavelength of the light. Therefore if light of many wavelengths is used, each wavelength forms an image at a different position. As a result in whatever plane one focuses the viewing device (the eye or a camera) one can see an image in focus and many other images, formed with other wavelengths, somewhat out of focus. This defect in the resulting image is called longitudinal dispersion. The spectral bandwidth of the illuminating source should be narrow enough so that such blurring effects are negligible. In fact, it can be shown that the spectral bandwidth should not be larger than the average wavelength times the compression ratio, which is exactly the criterion followed in the formation of the hologram.

The compression ratio (or the expansion ratio) is the only factor involved in determining the required temporal coherence. Even the size of the object is irrelevant; hence light of limited coherence in no way limits the size of the object that can be holographed.



BEST-KNOWN FORM OF HOLOGRAPHY, depicted in this schematic diagram, relies on coherent laser light for making the hologram. In the formation stage (*top*) a laser beam is split into two parts. One part is used to illuminate an object, and some of the reflected light falls on a photographic plate. The other part of the beam is aimed directly at the plate by means of a mirror. The plate records the complex pattern caused by the interference of the two beams. In the readout stage (*bottom*) the hologram is illuminated with the reference beam alone, creating in the emerging beam an exact duplicate of wave fronts reflected from the object. The reference beam can be produced by a laser or by a quasi-coherent source such as a mercury-arc lamp. The image one perceives by looking through the hologram is known as the virtual image.

Spatial coherence, the other aspect of coherence, is related to the size of the light source. An extended source can be regarded as composed of many separate point sources, each of which forms a diffraction pattern for each resolution element of the object [*see lower illustration on page 88*]. All these patterns are projected along the propagation direction of the incident light beam. Thus the different resolution elements of the light source form slightly displaced bull's-eye diffraction patterns, and if the source is large enough, the resulting diffraction pattern due to the entire source is blurred, particularly at the edges where the fringe spacings are finer. Obviously the light source employed to form a hologram should be small enough for this kind of smearing to be negligible. Similarly, in the viewing process an extended light source leads to image-smearing, and the spatial-coherence requirement is the same as in forming the hologram.

Our group at Michigan became involved in optical holography in 1960, when we attempted to improve the image quality, which was then rather poor. One basic problem was that each resolution element of the hologram forms two waves of equal but opposite curvature; the additional wave is convergent, forming an image in front of the hologram called the real image. One can view either the virtual image or the real image, but always with the other image out of focus in the background. This defect arises because in the recording process the sense of curvature of the wave front is lost. (A spherically diverging wave and a spherically converging wave would form the same recorded diffraction pattern.) The reconstruction process resolves this ambiguity by producing two waves, one converging and one diverging, but the quality of the image is thereby degraded. Other defects are also present, leading to additional components of extraneous light.

Upatnieks and I solved these problems by means of an approach better known in communication technology. Resorting to a technique somewhat akin to the modulation of a radio wave onto a carrier frequency, we caused a portion of the illuminating beam to bypass the object and to impinge on the recording plate obliquely. The interference fringes consequently became much finer and more numerous, and the hologram more nearly resembled an ordinary parallel-line diffraction grating. Just as a diffraction grating diffracts an incident beam of light into many separate beams (called orders) traveling in different directions, so our holograms caused diffraction in several directions. The beams forming the two images now propagated in two different directions, and therefore they became separated. In addition the various other extraneous beams propagated in still other directions. As a result each of the two images, real and virtual, was now free from extraneous effects and was of greatly improved quality.

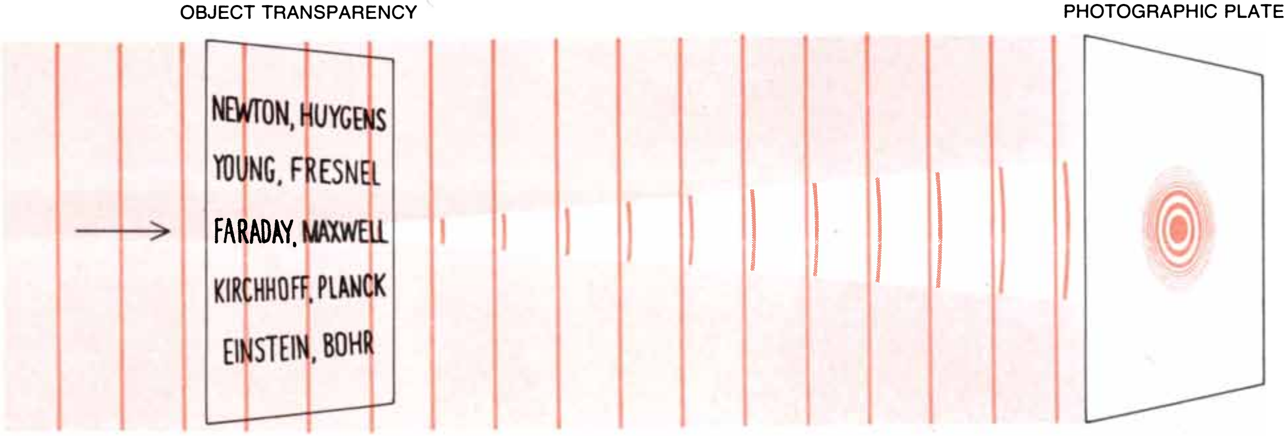
With many more fringes and a finer fringe spacing, one might suppose that a much greater coherence requirement would be placed on the light source used for making such a hologram. Furthermore, if the hologram is now like a diffraction grating, might not the hologram tend to disperse the incident light into its component colors, as diffraction gratings inherently do? (Diffraction gratings deviate the transmitted light in proportion to wavelength, so that within each order the red component, say, is deviated much more than the blue one.) It would therefore seem reasonable to expect that this type of hologram, generally called the carrier-frequency hologram or the off-axis hologram, would form for each wavelength an image slightly displaced laterally from the image formed by a slightly different wavelength [*see top illustration on page 92*]. Even for a fairly narrow band of wavelengths the image would as a result be hopelessly smeared. This type of imaging defect, called lateral dispersion, exists independently of the longitudinal dispersion described above. It would seem that the coherence requirement for our improved form of holography should be greater than that for the Gabor type, and that it perhaps could be met only with the light from a laser.

Both suppositions turn out to be wrong. Our first carrier-frequency holograms were actually made in the prelaser era and with nothing more than a conventional mercury-arc light source. Since the coherence requirement for basic holography is itself not very great, even a severalfold increase in this requirement is still within the capability of the mercury-arc source. Furthermore, the increase is not fundamental and can easily be avoided.

Suppose, for example, that in the reconstruction process one places between the source and the hologram a diffraction grating with a spacing equal to the average spacing of the fringes of the hologram, and suppose further that for illuminating the hologram one uses precisely that diffracted order which bends the light oppositely from the way the hologram does. The dispersions would then compensate. As viewed from the hologram, each wavelength component of the incident light would impinge from a slightly different direction, the angular displacement being just the right amount to cause the images formed by the various wavelengths to be exactly in register. If the adjustments are made properly, the coherence requirement can be reduced to exactly what it was in Gabor's method [*see middle illustration on page 92*].

The same principle applies in making the hologram. The problem with causing beams of polychromatic light to interfere is that the spacing of the interference fringes is proportional to the wavelength. Therefore if the fringes are aligned at one place in the overlapping beams, they become progressively misaligned as one moves away from that place; as a result an inadequate number of interference fringes are formed. By placing a grating between the source and the object, however, one can predisperse the light to make the spacing of the fringes the same for all wavelengths; the number of fringes that can be obtained then is independent of the wavelength. The grating compensates for lateral dispersion, thereby reducing the coherence requirement for both making and viewing the hologram back to the criterion of Gabor's original scheme for holography. The system described in our original paper on holography incorporated such a grating, and indeed it gave wavelength compensation, although that had not been our aim, and the simplicity of the object made such compensation unnecessary. Robert E. Brooks, L. O. Heflinger and Ralph F. Wuerker of TRW Inc. suggested and demonstrated similar methods of making pulsed-laser holograms, but at that time the pulsed laser lacked sufficient coherence for the task.

There are a number of other techniques that reduce the coherence requirement of the new and more advanced forms of holography back to that of the Gabor method, a circumstance suggesting that the original coherence requirement is a basic, irreducible minimum. This supposition, it turns out, is also untrue; in fact, one can reduce the coherence requirement even further. For example, in illuminating a Gabor hologram one could interpose a



ORIGINAL HOLOGRAPHIC METHOD, invented by Dennis Gabor in 1947, employed simple transparencies with opaque lettering as objects and a mercury-arc lamp as a light source. The diagram at the left shows how the holograms were made. Most of the light passed through the transparency and served as the reference beam. In the vicinity of the dark lines that formed the letters, however, a portion of the transmitted light was diffracted, or scattered, into a spherical wave front by every resolution element (or smallest resolvable point) on the transparency. The interference of the scattered wave fronts from each resolution element with the unscattered wave

Fresnel-zone plate between the hologram and the illuminating source. One could then take the real image formed by the zone plate as a source for illuminating the hologram. (The zone plate forms a longitudinally dispersed image of the source, with the violet, the shortest wavelength, being the most distant from the plate.) Such an arrangement compensates for the longitudinal dispersion of the hologram, causing the image for each wavelength to form at the same distance from the hologram [*see bottom illustration on page 92*]. It follows that such a hologram could be viewed in white light. A similar argument can be made for the hologram-forming process, which leads in turn to the suggestion that one might also be able to make holograms in completely white light. Indeed, if one were to use a zone plate and a grating in combination, one could simultaneously compensate for both lateral and longitudinal dispersion, thereby making it possible for high-quality, off-axis holograms to be both formed and viewed in completely white light.

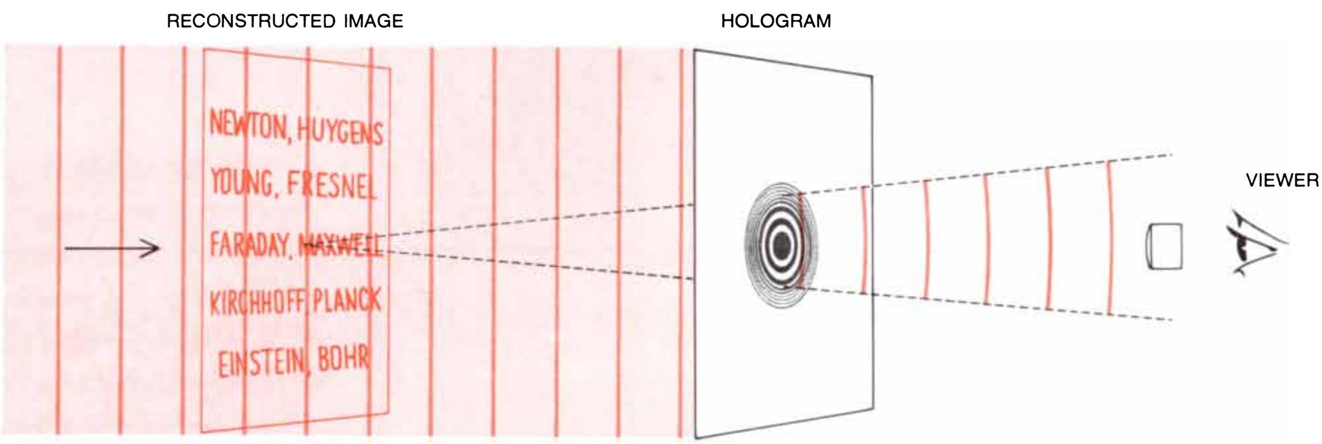
There is an interesting historical aspect to this suggestion. In 1955, while my colleagues and I at Michigan were working on radar imaging systems, it occurred to us that if we recorded radar data on photographic film or a similar medium and illuminated the resulting record with a beam of coherent light, we would regenerate miniature optical replicas of the microwaves that had originally impinged on the receiving aperture of the radar system. We could accordingly obtain microwave images in a simple and convenient way. From this viewpoint we developed a rather complete theory of holography, paralleling in many respects Gabor's earlier work, which was not known to us at that time. Our theory differed from Gabor's in several ways. It introduced for the first time the idea of off-axis holography as a scheme for avoiding the twin-image problem, and it also introduced the concept of using a grating and a Fresnel-zone plate in combination for correcting lateral and longitudinal dispersion, thereby allowing white-light operation. Thus although off-axis holography is often identified with the laser, this form of holography was in fact developed in the context of completely white light.

The zone-plate compensation does have one important shortcoming. Longitudinal dispersion is corrected only for a single image point, namely that point whose hologram component is aligned with the compensating Fresnel-zone plate. For display holography this limitation is inconvenient, since one generally wants to view the entire image, not just a single small region of it. For the radar application the defect presented no great problem, because one records the image a point at a time while moving the radar data record through an aperture, thereby bringing each image point in turn into position for dispersion correction.

This chapter in the story of holography closes somewhat anticlimactically. In 1958, when our airborne experimental radar system was completed, we were able to verify these theories, along with many others that our group had formulated. From the standpoint of engineering design, however, we preferred the mercury-arc source to a white-light source, and since our system utilized compression ratios on the order of only 100 we found that no correction was needed for either lateral or longitudinal dispersion. The mercury-arc source was entirely adequate, and we made high-quality reconstructions with it for several years until we switched to laser sources in 1962.

A further improvement in holography was introduced in 1963, when we set out to correct another defect in the holographic process. With coherent light diffraction patterns are formed from any scattering center in the path of the light. Thus dust on optical surfaces, bubbles and other occlusions in lenses all add their diffraction patterns to the hologram, which then transfers them to the image. We introduced a diffuser between the object and the light source in order to obliterate these patterns. The technique worked well and enabled us to form images of a quality genuinely comparable to those of conventional photography. Since the diffuser tends to spread the object-scattered light into wider angles, however, the expansion ratios with this technique are significantly larger, thereby increasing the coherence requirement, although not beyond the range of the mercury-arc light source.

Our next step was to make holograms of three-dimensional reflecting objects, instead of the transparencies that had until then been the objects for holography. Making a hologram of a three-dimensional object is much more difficult for several reasons, one of which is the tremendously increased temporal-coherence requirement. This reason can be readily understood if one thinks of the temporal-coherence requirement in terms of coherence length. In preparing our setup we split a laser beam into two parts. One part, serving as the reference beam, was directed to the photographic plate by means of a mirror. The other part of the beam was reflected from the object to the photographic plate. In order for these two beams to interfere the paths they traverse should not differ in length by more than the coherence length of the light. We therefore measured the paths and adjusted them so that they were roughly equal. Since the coherence length of laser light is typically many centimeters, we could afford to be casual. If we had been using a mercury-arc source, for which the coherence length is typically less than .1 milli-



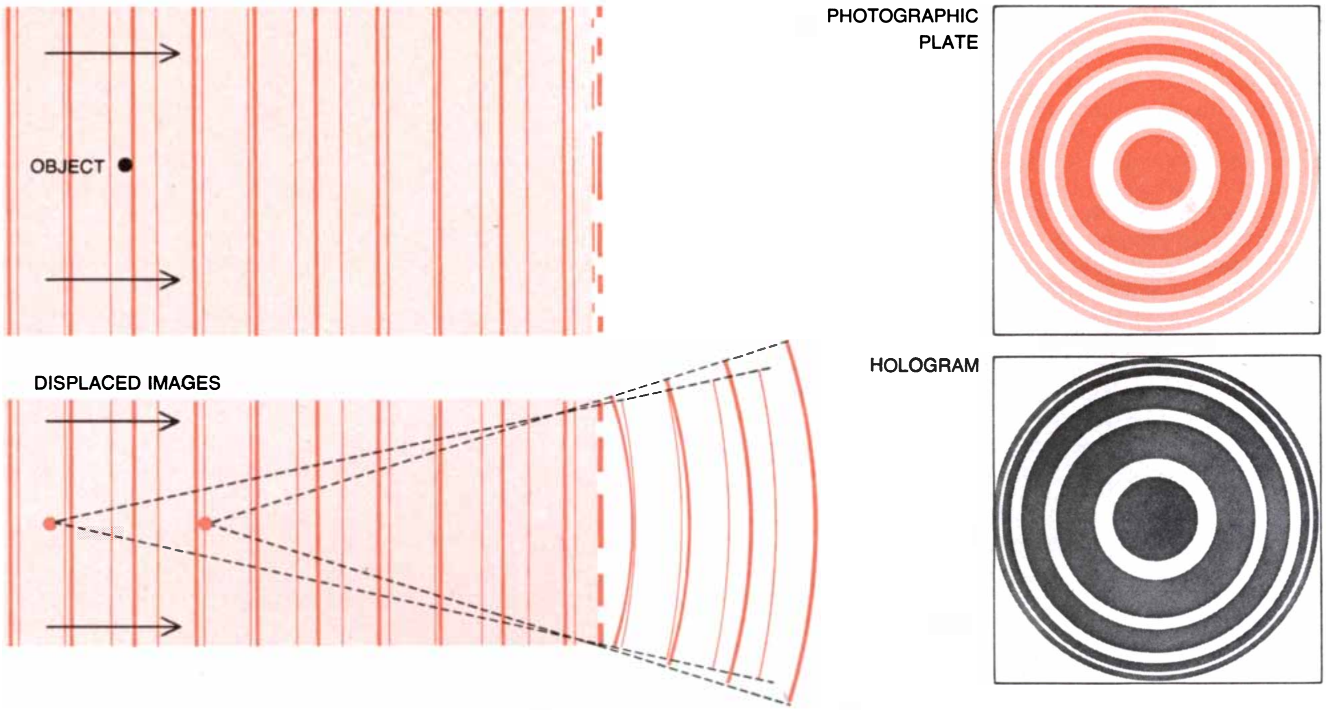
fronts of the reference beam at the photographic plate created a "bull's-eye" pattern resembling a Fresnel-zone diffraction plate. The resulting hologram is an overlay of many such patterns, one for each resolution element of the original object. The diagram at the right shows how the developed holograms were viewed. The diffraction of a portion of the transmitted reference beam by a bull's-eye interference pattern is seen re-creating a virtual image of a resolution element of the original object. The names of the scientists that appear in the hologram represented in these drawings were in Gabor's original hologram, published in *Proceedings of the Royal Society* in 1949.

meter, we would have had to equalize the paths much more accurately.

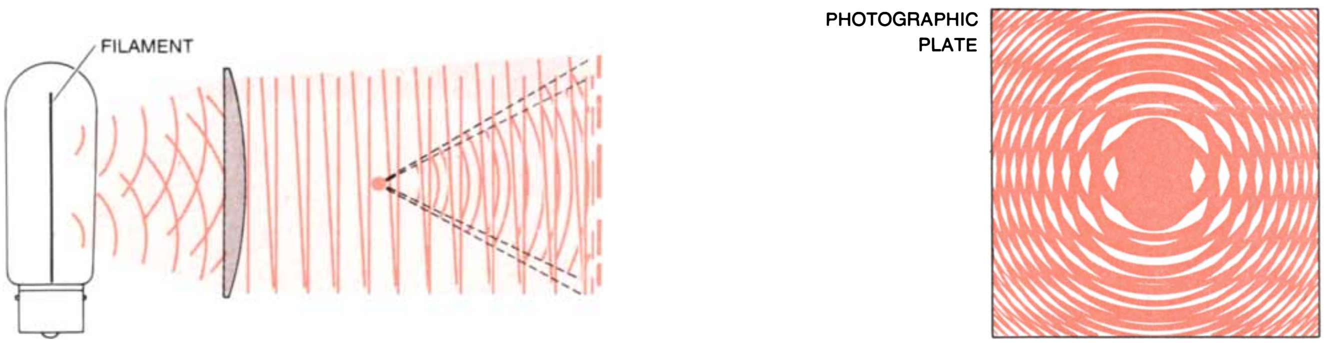
Assuming that the object has depth, however, the length of the object beam depends on the location of a point on the object, and it would therefore not be possible to equalize for all object points. We would then need a light source with a large enough coherence length to encompass the depth of the object. If the object has a depth of half a meter, say, the coherence length should be at least a meter, since that would be the difference in the length of the light path between a light ray reflected from the front of the object and a ray reflected from the most distant part of the object. The corresponding coherence length required to make holograms of transparencies in general would be about .01 millimeter. This new coherence requirement is independent of the one described above in connection with transparencies, but it is so great that the previous requirement is negligible by comparison. This step in our work therefore required the laser, which had fortunately been under development during the same period, and it came on the scene at a particularly convenient time for us.

Up to this point the hologram-making and hologram-viewing processes had had similar coherence requirements, a fact readily appreciated by regarding the latter simply as being a retracing of the rays that mediated the formation process. In the viewing process, however, there is no counterpart of the phenomenon of large optical paths generated from reflections from different object depths, since the process does not involve any such reflections; hence this new coherence requirement does not apply to the viewing process. It follows that holograms of three-dimensional objects, which require a laser for their formation, can be viewed in the light of a mercury-arc source, and indeed this is often done.

Meanwhile in 1962 the Russian investigator Yu. N. Denisyuk reported a ma-



SCALE OF THE BULL'S-EYE diffraction pattern formed by a point object is proportional to the wavelength of the incident light. In the convention adopted in these diagrams light of two different wavelengths is represented by wave fronts drawn at two different intervals using two different thicknesses of colored lines. The upper diagram shows that the pattern formed by the longer-wavelength, or redder, light is larger than the one formed by the shorter-wavelength, or bluer, light. A pattern formed by a light beam composed of a wide band of wavelengths would therefore be blurred at the edges, and the resulting hologram would also be blurred, even if it were viewed in perfectly monochromatic light. The lower diagram shows a hologram made with perfectly monochromatic light but viewed in polychromatic light. The hologram acts like a lens, causing a portion of the incident light to be converted into a diverging wave, which projects back to a virtual image whose resolution is ideally the same as that of the original object. The distance between the hologram and the image formed at each wavelength, however, is proportional to the wavelength; hence for white light, which consists of all the visible wavelengths, there is no plane where all the light is focused, and the resulting image is blurred, a defect known as longitudinal dispersion.



EXTENDED LIGHT SOURCE can be thought of as a collection of point sources, each illuminating a point object from a slightly different direction. The bull's-eye patterns are projected in slightly different directions and hence are displaced on the hologram. Thus a spatially incoherent, or extended, light source yields a smeared pattern, with finest interference fringes being the ones obliterated first.

TR7 WINS AT CHARLOTTE, LIME ROCK, BRIDGEHAMPTON, POCONO AND NELSON LEDGES TO ALL BUT LOCK UP A SPORTS CAR CLUB OF AMERICA DIVISION CHAMPIONSHIP. CELEBRATE OUR TRIUMPH WITH THE

For a sports car to win this many victories in only two months of competition is unheard-of.

For the TR7 to have won these victories against such fine racing veterans as Alfa, Lotus, Datsun and Porsche, makes it truly a cause to celebrate.

So, we're celebrating. By offering a special TR7 Victory Edition with free competition-type spoker wheels, vinyl top and racing stripes. At participating dealers for a limited time only.

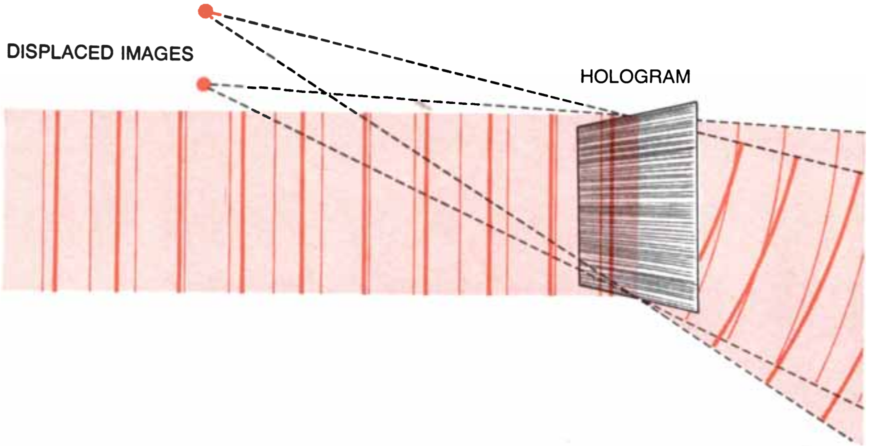
Considering what these extras would cost, this TR7 is an unbeatable value as well as a race-proven winner.

The Triumph TR7 Victory Edition. One of the few sports cars around today that actually earned its stripes.

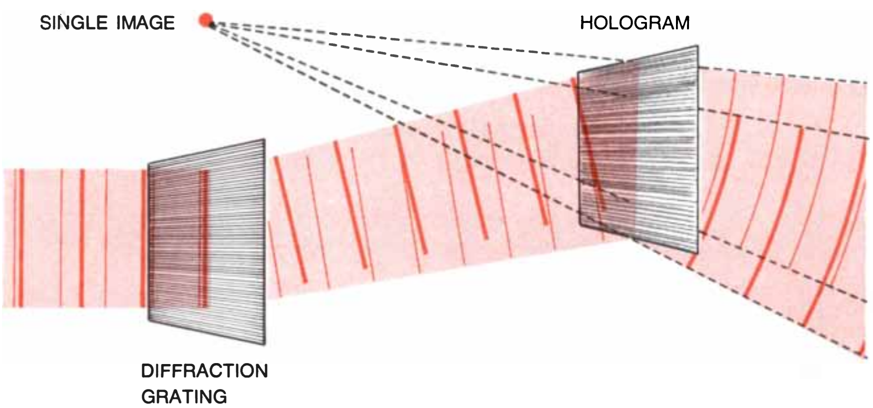
For the name of your nearest Triumph dealer call 800-447-4700. In Illinois call 800-322-4400. British Leyland Motors Inc., Leonia, New Jersey 07605.

BRITISH LEYLAND
TRIUMPH

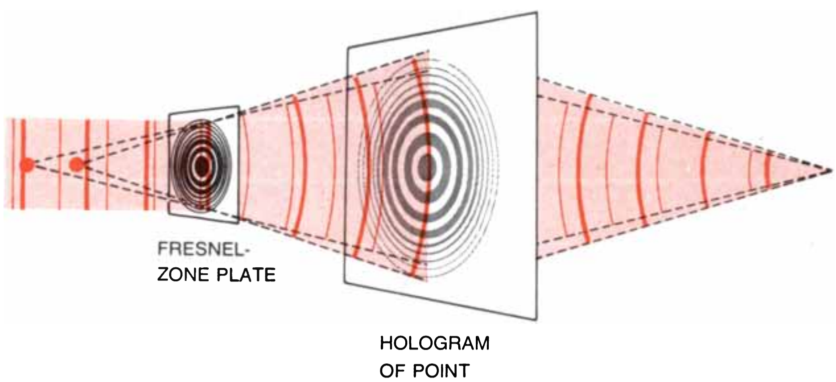

EASTERN LITTLE
102001000
4
SC
20



OFF-AXIS HOLOGRAM, which was invented by the author and his colleague Juris Upatnieks in 1960, behaves like a parallel-line diffraction grating, bending the light at an angle that is approximately proportional to the wavelength. Images formed at two different wavelengths are laterally displaced, with longer-wavelength image having greater displacement.



LATERAL DISPERSION IS CORRECTED by inserting an actual diffraction grating in the path of the incident light beam. The grating is carefully chosen to have a spacing equal to the average spacing of the fringes of the hologram. Hence it can be used to bend the light oppositely from the way the hologram does, exactly compensating for the lateral dispersion caused by the hologram. As a result images formed at two different wavelengths are focused at same point.



LONGITUDINAL DISPERSION IS CORRECTED by inserting a Fresnel-zone plate in the path of the incident light beam. The device predisperses a beam of white light longitudinally, forming a slightly displaced virtual image for each wavelength, with the shorter wavelengths focusing farther from the zone plate. This collection of displaced images serves in turn as the light source for the hologram (in this case again a hologram of a single point). The hologram, which acts just like such a zone plate, bends the longer-wavelength light more, bringing all colors to a common point. Combination of parallel-line diffraction grating and Fresnel-zone plate drastically reduces coherence requirement for making and viewing off-axis holograms.

jor advance, in which the process of holography was combined with a form of color photography invented by the French physicist Gabriel Lippmann in 1891. The Denisyuk hologram can produce either monochromatic or color images when it is viewed in white light from a point source.

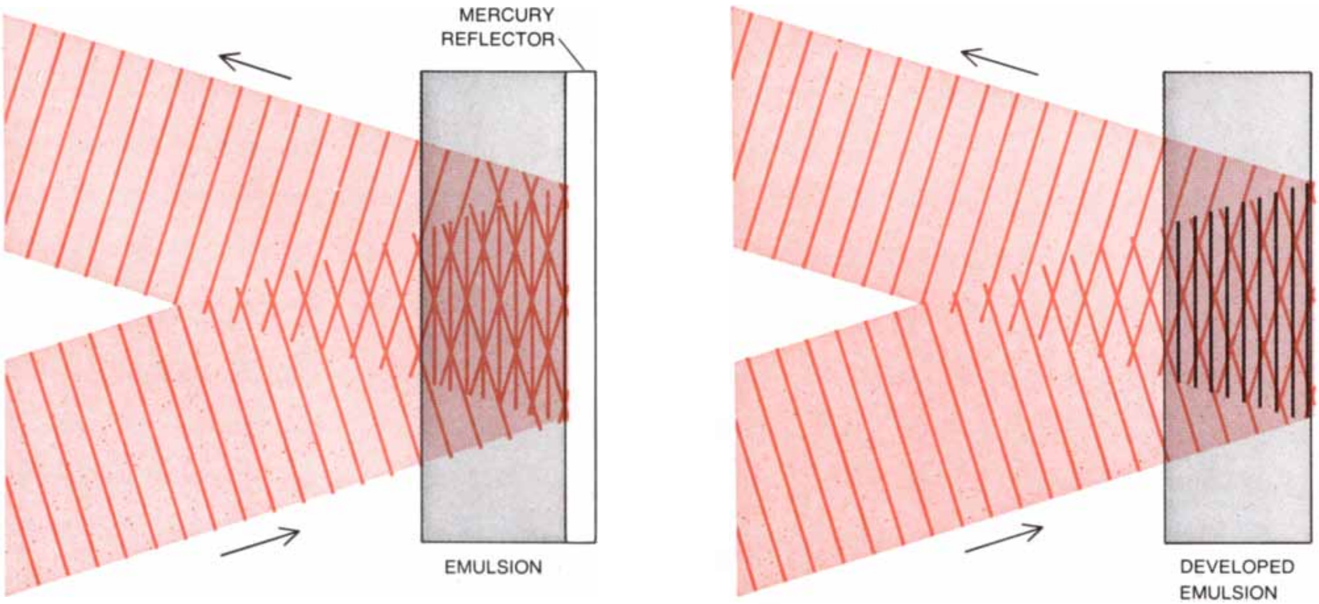
In the Lippmann process an image is formed on a high-resolution photographic plate. The plate is backed on the emulsion side by a pool of mercury, so that the light, which enters the emulsion from the plate side, is reflected and traverses the emulsion a second time. The two oppositely traveling beams of light readily interfere, even if the light source has a broad spectrum of wavelenths, since the path difference is nearly zero. The light source can also be spatially incoherent, because it is really the same rays, or closely adjacent rays, that are superposed. Thus fine fringes (actually surfaces of brightness interspersed with surfaces of darkness) are formed, aligned nearly parallel to the emulsion. The spacing between the fringes is only half a wavelength of light; hence some 30 fringes are formed across an emulsion with a typical thickness of 15 microns. The fringes result in silver deposits in the bright regions, as in the usual photographic process. The silver grains deposited at each location have a concentration related to the brightness of the image in that region; moreover, their spacing is related to the color of the image at that position [*see upper illustration on opposite page*].

When this structure is illuminated with white light, each layer of silver grains acts as a mirror, reflecting a small portion of the incident light while transmitting the greater portion to the next surface, where a little more light is reflected. The total light reflected from each location on the plate is proportional to the original light that created the fringes at that location; thus the observer sees on the plate an image of the original object, just as in a conventional photograph. More important, the image is in full color, because of the interference effects among the various layers. Each layer reflects light in all colors, but there is reinforcement only of certain colors, depending on the distance between the fringes. Since the spacing of the fringes was determined by the original light, the colors that are reinforced are those corresponding to the colors of the original light.

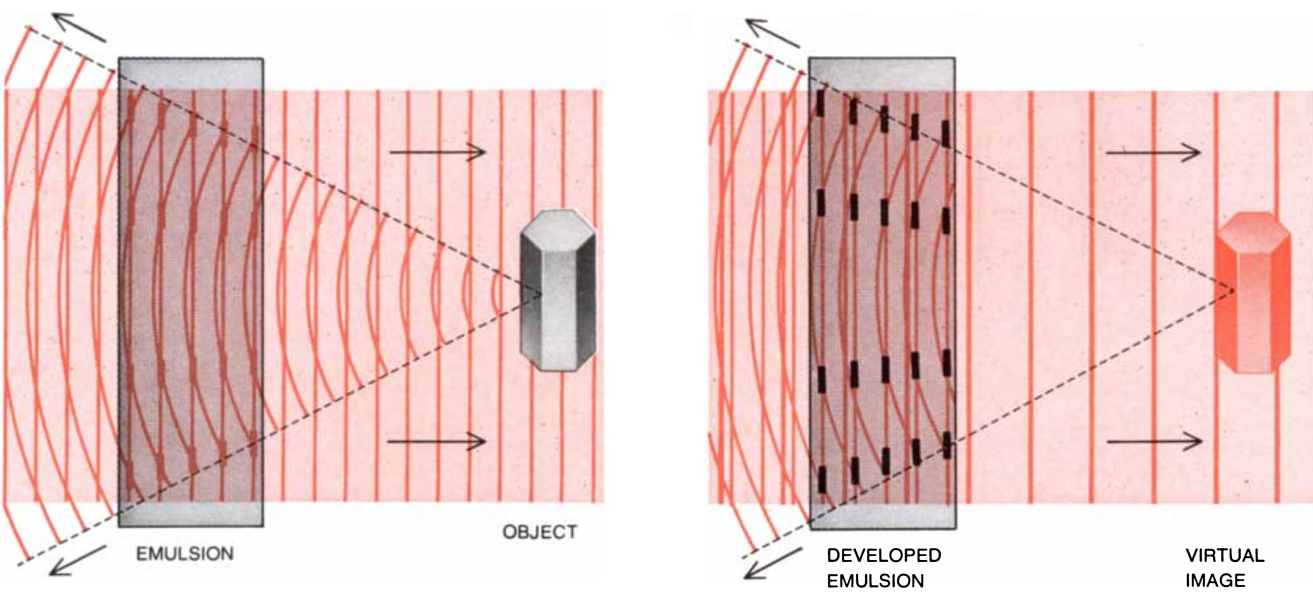
Denisyuk's work, which is one of the cornerstones of holography, combined the work of Lippmann and Gabor by using coherent light and recording instead of an image the diffraction patterns produced by the object. In Denisyuk's method light passes through the plate, falls on an object and then is reflected back to the plate, traversing it a second time but in the opposite direction. As in the Lippmann process, fringes are formed throughout the depth of the emulsion. The result is a hologram that has the characteristics of the Lippmann plate and can be viewed with a white-light source of limited spatial extent, such as a projector lamp or an ordinary incandescent bulb [*see lower illustration below*]. Moreover, by using three beams of coherent light at three discrete wavelengths, one for each of the primary colors, a hologram can be made to yield a full-color image.

The next step was to combine our technique, employing a separate reference beam, with the Denisyuk technique. By early 1964 reports of our technique had become widely disseminated, and Denisyuk's papers had been translated into English and were readily available. The actual combination, however, was not achieved until the latter part of 1965, and then it was done almost simultaneously by three groups. Our group reported and demonstrated such holograms at the spring 1966 meeting of the Optical Society of America; so did another group at Michigan, headed by G. W. Stroke and A. E. Labeyrie. As it happens, Nile Hartman, working at the Battelle Memorial Institute, had achieved the same result several months earlier, and consequently he was awarded the contested patent.

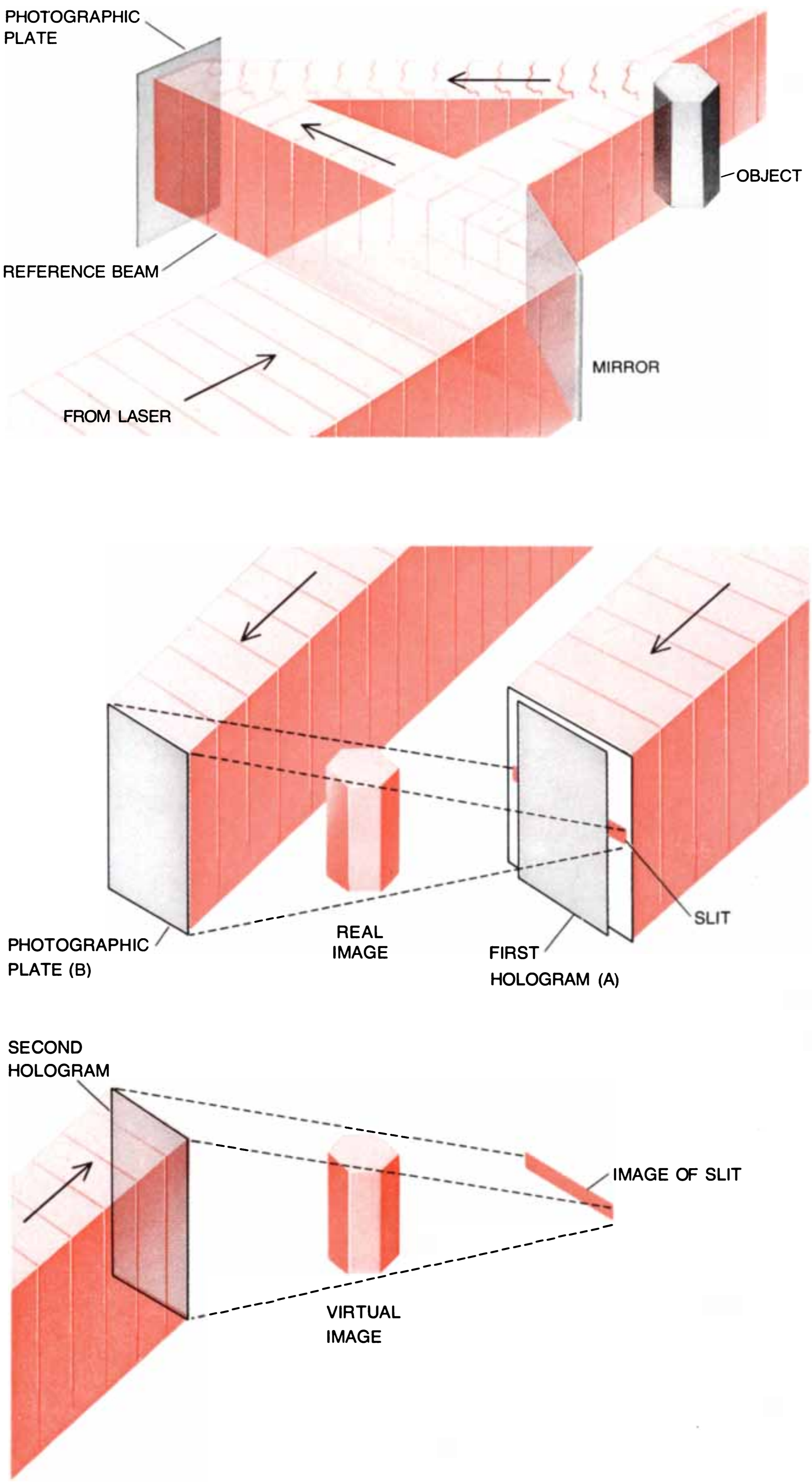
The Denisyuk white-light hologram is in effect a hologram and a narrow-band color filter combined into a single structure. Indeed, the same white-light viewing can be accomplished with a narrow-



LIPPMANN PROCESS for color photography, invented by the French physicist Gabriel Lippmann in 1891, produces interference fringes in a thick photographic emulsion by allowing the incident light to reflect back on itself with the aid of a pool of mercury on the back side of the emulsion (*diagram at left*). Silver deposits formed at the surfaces of maximum brightness are then fixed by developing the plate. In the readout process (*diagram at right*) the incident light is partially reflected at successive layers of silver grains in the emulsion. The reflections reinforce one another at the wavelength (or wavelengths) corresponding to the original light used to expose the plate.



DENISYUK METHOD for holography, basically a combination of the Lippmann process and the Gabor holographic process, operates by passing a beam of coherent light through a photographic emulsion, reflecting the light from an object and recording the interference created in the emulsion between the incident beam and the reflected beam (*left*). The hologram is viewed by directing white light at the hologram, which selects by reinforcement light of only one wavelength, forming a three-dimensional image of the object (*right*).



RAINBOW HOLOGRAMS ARE MADE in two steps. First a "master" hologram is produced in the conventional way (*top*). Next a horizontal slit is placed over the hologram, and a real, or converging-wave-front, image is formed by illuminating the hologram with the conjugate, or time-reversed, version of the reference beam (*middle*). This image then serves as the object for a second hologram, which is recorded near the real-image space with the aid of a converging reference beam that is vertically inclined. When the second hologram is illuminated with a white-light source, each wavelength forms an image of the slit at a different vertical position (*bottom*). When the viewer looks through any one of these slit images, the original scene is perceived in three dimensions, complete with horizontal parallax, in a single color. As the viewer moves his head up and down the image appears successively in the various colors of rainbow, although it lacks vertical parallax. Technique shown here was invented by Benton.

band filter in combination with other types of holograms, so that the main advantage of this type of hologram lies in avoiding the inconvenience of a separate color filter.

In 1969 Stephen Benton of the Polaroid Corporation reported another significant advance in display holography. His method involves two steps. First a hologram (typically of a three-dimensional object) is formed in the conventional way. Next a second hologram is made, using as its object the real image from the first hologram. In this process a narrow horizontal slit is placed at the first hologram. Since each point on a hologram reproduces the entire image, but from only one viewpoint, the slit eliminates all parallax in the vertical plane. (Parallax, an essential element in the perception of three-dimensionality, is the apparent displacement of an object when seen from two different points.) The second hologram, under coherent illumination, forms an image (either real or virtual) of the original object as well as a real image of the slit. To see the entire image the observer must place his eye at this slit image [*see illustration at left*]. If he moves his head vertically, the image disappears; hence there is no vertical parallax. The image, however, can be quite bright; all the light that ordinarily would be diffracted over a vertical distance of perhaps a meter at the observer's position is now confined to the slit image, and the light gain can be several hundredfold.

There is, however, an even more important advantage to this method. If one were to illuminate the hologram with light of two different wavelengths, the lateral dispersion would result in two vertically displaced slit images, one for each color. The observer would see the entire image in one color when he positioned his eyes at the one slit image and in the other color when he positioned them at the other slit image. Extrapolating this process to white-light illumination, where the slit image would be smeared into a continuous spectrum, one finds that to view the hologram the observer may position his eyes at any point on the spectrum; wherever he positions them he will see the image sharp and clear in the color corresponding to that part of the spectrum. As he moves his head vertically the perceived image changes color but remains sharp and clear. If his head is not in the image plane of the slit, he will still see the entire image, but the color will change from top to bottom. Accordingly such holograms have sometimes been called rainbow holograms.

In exchange for giving up vertical parallax while retaining normal horizontal parallax one thus obtains holograms that can be viewed in white light without the need for any narrow-band filter, either external or built in. Moreover, all the white light is utilized, rather than just a narrow band of the spectrum, making the Benton holograms exceedingly bright. The loss of vertical parallax is in general unimportant, since an observer normally moves his head horizontally, not vertically. Of course, if the observer moves his head so that his eyes are displaced vertically rather than horizontally, each eye would receive the same image and there would be no perception of three-dimensionality. The Benton method is in short a favorable trade-off in which comparatively unimportant features are exchanged for important ones.

Another noteworthy development, the composite or multiplex technique, overcomes some of the basic limitations of the more conventional holographic techniques and has led to impressive results. In this technique a hologram is synthesized from a great many ordinary photographs made in a conventional way. The pictures, made from different positions, constitute many views of the object, and they have in their totality all the essential information contained in a hologram. The composite method synthesizes the many pictures into a single hologram.

Composite holograms were first reported in 1967 by Robert V. Pole of the International Business Machines Corporation. In the next few years many advances were made, culminating in a system perfected over the past five years by Lloyd Cross and his colleagues at the Multiplex Company, who introduced a number of ingenious innovations. In the method devised by Cross the object is placed on a slowly rotating platform and a motion-picture camera makes hundreds of pictures, just as in conventional cinematography. These pictures could then be made into a single hologram in the following way. The first picture could be imaged with coherent light onto a screen consisting of a piece of frosted glass. The light transmitted by the screen would be recorded as a hologram in the form of a long, thin strip about a millimeter wide and 20 centimeters long, the purpose of the screen being to direct light from all parts of the image to the recording slit. The actual system differs from this in that instead of the diffuser a large cylindrical lens at the image plane is used for directing the light to the slit; the image from the resulting hologram is much superior. After one frame is recorded as a hologram the next frame is similarly projected, and a second hologram is made in an adjacent position contiguous with the first, and so on. This is readily done by means of a vertical slit placed in front of the recording film, the film being advanced the width of one slit between exposures.

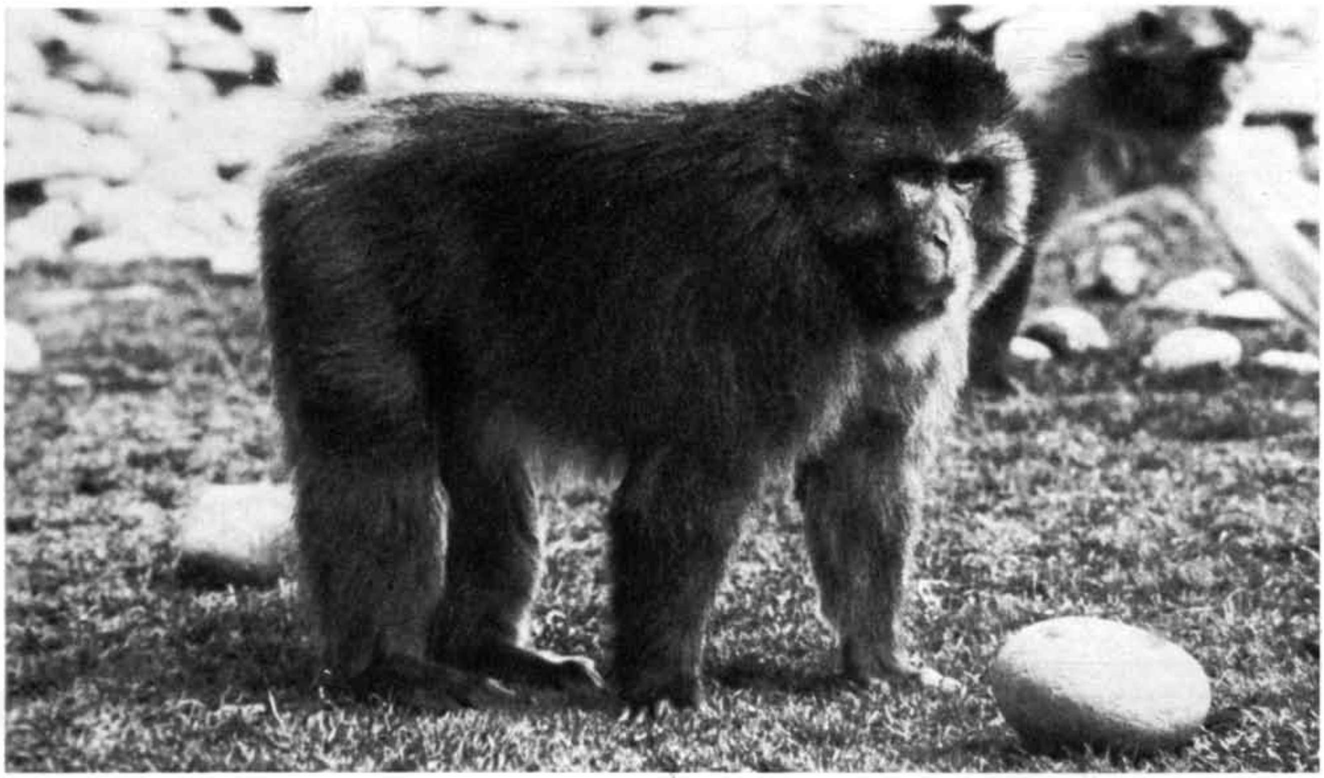
The resulting composite hologram is then wrapped into an arc corresponding to the angle of rotation of the object and is illuminated as a conventional hologram. Each component hologram projects a virtual image without depth or parallax, just like the image from which the hologram was made. When an observer views the film strip, however, each eye looks at an image point through different holograms. The images formed on the two retinas are slightly different, as in a stereoscopic viewer, and consequently a three-dimensional image is perceived. As the observer moves his head laterally, different holographic images are seen, thereby producing the parallax relations of real life. The many individual frames, containing many different views, have in effect been formed into a single hologram. The parallax changes appear in small jumps, of course, but these are no more noticeable than the small jumps between successive frames in conventional motion pictures. There is no parallax in the vertical plane, but again that is of little consequence.

The Cross technique can be combined with the Benton technique, thus allowing white-light viewing. Essentially nothing is given up in doing so, since the composite holograms lack vertical parallax anyway.

The composite technique has made it possible to make holograms of live subjects without the need for illumination with laser light. Moreover, since some motion of the object is permissible, the technique has led to holograms where the subject can be seen moving his hand, shifting his position or changing his facial expression. Such motion is perceived by viewing different parts of the hologram in sequence, either by having the viewer walk past the hologram or by rotating the hologram.

Composite holograms are formed from pictures made in a conventional way with conventional white light. When they are combined with the Benton technique, they can also be viewed in ordinary white light. Since the hologram is formed from transparencies only, and since the cone of rays striking each part of the film is quite narrow, it is apparent that the hologram could be made with the quasi-coherent light of a mercury-arc source or, taking advantage of our original scanning technique, with ordinary white light. In short, the Cross process, which produces some of the most dramatic and sophisticated of all holograms, requires no laser at all and therefore could have been implemented even in the late 1940's, when Gabor first conceived of holography.

From this account of holographic methods and their coherence requirements one fundamental observation stands out: The process is by no means dependent on highly coherent illumination, either in the making process or in the viewing process, and indeed it never has been.



"ALPHA," OR DOMINANT, MALE of a troop of Japanese macaques kept under seminatural conditions in a two-acre corral at the Oregon Regional Primate Research Center is named Arrowhead. The alpha male is responsible for directing the movement, defense and policing of the troop. Arrowhead is about 25 years old, and he was probably the alpha male when the troop roamed wild in Japan. Although he is one of the smallest males, has no canine teeth and is missing an eye, a challenge to his leadership has never been observed.

ADULT FEMALE MACAQUES huddle with their infants atop a dead tree stump in the Oregon corral. The single offspring are born in April and May and are nursed throughout the summer. After the infant is about two weeks old, the mother teaches it to walk and which foods to eat. The female macaques are very protective and do not allow other females to touch the infant until it is several weeks old.

The Social Order of Japanese Macaques

Long-term observation of a troop of these monkeys confined in an Oregon corral indicates that the biological component of their behavior is much modified by the social component

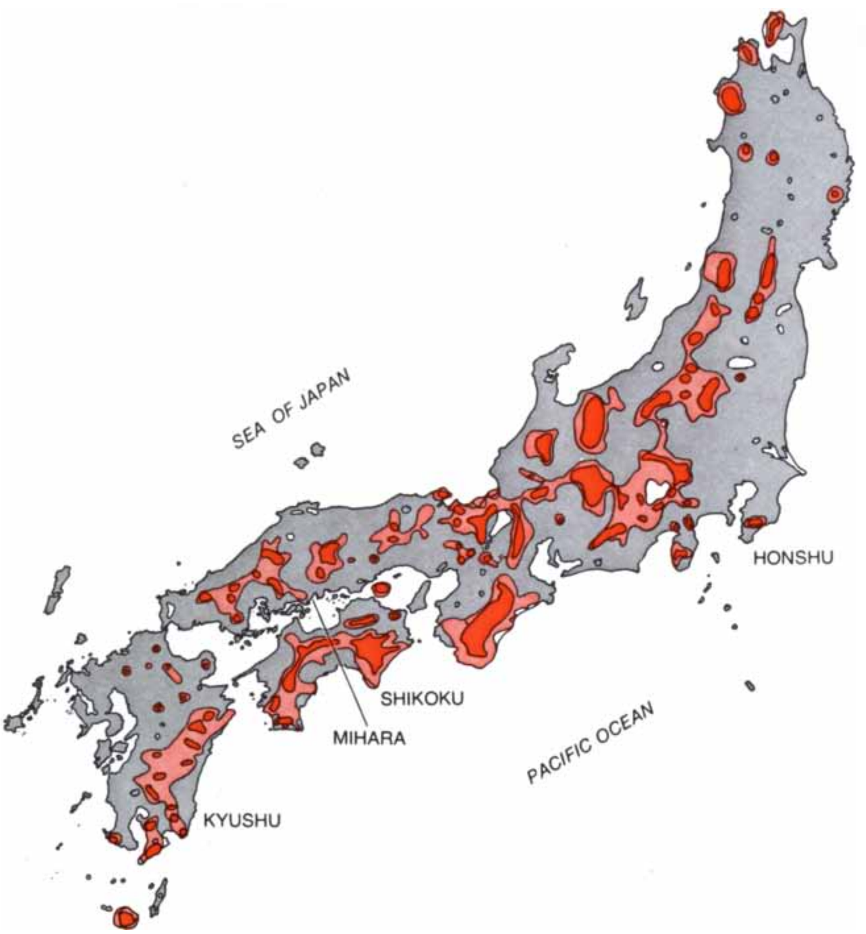
by G. Gray Eaton

Japanese macaques are perhaps best known in the form of the three wise-looking wooden monkeys that mime the Buddhist proverb "See no evil, hear no evil, speak no evil" on the world's souvenir stands. The choice of models is apt, because in real life macaques are among the most intelligent of primates. Relatives of the rhesus monkey and the Barbary ape, they live in cohesive troops of 50 to 150 individuals. They have a complex social order that like all other social orders has been created by the interaction of biological and social forces. What is particularly interesting about the social order of macaques is that the social forces play a larger role than one might expect.

In their natural forest habitat Japanese macaques are shy and difficult to observe, and ethologists interested in their social behavior have been obliged either to provision wild troops or to study them in captivity. Over the past six years my colleagues and I have spent more than 8,000 hours observing a typical macaque troop in a grassy two-acre corral at the Oregon Regional Primate Research Center. Our troop had originally roamed free near Mihara City in the Hiroshima prefecture of Japan, until complaints from local farmers prompted the Japan Monkey Center to capture nearly all the 49 troop members in 1964. The following year they were transplanted to our corral in Oregon, and since then they have thrived.

The macaques in our troop were not studied before their capture, so that it is difficult to assess the effect confinement has had on them. We assume, however, that provisioning and confinement tend to increase the frequency of all types of social interaction, since the restriction of movement and the freedom from having to spend hours foraging for food have increased the animals' unoccupied time. When on occasion we have fed our macaques by scattering grain throughout the corral instead of doling out their usual ration of monkey chow, we have noticed how quiet they become when they are foraging for food.

The most striking feature of social behavior in the troop is the fact that a few males dominate all the other animals. Next one notices several old females that attack other females without retaliation, and many adult females that threaten and chase males. It soon becomes apparent that there is a rigid dominance hierarchy, analogous to a pecking order. The top position, that of the "leader," or "alpha" monkey, is almost always occupied by a mature adult

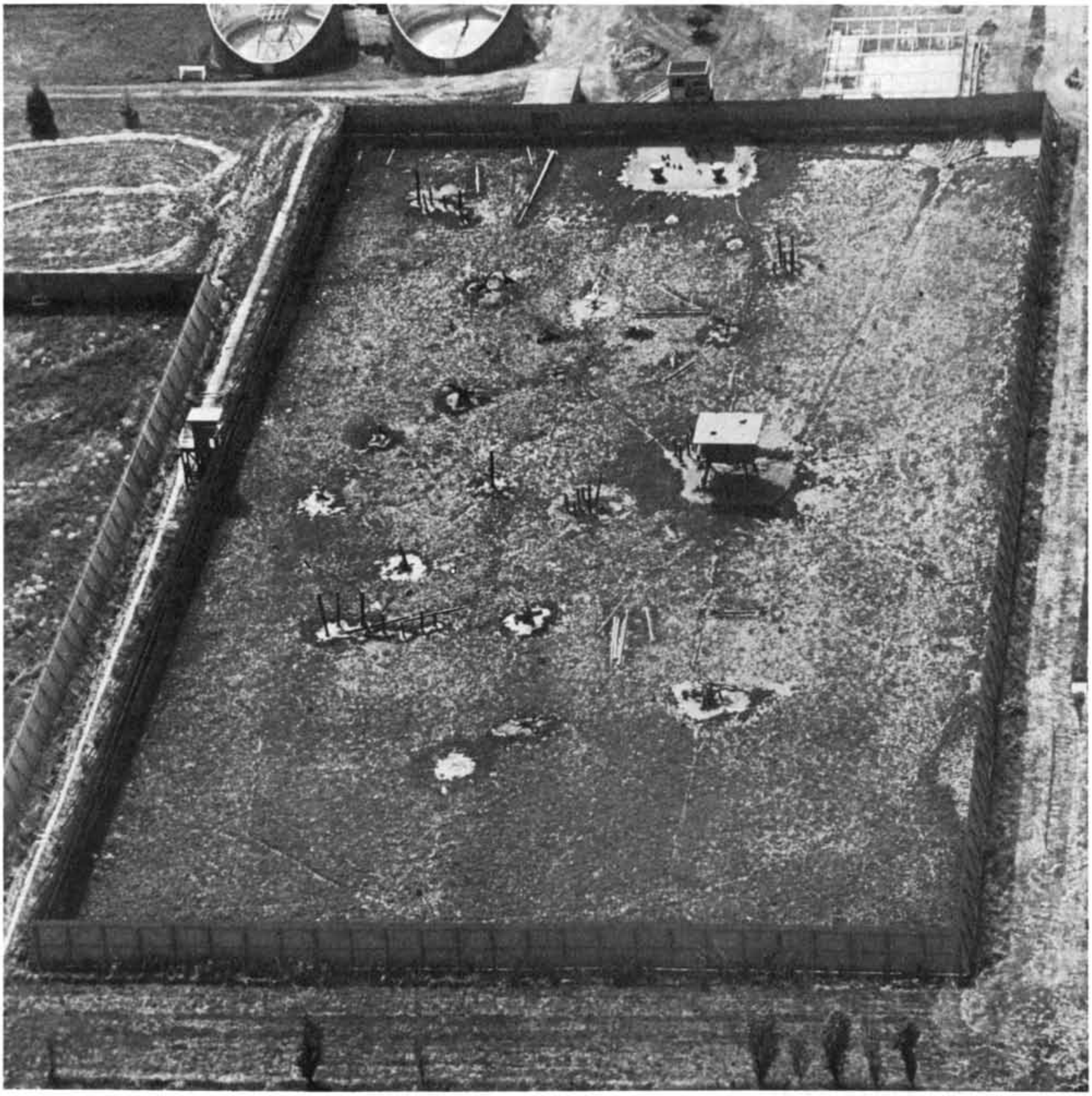


SHRINKING RANGE of Japanese macaques is shown in this superposition of distribution maps made by K. Hasebe in 1932 (*light color*) and by K. Kishida in 1953 (*dark color*). The most northerly-living of nonhuman primates, their distribution extremes have remained unchanged during the past half-century, but deforestation has caused decreases in every area. Today macaques are considered a national treasure, and the Japanese government bans their export.

male that sometimes does not attain this rank until he is 18 or 19 years old. (Males normally reach puberty at four and full body growth at eight to 10.) Immediately below the alpha male are typically five or six "subleader" males, followed by most of the adult females, which reach puberty at three years and full body size at six to eight and together with their infant and juvenile offspring form the middle of the hierarchy. The remainder of the adult males are at the bottom of the hierarchy, and in the wild they live on the periphery of the troop.

Rank is not highly correlated with aggressiveness, so that we cannot determine the order of the hierarchy simply by counting the number of attacks. Instead the troop is organized in such a way that the highest-ranking animal is attacked by the smallest number of other animals, with the progressively lower-ranking animals being attacked by an increasingly large number of others. Dominance rank is basically linear, but there are occasional reversals, that is, animal *X* chases animal *Y*, *Y* chases *Z* and *Z* chases *X*.

By the same token, high rank does not necessarily entail a high frequency of aggressive behavior. Our second- and third-ranking males, named Greater Than and Perfect, are both more aggressive than the alpha male, Arrowhead, because, among other reasons, Arrow-



GRASSY TWO-ACRE CORRAL at the Oregon Regional Primate Research Center confines the author's 230-member troop of Japanese macaques. (The monkeys appear as black dots scattered throughout the enclosure.) The walls are of corrugated steel, 3.5 meters high, and slant inward about 15 degrees. Poles and tree stumps provide climbing surfaces for this largely ground-dwelling but occasionally arboreal species; the trails in the grass were made by the monkeys as they wandered between preferred areas. Three observation towers are visible, and at left is part of a newly built corral identical with the first that will be used in the near future for a population experiment.

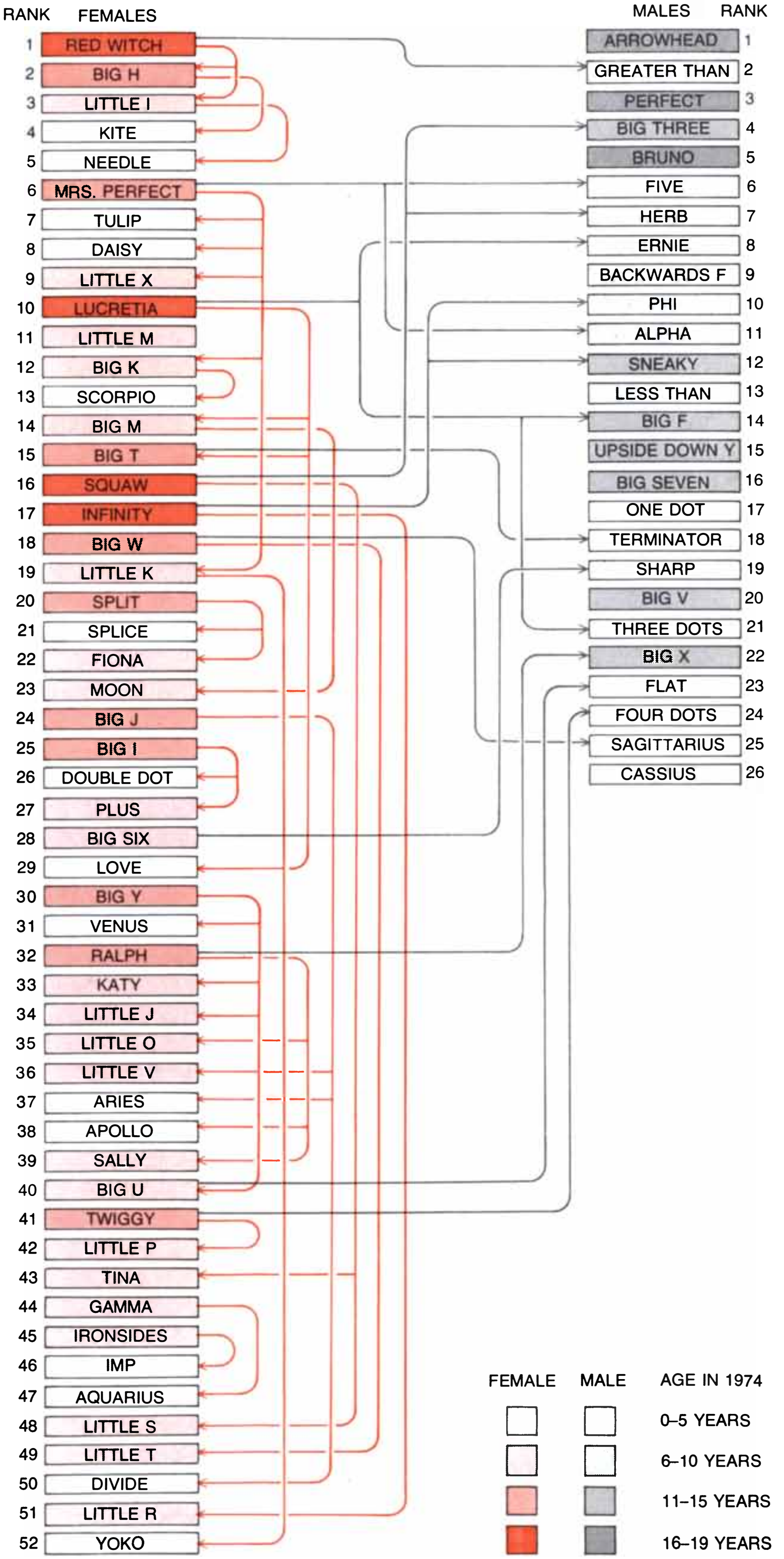
head receives more "respect" from the other troop members and therefore does not need to maintain his position by constantly attacking other animals or even threatening them. For example, when the monkeys are fed, Arrowhead is given more personal space than Greater Than, and when a prized food such as a peanut is thrown near either of them, Arrowhead's mere presence is enough to keep the other animals away, whereas Greater Than often has to chase them away.

Paradoxically the dominant males are not at the top of the hierarchy because of fighting ability or physical characteristics such as size. Arrowhead is one of the smallest of the adult males, and he has no canine teeth and only one eye, but we have never observed any challenge to his authority as the alpha male. Moreover, when he attacks other males, they do not fight back but merely struggle to get free. Of the subleader males, Bruno and Perfect also lack canine teeth, and Greater Than achieved his rank before his canine teeth had fully erupted. If size and canine teeth play no role in attaining or maintaining a high dominance rank, what is the basis of the macaque "pecking order"?

After observing wild troops Japanese ethologists concluded that the dominance rank of an animal is closely correlated with the rank of its mother. The role of the mothers in determining the rank of their offspring can be observed during juvenile fights, most of which begin when play gets too rough. One juvenile screams, inciting its mother to run over and bite the other juvenile, which is in turn aided by its mother. The two mothers then fight, and the dominant female and her offspring chase the lower-ranking female, whose son or daughter flees with her. Once this has happened several times, the offspring of the lower-ranking female will run away from those of the higher-ranking female even when she is not nearby. Occasionally a young macaque will be unusually aggressive and will rise above its mother's rank, but that is exceptional.

We have several adult males in our troop that are still defended by their mothers. When the current second-ranked male, Greater Than, was third in the male hierarchy, he was occasionally attacked by the second-ranked male, One-Eye, and the alpha male, Arrowhead. At those times Greater Than's mother, Red Witch, would distract the dominant monkeys by leaping on their back, enabling her son to escape.

In the wild only the sons of very high-ranking females are allowed to stay in the center of the troop; the other males are driven to the periphery when they are about five years old, about a year after they have reached puberty. At that age they probably have to fight for a

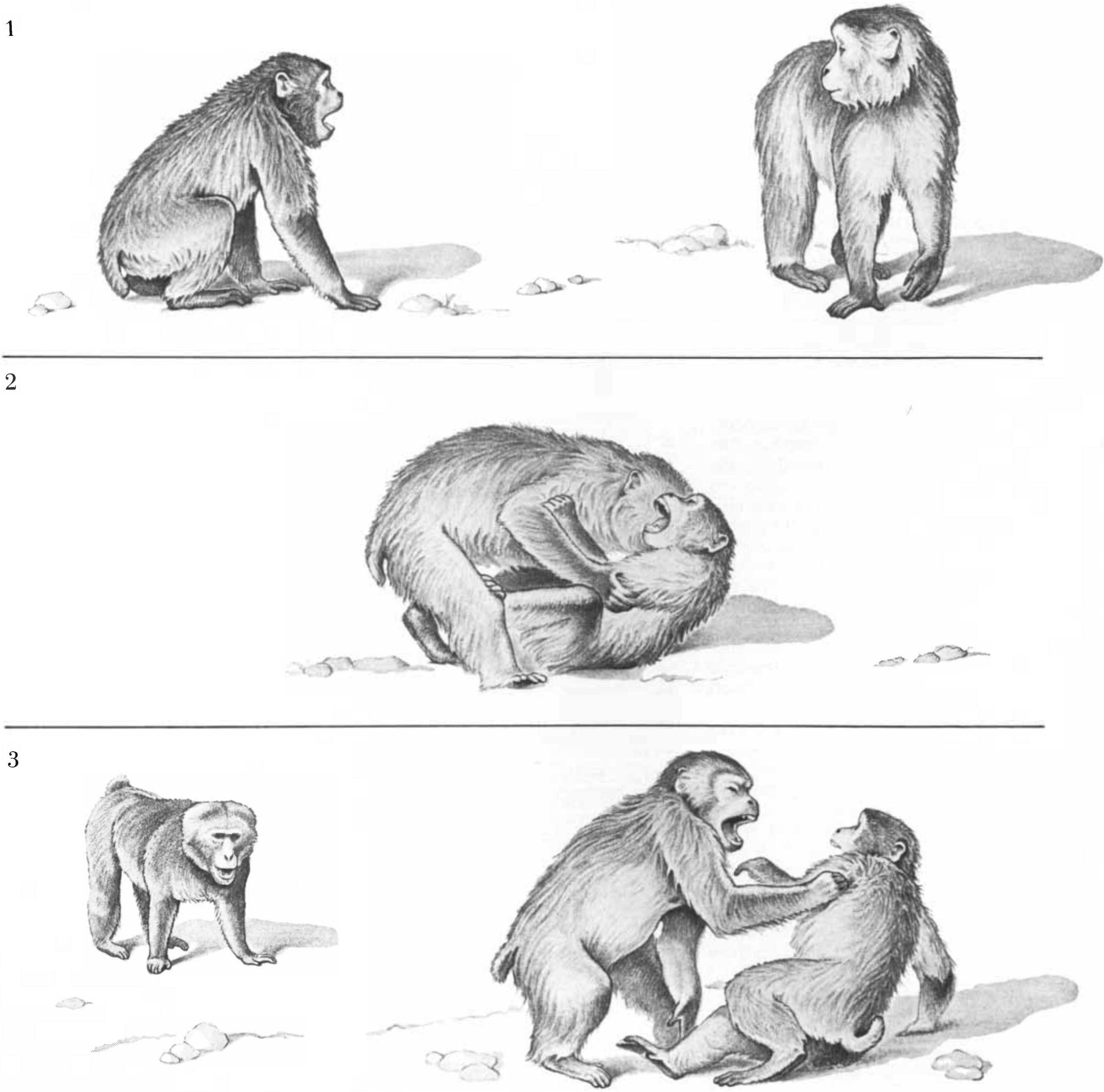


ADULT DOMINANCE HIERARCHIES for the Oregon troop are shown as they stood in 1974. Female hierarchy (*color*) and male hierarchy (*gray*) are listed separately because dominance relationships change during the mating season, when low-ranking males attack females. Arrows indicate maternal relationships to show how rank is influenced by maternal status.

place in the peripheral male hierarchy without the physical assistance of their mothers. The mothers' influence may still carry over, because many macaque fights are bluffing matches, and the sons of high-ranking females would presumably be more self-confident than those of low-ranking females.

Superimposed on the maternally determined dominance hierarchy is a second system of organization we have found to be even more important in determining the overall social order: each monkey has a specific function or role to perform within the troop that is dependent on its age, sex and dominance rank. These roles can be as simple and interchangeable as watching for predators or as complex and fixed as rearing an infant, but together they constitute the essence of a society as opposed to a simple aggregation of individuals.

The role of the alpha male appears to be one of directing the movement of the troop and defending it. Since movements of our troop are restricted by a fence, Arrowhead no longer leads, but he still defends. Once a year, when we round up the entire troop for general



SUBLEADER POLICING EPISODE lasting one and a half seconds was drawn from a series of photographs taken with a motor-drive camera by Kurt B. Modahl. The sequence begins when Tulip, a juvenile female, threatens Infinity, an adult female (*1*). Infinity responds by fiercely attacking and punishing Tulip (*2 and 3*). Tulip's screams prompt Greater Than, the second-ranking subleader male, to rush in from the rear left (*4*) and break up the fight by pursuing Tulip a short distance (*5*) and then threatening her with a stare (*6*). The subleaders

housekeeping purposes, he is conspicuous in his disregard for his own safety in defending the troop. Because of the limited size of the catching pen at the west end of the corral we try to detach small, manageable groups of animals from the main troop and drive them along the corral's north wall. As we do so Arrowhead leaves the main troop and joins the group being herded into the catching pen. Instead of facing away from us and toward the pen as the other animals do he faces toward us and makes threats as he moves backward with the others. Once the other monkeys have been driven into the pen, Arrowhead breaks through the line of animal handlers, returns to the main troop and repeats the performance with the next small group.

Occasionally we have to remove a sick or wounded animal from the corral. As we attempt to do so Arrowhead leads many of the low-ranking males, one or two of the subleader males and some of the adult females in threatening us, while the rest of the subleader males move away and stay close to the mothers and infants in the main troop. The cry of an infant makes Arrowhead and the other adult males appear to be par-

4

5

6

usually attack the dominant combatant; in this case, however, Greater Than chased the losing Tulip. He did so apparently because she may have inadvertently threatened him when she started the fight, and because he seems to like Infinity, although this interpretation is speculative. In any case macaque behavior is variable and complex and cannot be reduced to formulas. The presence of the alpha male, male-female alliances and the personal likes and dislikes of the individual subleaders are all factors that may influence policing behavior.

ticularly enraged, and they advance shoulder to shoulder, growling, bobbing their heads and slapping the ground with their hands.

The principal role of the subleader males is to stop fights. This they usually do by chasing away the more aggressive of the combatants. Such behavior favors the underdog, but it cannot be regarded as "ethical" because it seems to be an automatic response to the dynamics of the fight situation. The screams, grimaces and cringing posture of the losing monkey are signs of submission that tend to inhibit aggression by the subleader males, whereas the growls, gapes and ear-flattening display of the attacking monkey are threatening gestures that are likely to incite the subleader males to attack.

The alpha male occasionally plays a secondary role in this policing behavior. In his presence the subleader males appear uncertain and alternately threaten the combatants and turn with exaggerated motions to look at him. If the alpha male turns away as though ignoring the entire affair, the subleader males may or may not attack the aggressor in the fight; if he threatens the subleaders themselves, they will leave the area; if he joins them in threatening the combatants, they will attack and end the fight.

Role-specialization in a troop of Japanese macaques is largely determined by class membership. Some learning nonetheless seems to be necessary. When Big Three rose from the peripheral class to the subleader class after defeating the subleader Bruno, he appeared to assume the subleader role of policing, but his behavior differed from that of the more experienced subleader males. When a fight started, he would become overly excited and attack animals that were not involved. Over a period of several months, however, he gradually became more proficient, and he is now one of the most effective subleaders in checking aggression.

In the wild the primary role of the peripheral males is apparently to give warning of predators and to help the alpha male defend the troop. They also play with and discipline juvenile males and are groomed by them; the juveniles undoubtedly learn a great deal about social behavior through this interaction. In addition the peripheral males move from one troop to another and thus provide a mechanism of genetic exchange that prevents continuous inbreeding. In our corral the peripheral males are not really on the periphery because the corral is not large enough. They are generally accepted by the rest of the troop and are only occasionally harassed.

The role of the adult females is to raise their offspring and protect them. A newborn macaque is relatively helpless, able only to cling to its mother's breast, and for the first few days the mother supports the infant with one hand as she moves about. The infant develops rapidly and soon attempts to crawl away from its mother. It rarely gets far; the mother is adept at controlling the infant with her foot while she uses both hands to eat or groom. When the infant is about two weeks old, the mother appears to teach it to walk by placing it on the ground and then backing away, encouraging it by smacking her lips. During these developmental stages wide individual differences are apparent. Some females prefer having their infant ride jockey style on their back. Others seem not to care; we have occasionally observed a mother dragging her one- or two-year-old offspring along the ground as it clung to her abdomen.

Infants are a great source of interest to other females, particularly to older sisters and females that do not have offspring of their own. The mothers are extremely protective, however, and do not allow another female to pick up the infant until it is several weeks old. Some females use a rather deceptive stratagem to circumvent this protectiveness: they patiently groom the mother until she is sufficiently distracted and then surreptitiously touch the infant while it is still at its mother's breast.

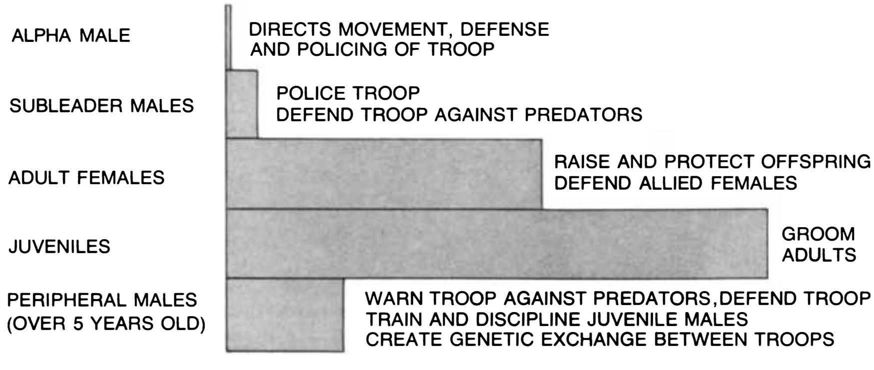
Sex differences in behavior develop as the infants begin to venture away from their mothers. Juvenile males tend to spend more of their time roughhousing in play groups consisting of their peers and an occasional adult male than juvenile females do; the females are mostly occupied in grooming activities with their mothers and sisters. Occasionally a female joins a male play group, but she usually does not stay long before she returns to less strenuous activities. Juvenile males seldom groom other monkeys, but they groom themselves more frequently than females do. When juvenile males and females groom others, they both tend to groom adults of their own sex.

Siblings defend one another and their mother; after puberty, however, the males no longer defend the mother. The bond between mother and daughter remains strong throughout their lives. They spend much time grooming each other; it is not uncommon to see two or three fully grown sisters deftly picking through their mother's fur as their own offspring play beside them.

Adult females also form alliances with one another, a kind of behavior that is rarely observed in unrelated males. These alliances may be the reason many males are driven from wild troops or are low-ranking in our troop. If a male fights with a female, other females will usually come to her aid, whereas adult males rarely assist one another. Groups of females occasionally fight, and in the wild the social disruption caused by continued fighting may result in the fission of the troop, that is, the departure of several families to form a new troop. Female alliances therefore appear to be more important in regulating the cohesiveness of the macaque social order than sexual attraction between males and females, which in this species is seasonal and transitory.

A male may, however, form an alliance with a female and defend her against attack. In our troop the female Gamma is treated in this way by two of the subleader males. One of them, the old male Bruno, "adopted" Gamma when her mother died. Bruno groomed, cuddled and defended her; even now, eight years later, Gamma still runs to him for support. The fact that Bruno has never mated with Gamma leads us to believe that in Japanese macaques the "incest taboo," as a result of which male macaques do not mate with their mothers or sisters, is not genetically determined but has its source in a developmental process. Gamma's other protector, Big Three, does mate with her, but he also defends her during the nonbreeding season.

The seasonality of the macaques' behavior is striking. Their mating is confined to fall and winter; October and November are the months of peak activ-

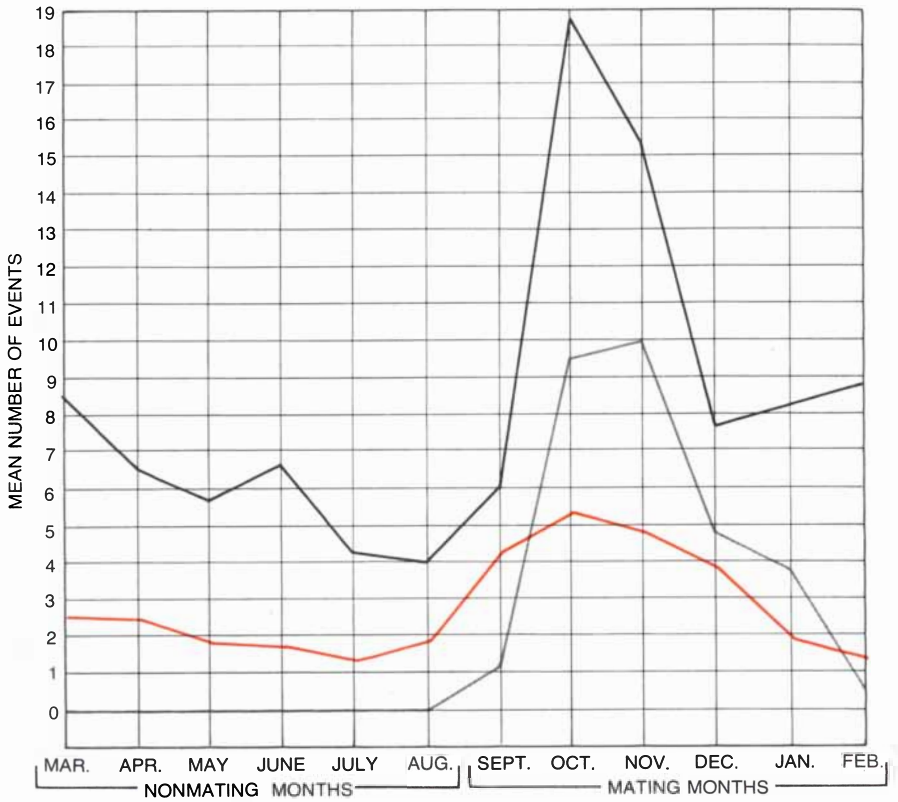


CLASS STRUCTURE of the macaque social order is determined by age, sex and dominance rank. Each class has specific social roles to perform. Bars show relative proportions in 1974.

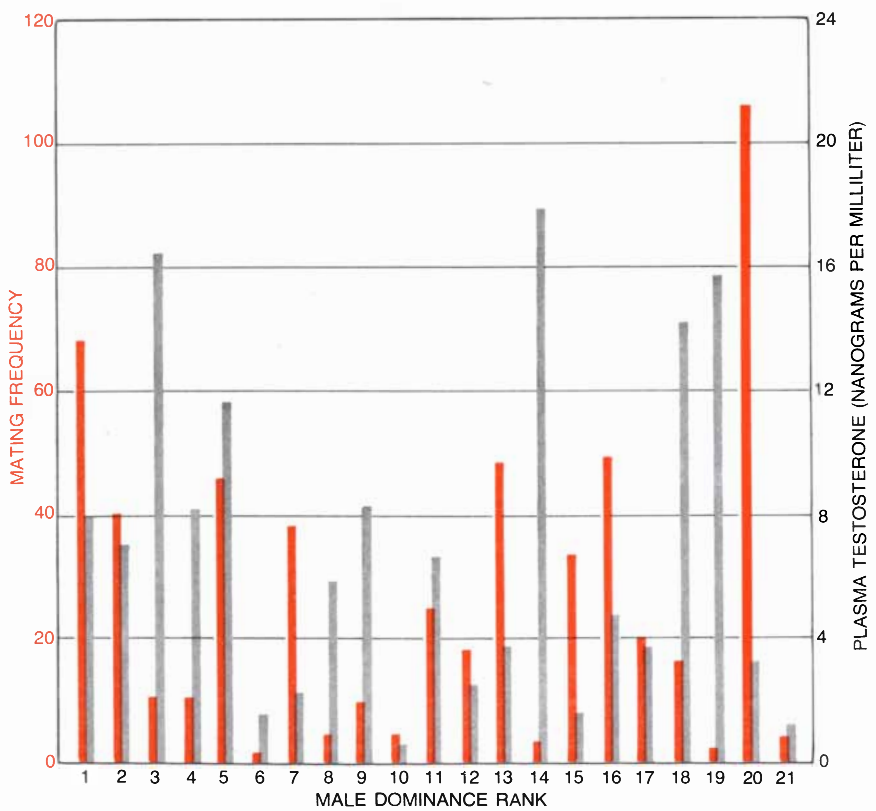
ity. The majority of the young are born about six months later, in April and May. Summer is a quiet season, with the females nursing their infants and grooming one another while the males lie about in the sun. As fall approaches there is a change in behavior, first in the males. Holding their stubby tails erect, they begin to strut around the corral. From the rear the view is striking: the bright white fur on the underside of the male's tail contrasts with the deep scarlet of the skin of the scrotum and perineum. This display of color is a form of courtship behavior, resembling the behavior of a peacock fanning his tail in front of the peahen. The males also shake objects such as the uprooted tree stumps scattered around the corral, and they leap rhythmically up and down and drum their feet. The shaking, leaping and drumming seem to serve a courtship function, since the males that do them most frequently mate most frequently.

As time passes, the males begin to groom the females more often than they do during the nonmating season, and they often fight among themselves and threaten or attack the females. Far from being repelled by this behavior, the females become increasingly attracted to the males until mating pairs form. Not all the animals pair, however, and among those that do the relationship is somewhat casual. The male and female may stay together for a few hours, seldom longer than a few days, and then the pair breaks up; one partner or both may enter into a new pairing. Most of the animals mate with several different partners in the course of the breeding season.

Dominance rank is not correlated with mating either in the males or in the females. Arrowhead, the alpha male, is one of the more active males, but Big V, one of the lowest-ranking males, has mated much more frequently in the years we have observed the two. High-ranking females also mate with low-ranking males in an apparently similar random fashion. The male takes the active role in courting, but the ultimate choice of partner appears to be made by the female. Over the years we have seen a number of females refuse Arrowhead simply by not standing still when he indicated (by pushing lightly on the female's back) that he was ready to mount. In each of these cases Arrowhead followed the female for a few days to a week before turning his attention to a more willing partner. Meanwhile the original female might have mated with several other males she apparently preferred to Arrowhead. Because we do not have a reliable paternity test we do not know which males are making the greatest contribution to the gene pool, but a preliminary data analysis indicates that the dominant males were not involved



SEASONALITY of certain kinds of macaque behavior is striking: male aggressive attacks (*color*), mating (*gray*) and male courtship displays (*black*) are largely confined to fall and winter.



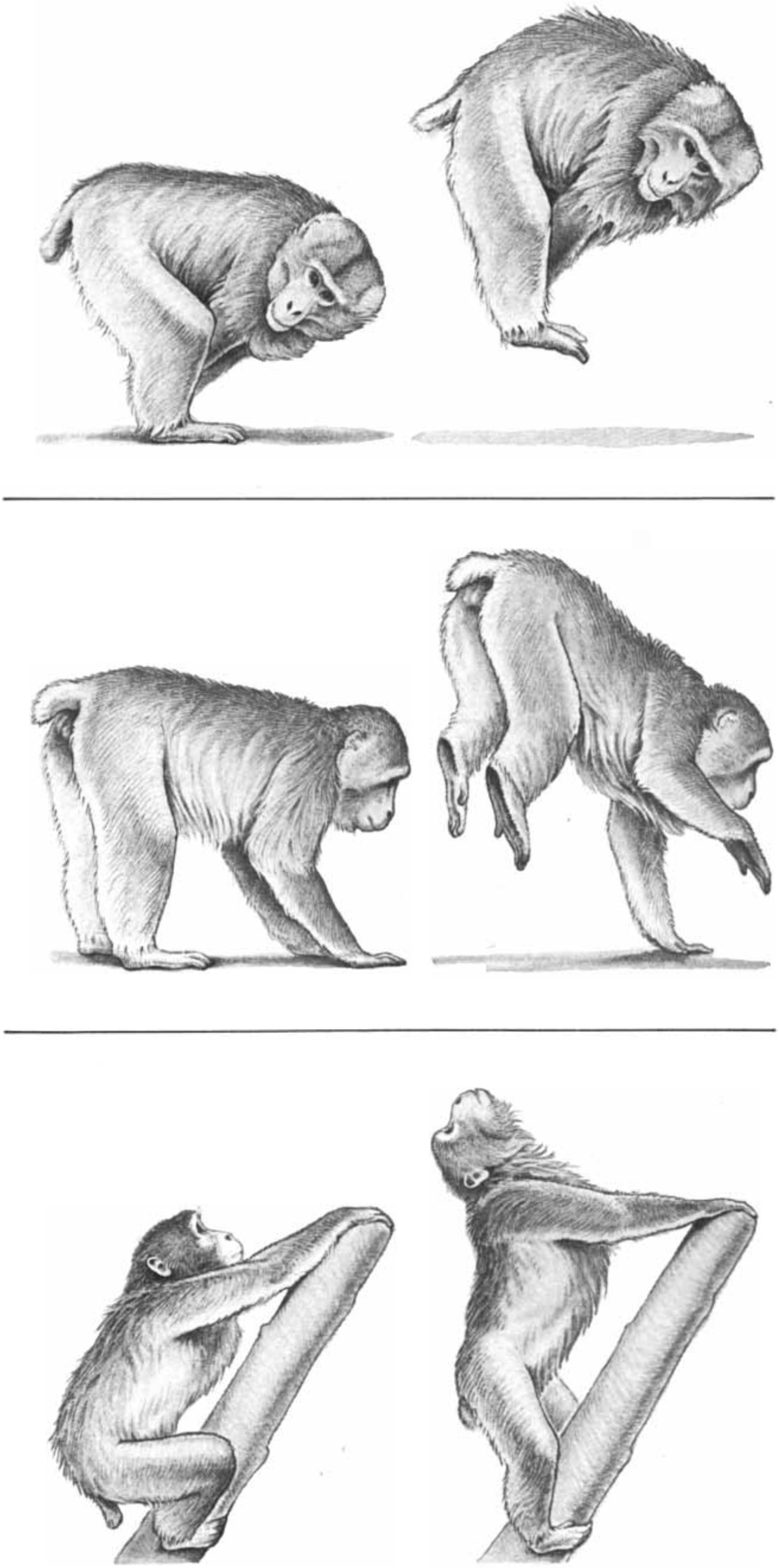
LEVELS OF MALE SEX HORMONE (testosterone) in the blood of male macaques (*gray bars*) are not correlated with dominance rank or mating frequency (*colored bars*). Although high hormone levels may lower the threshold for aggressive behavior, social and developmental factors are more important than biological factors in determining the macaque social order.

in a disproportionately large number of matings about the time of conception.

The changes in male aggressive behavior during the mating season are presumably caused by an increase in the male sex hormone testosterone. The relation between testosterone and aggression is complex, however, and is subject to modification by social factors. Although we have not measured the hormone throughout the year, we sampled every adult male at the beginning of two separate mating seasons and compared the measurement with the dominance rank of the animal. To our surprise there was no correlation between a male's dominance rank and his testosterone levels. We concluded that although testosterone may help to lower the thresholds of aggressive reactions during the mating season, it is less important than social stimuli in eliciting masculine patterns of aggressive behavior.

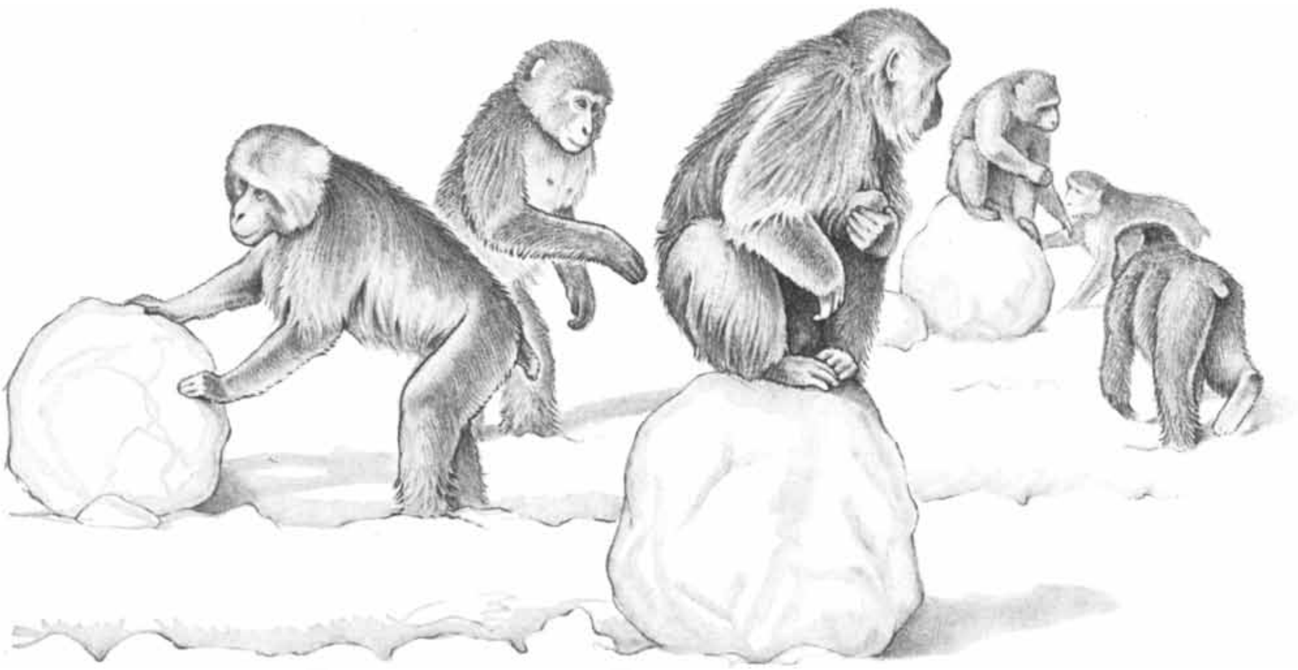
The end of the mating season is signaled by the cessation of pairing and by play activity among the adult males. The same monkeys that viciously attacked one another during the mating season now spend much time wrestling and tumbling with one another and with juvenile males. It is amusing to see one of the large males lying on his back waving his legs in the air while a juvenile jumps up and down on his belly and then executes a mock diving attack on his throat. The adult males also show considerable parental concern at this time and not only play with the juveniles but also groom them, defend them, huddle with them and carry them around on their backs. According to Junichiro Itani of Kyoto University, this paternal care, which is unrelated to dominance rank in our troop, is displayed only by high-ranking males in the wild Takasakiyama troop in Japan.

The difference is not necessarily related to the confinement of our troop but may simply reflect protocultural differences. Shunzo Kawamura of Osaka City University, who has studied protocultural behavior in free-ranging troops in Japan, has found that different troops have different diets. The Shodoshima-O troop, for example, normally feeds on rice being grown by farmers, whereas the Takagoyama troops never pillage rice paddies even though they pass through them in the course of their wanderings. Kawamura also observed a young female of the Koshima Island troop washing sand off sweet potatoes in a stream that runs through the island beach. The habit then spread to other members of the troop, until all the monkeys began to wash the sweet potatoes in the sea, apparently to season them.



MALE COURTSHIP DISPLAYS occur primarily during the fall and winter mating season. Individual males tend to prefer one type of display and perform it repetitively on a stump, a log or the ground. The most common display is leaping (*top*), followed in frequency by kicking (*middle*) and tossing (*bottom*). Some macaques combine kicking and tossing in a single motion.

Protocultural differences in macaque courtship have been recorded by Gordon R. Stephenson of the University of Wisconsin, who found significant dif-



SNOWBALL-MAKING is a protocultural behavior observed in the Oregon troop. It first appeared on January 13, 1971, when Big X, a low-ranking male, made a small snowball while eating snow and proceeded to roll it until it had reached a diameter of about one and a half feet. Every winter since then he and other macaques have constructed several large snowballs that have become centers of attention for the troop. Although unconnected with aggressive or feeding behavior (unlike most tool use in animals), the improvisation of snowballs and other objects may serve an adaptive function by preparing the macaques to adapt to new situations and a variety of habitats.

ferences in courtship behavior among the Miyajima, Arashiyama and Koshima troops. Female macaques are known to court males by mounting them, yet the three troops differed in this behavior: only low-ranking females mounted males in the Miyajima troop, all females did so in the Arashiyama troop and no female was ever observed to mount a male in the Koshima troop. In our troop only a few females mount males, and they do so without relation to their rank or the male's.

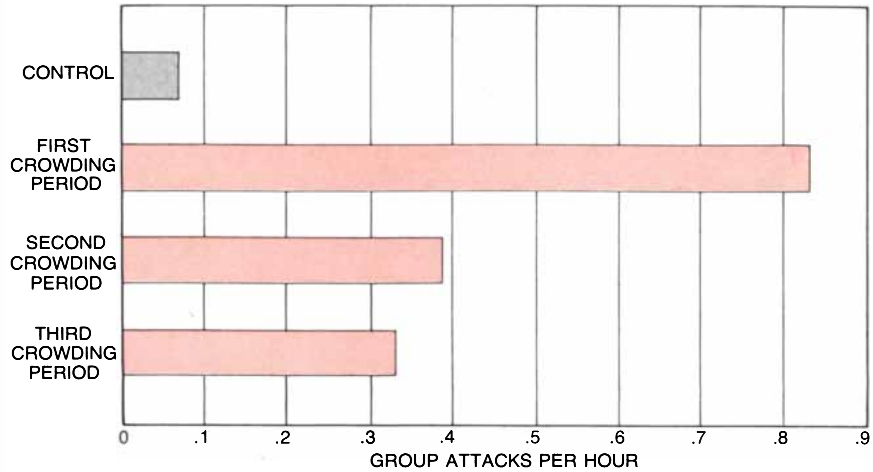
We have observed an unusual form of protocultural behavior in our troop. In the winter of 1970–1971 one of the peripheral males, Big X, rolled a snowball along the ground until it was about a foot and a half in diameter. Every winter since then other macaques have made similar snowballs; they have been centers of attention for infants and juveniles, who play on them, and for adults, who simply sit on them. The snowballs do not seem to have any functional significance for the macaques but appear to be the product of their manipulative skill, intelligence and leisure time.

Since the fence around our corral prevents a monkey's escape from an aggressor that is chasing it, the frequency of aggression is undoubtedly higher than it would be in the wild. We have tried to provide a form of escape by placing small steel drums on their side in the corners of the corral, bolting them to the ground and fitting them with a handle on the inside to which the aggressee can cling if the aggressor tries to pull him out. These barrels are often used as hiding places by the low-ranking males and seem to effectively stop group attacks.

Occasionally many members of the troop attack one animal. Why they do it is a mystery. The mobbing does not seem to be caused by crowding. Bruce Alexander, who is now at Simon Fraser University, looked into the matter by crowding our troop into a pen 50 feet square adjacent to the corral. He recorded the type and frequency of aggressive interactions during three crowding periods in the pen and three control periods in the corral. Surprisingly the crowding did not result in a breakdown of the social structure of the troop; indeed, the dominance hierarchy was stabler in the pen than it was in the corral. Mobbing and severe aggression increased during the crowding periods, but their frequency decreased over the span of the three crowding periods. This suggested that the increased frequency of mobbing observed under crowded conditions was due more to the removal of the animals from a familiar habitat than to increased density. Alexander's hypothesis is supported by the fact that our troop's initial move from an 1,100-square-foot holding pen into the two-acre corral was immediately followed by an increase in severe aggression that lasted for three weeks and left one low-ranking male dead and three others badly wounded.

My own research on the troop is concerned with the effects of increasing population density on both reproduction and aggression. I am following a procedure similar to the classic one of John B. Calhoun of the National Institute of Mental Health, who allowed populations of laboratory rodents to reproduce in a confined space until they developed abnormal patterns of behavior that invariably led to a disastrous decline in population [see "Population Density and Social Pathology," by John B. Calhoun; SCIENTIFIC AMERICAN, February, 1962]. We too have allowed the population of our macaques to grow without restriction within the confines of their corral.

Our troop gives no indication of developing abnormal patterns of behavior in spite of its rapid growth from the original 46 members captured in 1964 to 230 in 1976. In fact, we recorded a significant decrease in aggressive encounters among the adult males between the 1971 and 1972 mating seasons, when the population of the troop increased by 17 percent. (The number of adult males remained stable at 21.) A more detailed analysis showed that the



GROUP ATTACKS on a low-ranking monkey, involving up to 35 animals and often resulting in severe wounding or death, greatly increased in frequency when the Oregon troop was crowded into a 50-foot-square pen in an experiment conducted by Bruce Alexander. Over three sequential crowding periods, however, the frequency of mobbing decreased, suggesting that severe aggression was due largely to unfamiliar environment and not to crowding per se.

higher frequency of aggression during the 1971 season was associated with a change in the hierarchy that involved the six highest-ranking males, whereas in 1972 the hierarchy had stabilized. It therefore appears that our macaques are unlike rodents in their response to increasing population density because their social structure is a critical factor in regulating their aggressive behavior.

On the other hand, even though the macaques are confined in a space more than 100 times smaller than their original home range, which was probably a minimum of 250 acres, the crowding may not yet have reached the pathological limit. Daniel S. Stokols of the University of North Carolina has distinguished between the physical condition of density, defined purely in terms of spatial parameters, and the experience of crowding, a motivational state aroused through the interaction of spatial, social and individual factors. With this distinction in mind I have observed the Arashiyama West troop of Japanese macaques, which is confined in a 107-acre enclosure near Laredo, Tex. There I found that although the 150 troop members move through a minimum of eight to 10 acres a day when they are feeding and cover the entire 107 acres every few days, at most times of the day or night the entire troop appears to be grouped in a space of less than two acres. It is therefore possible that our troop is not yet "crowded" but is only limited in range.

It seems unlikely, however, that the corral population can continue to increase at the present rate without eventual pathological changes in the animals. Deterioration in social structure has been reported to precede troop fission in the Gagusan troop in Japan, which has maintained its size between 130 and 160 by frequently dividing. Such fission may be evidence that the sensitivity of the macaque social structure to increased numbers evolved as part of an intrinsic mechanism for population control. Alexander has speculated that once a threshold population level has been reached territorial interactions within the troop will lead to fission without a decrease in fertility.

We plan to test this hypothesis in the near future. We have just finished a new corral adjacent to the original one and shall soon connect the two by a runway in order to determine whether fission will occur when rising population density threatens the stability of the social order or whether the troop will simply expand its home range. If fission does take place without any indication of reduced fertility, it will suggest that Japanese macaques have evolved population-control mechanisms fundamentally different from those of rodents.

The fission of our troop, if it occurs, will also provide us with a unique opportunity for studying the formation of a new social order. Which families will emigrate? Which males will be subleaders? Which will be the alpha male? What effect will the old dominance relationships have on the new social order? What will happen to the old social order if and when its key members leave?

Beyond the intrinsic interest of studies of primate societies, they may shed some light on the vastly more complex structure of human social behavior. To survive is to adapt, and man's understanding of the macaques' successful adaptation to many different environments may ultimately contribute to his own survival.

Dust Storms

Although in many parts of the earth they are rarely seen, they are a major dynamic feature of the planet, changing its surface and having a heavy impact on human affairs

by Sherwood B. Idso

A motorist driving between Phoenix and Tucson on Interstate Route 8 may encounter a sign that reads: "Blowing dust, reduce speed." The changeable-message signs that carry the warning are spaced about five miles apart on the highway, which passes through a corridor where dust storms arise frequently and sometimes reduce visibility so severely that collisions result. A dust storm is always an awesome meteorological phenomenon; as the signs suggest, it can also be a dangerous one.

Various processes inject dust into the atmosphere. Probably the most dramatic process on the earth is volcanism. The eruption of a large volcano thrusts vast amounts of debris high into the stratosphere. At that level strong winds rapidly spread the dust over the entire globe. Some investigators who have studied the phenomenon believe such events are responsible for several periods of significant cooling of the earth's climate.

Human activities account for much of the dust that is injected into the lower atmosphere. Urban areas with concentrations of industry are major sources of dust. In rural areas too man plays a role through slash-and-burn methods of clearing land and through farming practices that expose the land to erosion by the wind. Whatever the source of the dust, wind is the common factor that lifts the dust particles from the ground and moves them through the air. Indeed, the dictionary definition of a dust storm is "strong winds bearing clouds of dust."

The earth is not the only planet with strong winds and dust. When *Mariner 9* approached Mars late in 1971, the entire planet was shrouded by a blanket of dust raised by a storm. Peter J. Gierasch and Richard M. Goody of Harvard University have proposed that such vast Martian dust storms are caused by a self-sustaining mechanism that converts solar energy into wind energy on a large organized scale, as in a terrestrial hurricane. Mars has localized dust storms as well.

The study of dust storms is probably very old, since they have played a crucial role in man's physical and economic well-being in many parts of the world. On the most direct level of interaction a dust storm can threaten life itself; at times both people and animals have died of suffocation during a severe dust storm. For example, a major dust storm in 1895 is reported to have caused the loss of 20 percent of the cattle in eastern Colorado.

Dust storms are also known to spread pathogens that are harmful to people, animals and plants. In the U.S. Southwest the parasitic protozoon *Coccidioides immitis,* which causes the human disease coccidioidomycosis (commonly known as valley fever), is spread far and wide in this way. Animals and occasionally people are afflicted throughout the world by dust-borne *Cryptococcus neoformans,* a yeastlike organism that attacks the central nervous system, the lungs and the skin. As for plants, dust not only carries pathogens but also causes lesions that give the pathogen entry to the plant.

The most direct threat of dust storms to man is probably soil erosion. The tendency of the wind to erode the soil is often aggravated by agricultural operations, although various kinds of evidence indicate that the Great Plains region of the U.S. was periodically subjected to tremendous dust storms long before the introduction of current methods of farming and ranching. Nevertheless, the tendency of agriculture to increase the hazard has led to programs in the U.S. and other countries for studying the characteristics of dust storms and for finding ameliorative countermeasures. At Big Spring, Tex., which has seen many huge dust storms, a team of workers with the Agricultural Research Service of the U.S. Department of Agriculture has developed a method for predicting at the beginning of a year the number of days during the year when blowing dust can be expected; the method is accurate to within five days 72 percent of the time. The system gives farmers an opportunity to take preventive action. At present the primary method of controlling erosion by wind is to maintain a surface cover, which is usually fairly effective. If rainfall is below normal for two years and the

DUST STORM bears down on the community of Big Spring, Tex., in September, 1930,

amount of crop residue produced is small, tillage may be required as an emergency measure. The plow produces a rough, cloddy surface that tends not to lose soil particles because they collect in the cavities formed by the clods.

The advent of meteorological satellites has afforded an expanded view of dust storms. The satellites reveal great integrated patterns of dust movement and also tremendous dune fields that have clarified the behavior and mechanisms of dust storms.

The smallest basic dust-storm unit of meteorological significance is probably the kind caused by a downdraft of cool air from a single cumulonimbus cloud. When such a cloud has developed to the point where rain begins to fall from it, the air is cooled significantly as the precipitation passes through it and either partly or wholly evaporates. Since this air is cooler than the surrounding air, it is also heavier, so that it descends in a downdraft having a velocity that is roughly proportional to the height of the cloud top. As the heavy cooled air hits the ground it is deflected forward and moves out in front of the cloud in a large, tongue-shaped pattern, flowing along the ground as a density current. If the ground is dry and covered with loose material, much surface debris will be swept up by the turbulent, churning head of the flow.

A dust storm of this type that forms over a desert during a period of convective instability is called a haboob, a term derived from the Arabic word for violent wind. Haboobs arise frequently in the Sudan, where around Khartoum they are experienced some 24 times a year. Being associated with the rainy season in the Sudan, they generally form between May and October, with a slight maximum in June and July. The amount of dust moved by haboobs in the Sudan is remarkable: in two months from 12 to 15 feet of material can pile up against an object that is exposed to the fury of these storms.

Quite similar dust storms appear in the U.S. Southwest, where they have come to be known as American haboobs. At Phoenix the frequency is about 12 a year. Much of the research on dust storms in the U.S. is done in Arizona.

As in the Sudan, haboobs are most frequent in Arizona during the rainy season (July and August). They occur with some regularity once the state is immersed in the tropical maritime air characteristic of the summer monsoon. During the second half of June upper-level moisture begins to move into Arizona from the Gulf of Mexico. For a few weeks it alternately advances and retreats, much like an ocean tide. After the first week of July, however, it is fairly well established. Then it is augmented at lower levels by periodic surges of moist tropical air from the Pacific that push up the Gulf of California. These surges often generate long, arcing squall lines, with haboobs fanning out below and ahead of each cell in the line. The

which was during the period when the Southwest was known as the Dust Bowl because of its numerous dust storms. The southern plains in particular were hard-hit by dust storms, which carried off large quantities of topsoil and thus degraded the region's agriculture.

individual outflows often merge to form a seemingly solid wall of dust that stretches for hundreds of miles.

Satellite photographs have shown that storm systems of this type often develop south and east of Tucson from dense masses of clouds over the Sierra Madre Occidental of northern Sonora in Mexico. The cloud masses form almost explosively just after noon, starting out as isolated cells that move slowly northwest. Thunderstorm activity seems to be generated in the convergence of moist air lifted by rising wind on east-facing slopes in the late morning and on highly heated west-facing slopes during the afternoon. At this time new cells usually develop and form an arcing squall line near Tucson. The systems generally intensify as they move up the Santa Cruz Valley toward Phoenix. Clouds in the line often rise to 40,000 feet and not infrequently go to 55,000 feet. The dust raised by the density currents in front of them rises all the way to the cloud base, which can range from 8,000 to 14,000 feet above the ground.

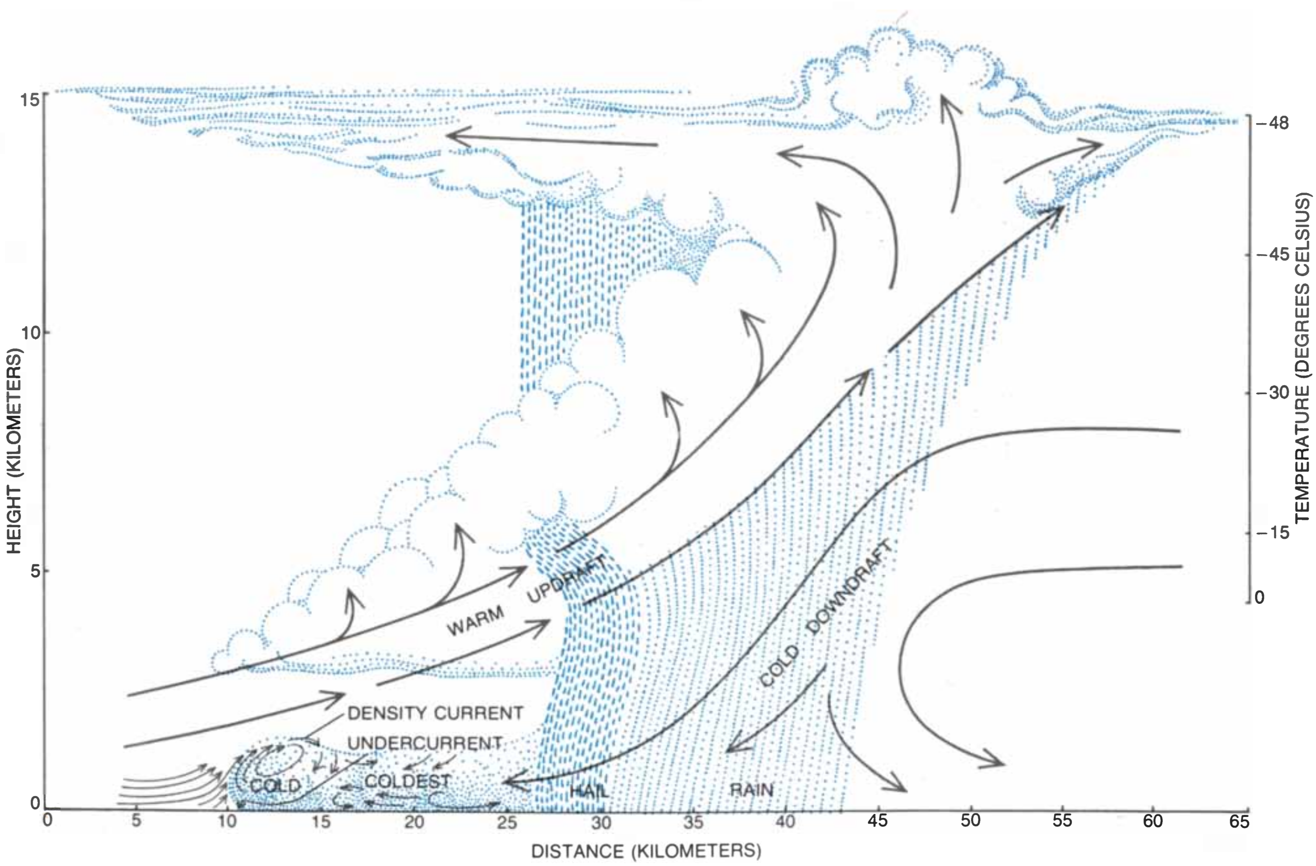
Haboobs generally have a mean speed of advance of 30 miles per hour. Maximum gusts within a haboob, however, are about 60 miles per hour. With the arrival of a haboob the relative humidity usually rises and the air temperature drops. Almost instantaneous temperature reductions of 15 degrees Celsius (27 degrees Fahrenheit) have been measured in a few severe storms as the front of dust arrives. Ordinarily, however, the mean reduction is only about half that amount.

Although the visibility can drop to zero in such a storm, the average is a quarter of a mile. After the storm it takes about an hour for the visibility to return to six miles; following a severe storm it can take as long as three hours. If the parent thunderstorm arrives behind the dust storm, its precipitation will clean the air effectively. Often, however, the trailing thunderstorm does not arrive or the precipitation evaporates before it reaches the ground, and then the dust can linger for hours or even days.

The leading edge of a haboob is always changing in its appearance. According to T. J. Lawson of the University of Reading, who has studied haboobs in the Sudan, "a push of cold air from within the body of the haboob will cause the formation of a bulge or lobe, which appears to grow forward at a slightly faster rate than the average speed of the front. The lobe will expand vertically and horizontally until, when its forward movement relative to the front has decreased, irregularities will start to appear in its leading edge, causing any fresh surge of air moving up from behind to flow to one side. In this way new lobes are continually appearing out of the dying stages of a previous one."

I have often seen two adjacent lobes expand simultaneously and engulf a large volume of intervening warm air. Expanding cold lobes also overrun a certain amount of the warmer air in front of them. Both movements can create unstable situations that are conducive to intense convection, which can give rise to small whirlwinds. Although the whirlwind thus formed is short-lived, it can damage buildings in its path. Such tornadic vortexes form not only inside dust storms but also in front of them, where warm air is pushed up over the nose of the advancing front of cold air.

Next on the size scale of dust storms is what one sees from the perspective of a satellite. Photographs from satellites have revealed much information about the extent of dust storms and



GENERATIVE MECHANISM of a dust storm of the squall-line type is depicted schematically. Rain and hail produce a downdraft of cold air that spreads out along the ground and moves forward to form a density current, that is, a body of moving air that is heavier than the surrounding air because it is cooler. The current picks up loose dust and sand from the surface. Moreover, the heavy cold air pushes up the warmer and lighter air it meets, thereby reinforcing the warm updraft that creates new rain clouds that renew the cycle.

about where they originate. In particular A. A. Grigor'yev and V. B. Lipatov of the Leningrad State University have been able to locate five major regions in Africa and Eurasia where dust storms are generated.

The first of these regions, which is in central and western Africa, is often characterized by extremely long dust currents over the southern Sahara. They appear to result from an enormous airstream moving east-northeast over the sandy deserts of Mauritania and Niger. Giant dust bands 2,500 kilometers long and 600 kilometers wide often move across the area with the cold fronts.

Some large systems have even carried this dust across the Atlantic to the eastern coast of South America. The masses of dust thus transported are raised originally when strong northeasterly trade winds (the local term for them is harmattan) meet the colder southwesterly monsoon. This convergence can occur as far north as 20 degrees north latitude in summer and as far south as six degrees north latitude in winter. The dust is raised to high altitudes, where westward-moving wind currents transport it across the Atlantic.

ADVANCING DUST STORM has a characteristic structure of lobes along its front edge. This type of storm is called a haboob, a name derived from the Arabic term for violent wind. The photograph, made in Phoenix in 1971, shows what is known as the great Arizona haboob.

The second major region where dust storms originate is the southern coast of the Mediterranean. Here the storms begin with the passage of cold fronts connected with low-pressure troughs that stretch toward North Africa from western Europe. Deep cyclones often form in the troughs. The very hot, dry winds that carry clouds of sand ahead of these eastward-moving depressions are known by a variety of local names.

The third major region is the northeastern Sudan. In addition to the haboob type of dust storm that is prevalent just south of this area, Grigor'yev and Lipatov have described vast dust storms that form when cold northwesterly air currents meet hot southwesterly monsoons. At such times masses of dust may be raised over large areas encompassing the region of the northeastern Sudan from the Nile River to the Red Sea. High-altitude winds often transport this dust across the Red Sea to the Arabian Peninsula.

That peninsula is the fourth major dust-storm region. Its storms form in a way that differs somewhat from the mechanisms in the first three regions. In the Arabian Peninsula they result from contact between two formations that have different barometric pressures; the merger gives rise to large storm areas. Such storms generally develop when the wind increases on the western periphery of a pressure depression centered over southern Iran. Cone-shaped streams of airborne particulates form and move southward, often expanding. At their inception, which Grigor'yev and Lipatov ascribe to whirlwinds, several small parallel streams, from three to five kilometers in size, develop. They stretch for about 100 kilometers and then merge into more powerful streams. The larger bands, swirling and expanding, stretch up to 500 kilometers. These streams, moving with the trade current, are carried through a corridor bounded on the north by the mountains of the southern edge of Asia Minor and Iraq and on the south by the plateaus and mountains of Saudi Arabia.

The last of the five dust-storm regions studied by Grigor'yev and Lipatov encompasses the lower Volga and the northern Caucasus. The dust storms here are also of the storm-zone type. They are typically caused by an increase in the gradient of barometric pressure on the northwestern periphery of high-pressure ridges that extend to the Volga and central Asia. A gradient increase of this kind is caused by the movement of cyclones into northern Europe and the spread of low-pressure troughs to the south of the cyclones.

Dust storms of similar size over central China have been described by G. K. T. Ing of the University of Hawaii. They arise most often in the early spring, when the conditions of freshly planted soil, low rainfall and increased surface winds associated with cold fronts combine to create the conditions for dust to be blown. Most of the dust storms in China, however, originate in the vast barren land between 32 and 47 degrees north and 75 and 115 degrees east, which contains the great deserts Takla Makan, Gobi and Ordos and also the major loessial-soil lands of China. It is estimated that several thousand tons of soil is transported by wind each year in these areas, some of it going several thousand kilometers downwind and causing deep red sunsets as far south as Hong Kong.

Two different mechanisms are responsible for the main types of dust storm in China. In the first mechanism dust is lifted by winds behind a front. Storms originating in this way are generally of limited size, and dust from them is usually confined below an altitude of about 1.5 kilometers. In the second mechanism dust is lifted much higher (to about three kilometers) by vertical motion ahead of a front. This material is often transported across distances on the order of 2,500 kilometers. Meteorologically the conditions for the first mechanism are strong winds of low humidity blowing at the surface and at moderate heights above it; the second mechanism requires a vigorous dry front with strong upward motion ahead of it.

A considerable amount of knowledge about dust storms can be gained indirectly by studying the surface characteristics of deserts, such as the types of dune. Some of the factors that determine the size and form of dune bodies are known: the direction, strength and variability of the wind, the moisture content of the soil, the vegetation, the underlying topography and the amount of transportable soil accessible to the wind. Wind is by far the most important factor. The study of major sand seas in deserts therefore promises to shed con-

siderable light on the dust storms in those areas.

As a first step in this process Edwin D. McKee and Carol Breed of the U.S. Geological Survey have employed photographs made by satellites to develop a dune-classification system that can be applied throughout the world. It recognizes five basic patterns among groups of dunes: parallel straight or linear, parallel wavy or crescentic, star or radial, parabolic or *U*-shaped and sheets or stringers.

Parallel straight or linear dune groups are defined as bodies of sand in which the length of the dunes is much greater than the width. The area occupied by dunes is about equal to the area between dunes. Both sides of a dune are likely to be steep enough to give rise to avalanches. Dune patterns of this kind are found in the Simpson Desert of Australia, the Kalahari Desert of South Africa, the Empty Quarter of Saudi Arabia and the Sahara.

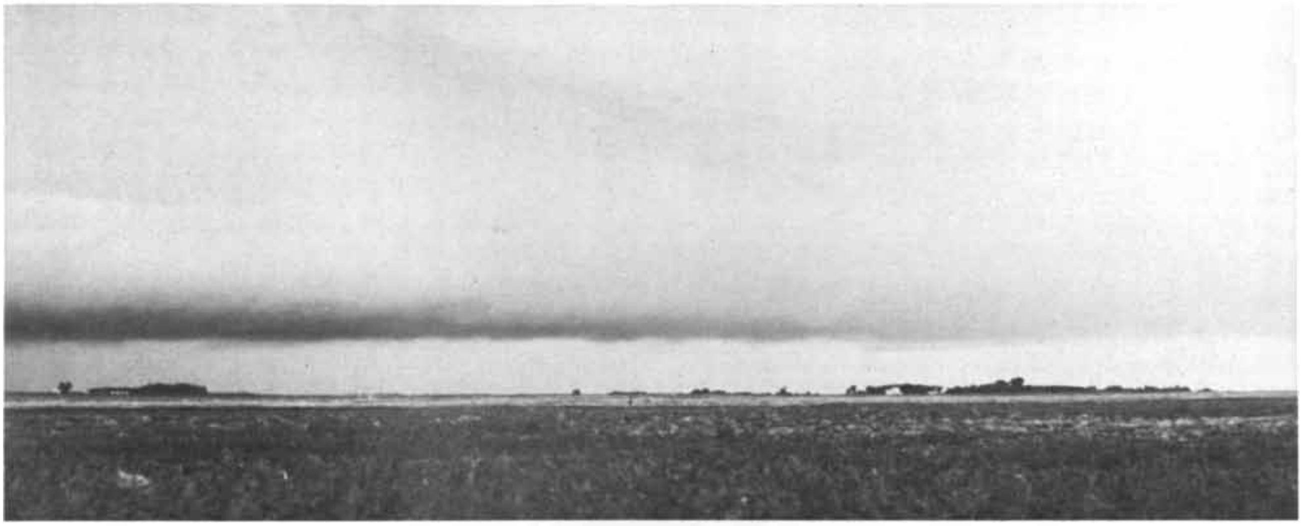
Parallel wavy or crescentic dune groups consist of nearly parallel rows of cuspate segments. They are best represented in the Kara Kum Desert of the U.S.S.R. and in the Nebraska Sand Hills of the U.S. Variants are found in the Great Eastern Erg of Algeria, parts of Saudi Arabia and the Gobi and Takla Makan deserts of China.

Star or radial dunes are composed of several arms that project out from a central cone in various directions; the dunes look much like giant pinwheels. They can be found scattered at random in parts of the Great Eastern Erg of Algeria and the Empty Quarter of Saudi Arabia. Chains of star dunes appear in other parts of Algeria and in the Gran Desierto of Mexico.

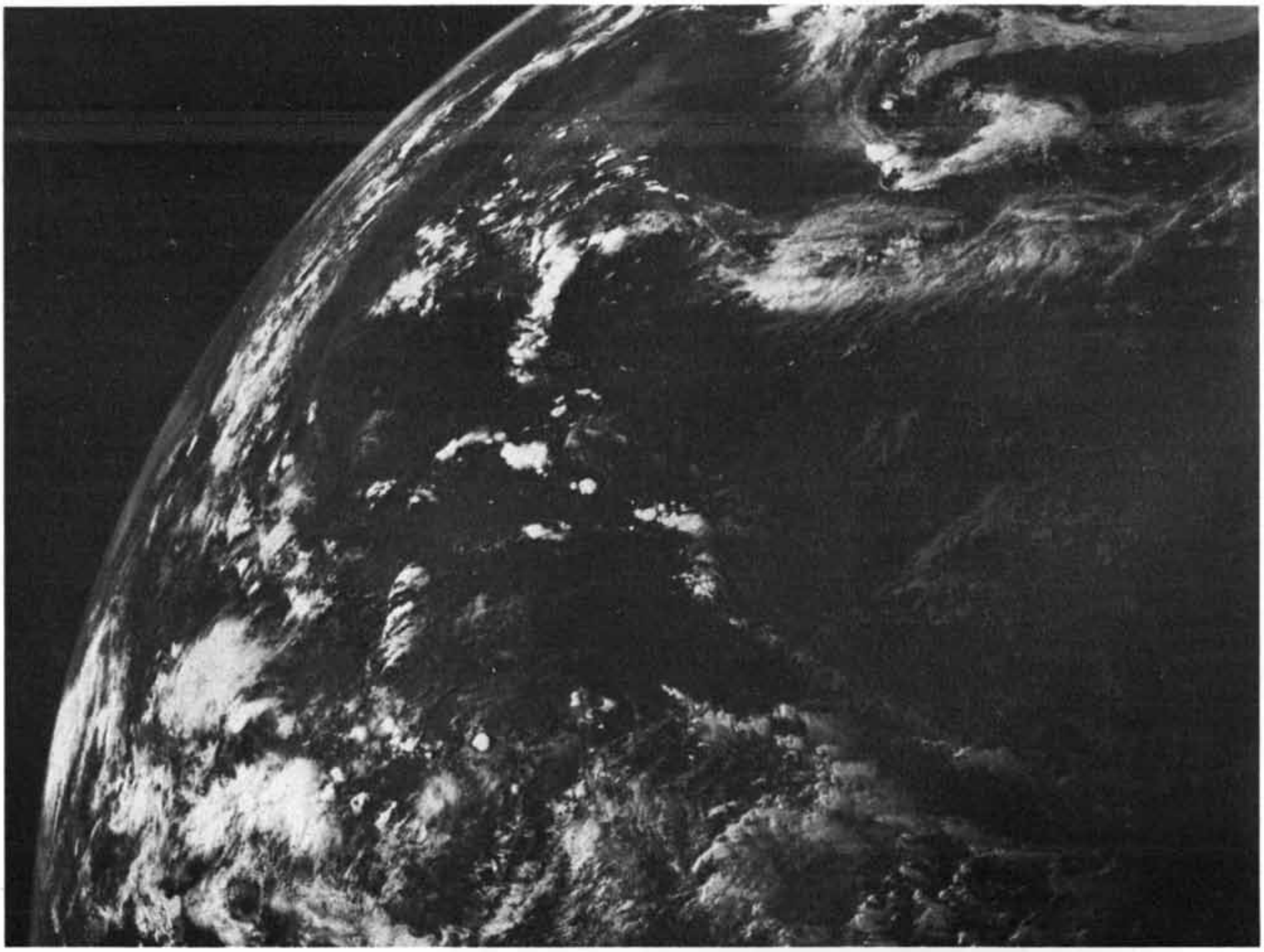
Parabolic or *U*-shaped dunes seem to develop in areas where vegetation or moisture or both tend to anchor the side arms while the center is blown out and its sand moves forward. This type of dune is found in the Thar Desert of Pakistan and at the White Sands Proving Ground in New Mexico.

Sheets or stringers are the terms employed to denote the type of relief in sand seas that lack a definite geomorphic form. Sand accumulates in flat sheets (as near Lima in Peru) or forms stringers downwind that do not exhibit any appreciable relief.

The basic work of dune classification has now been completed for the major sand seas of the world's deserts. The task of correlating the dune forms with patterns of wind and the characteristics of dust storms is still in progress. The work includes not only visits to deserts but also laboratory investigations. Observations of the surfaces and atmospheres of other planets such as Mars are also help-



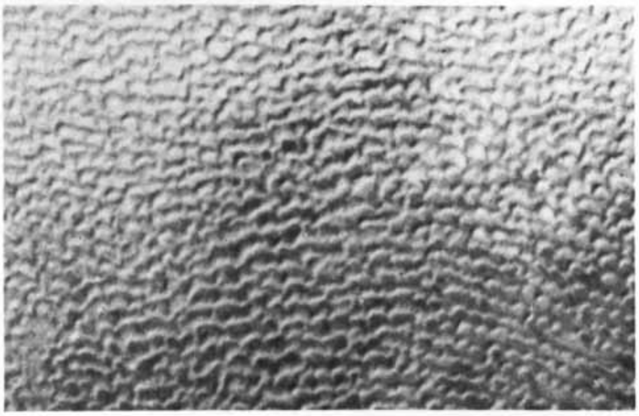
CONDITION OF SURFACE can determine whether or not a dust storm is generated when the meteorological conditions are right. Each of these photographs was made at a time when the meteorological conditions were identical: downdrafts of cool air created by precipitation were moving out in front of their parent thunderstorm as density currents. When the top photograph was made in Minnesota, the humidity was high and little loose material was available on the surface, so that there was no dust storm. The bottom photograph was made in Arizona at a time of low humidity; the surface was dry and bare. Loose surface material was swept up to create a dust storm.



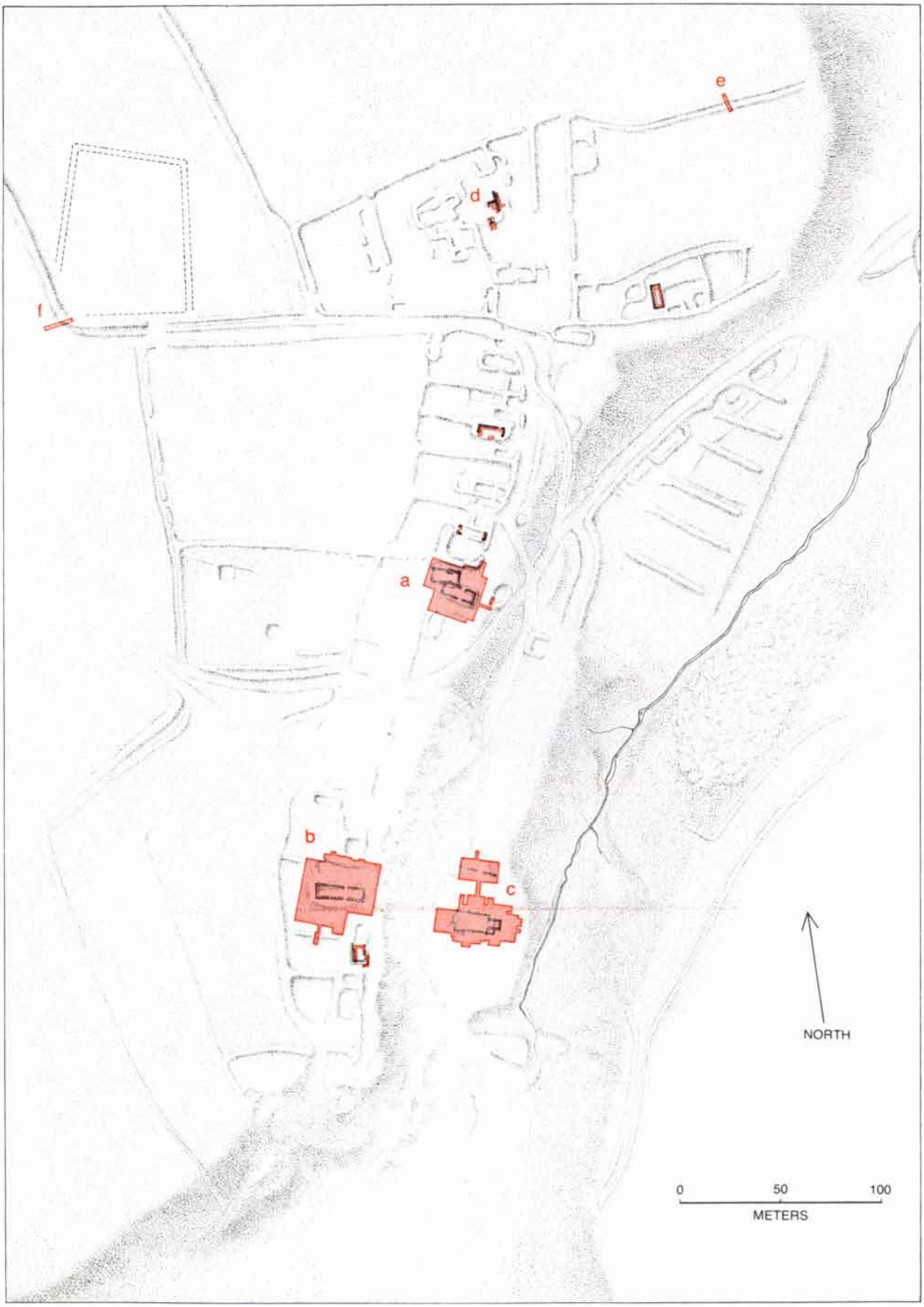
INTERCONTINENTAL DUST STORM appears in the eastern Caribbean in a photograph made from a weather satellite. The storm originated over the Sahara and was blown westward. The storm is the bluntly pointed gray area that takes up much of the right center portion of the photograph. Just below the dust pattern is a group of scattered white clouds delineating the northern part of South America.

ful. Together all three approaches are beginning to elucidate dust-storm patterns of considerable interest.

Although a good deal of information about dust storms is now in hand, much remains to be learned. The characterization of large-scale systems from satellite photography is only in its infancy. After characterization one can hope to be able to make reasonable predictions of where and when dust storms might arise. More work also needs to be done on small-scale features that are important in understanding soil erosion and localized but intense vortexes that often attend the passage of the frontal boundaries of density currents. Significant discoveries, important to plant, animal and human ecology, remain to be made concerning the spread of biologically active agents by dust storms.



SHAPES OF DUNES indicate patterns of wind and hence aid in the study of dust storms, since wind is the most important factor in a dust storm. These dune patterns were photographed in Saudi Arabia by an Earth Resources Technology Satellite. According to a classification system developed by Edwin D. McKee and Carol Breed of the U.S. Geological Survey, dune types shown here are parallel straight or linear (*left*) and parallel wavy or crescentic (*right*). Other patterns are radial, *U*-shaped and largely featureless sheets or stringers.


a
b
c
d
e
f
NORTH
0
50
100
METERS

A Deserted Medieval Village in England

In the final 200 years of the Middle Ages the rural population of England declined, and one village in every 10 was deserted. Excavations at one such site reveal the fabric of medieval life

by Maurice Beresford

To an American visitor the thatched and half-timbered cottages of an English village seem to be colorful relics of the Middle Ages. The fact is that not one medieval peasant dwelling has survived anywhere in England. Moreover, it is only over the past few decades that archaeologists have learned much about the structural plan of the dwellings that sheltered the large majority of the English people during the 1,000-year span of the Middle Ages.

To be sure, several thousand parish churches in England are medieval, if not in their entirety, then in part; so are many castles and monasteries and several hundred manor houses. These, however, were not the homes of peasants. When the medieval population of England was at its peak, just before the sequence of plagues that began with the Black Death of 1349, the number of villagers exceeded four million. A typical village consisted of 30 or 40 households, and each household numbered four or five individuals. The peasant dwellings were gathered around a green or along one or two streets; a church and a manor house usually formed part of such a nucleated group.

The majority of English village sites have been occupied continuously since the Middle Ages. Why, then, have no medieval village houses survived? The question is particularly pertinent because the archaeological evidence shows that the peasantry repaired and rebuilt their dwellings surprisingly often. The answer is that by the time the medieval period came to a close toward the end of the 15th century, economic circumstances in England were changing substantially. The massive rural rebuilding that began toward the middle of the following century differed significantly from the building and rebuilding of the previous millennium.

The villagers who rebuilt so drastically in the mid-16th century belonged to the upper stratum of the peasantry: the lucky and successful few who had survived a combination of plague and economic misfortune. Many of their fellow villagers were dead, and many of the less fortunate either found their landholdings shrunk to a few acres or, what was worse, were reduced to the status of landless laborers. Meanwhile the more fortunate peasants were enjoying the benefits of fixed rent in a period of rising demand for farm products. Increasingly they took to calling themselves yeomen, that is, freeborn men, a social rank just below the lowest level of gentle birth, designated esquire. The yeomen invested their new wealth in a great rebuilding that transformed their domestic environment.

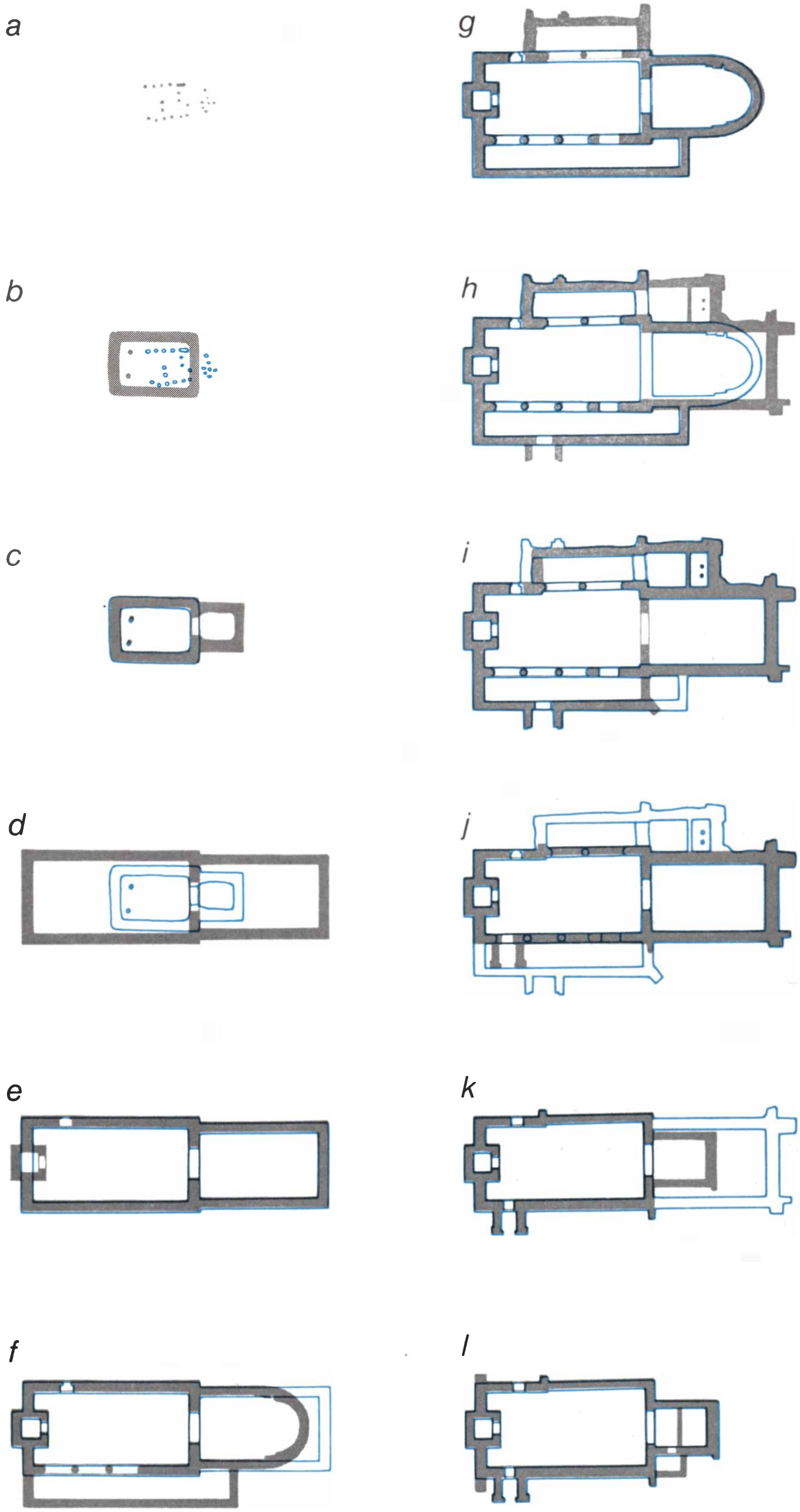
Brick was now available to supplement the two former principal construction materials: timber and local building stone. The number of rooms per dwelling increased, and two-story construction allowed for upstairs bedrooms. Farm animals that had commonly shared the same roof as their owners were now housed in separate structures. Many of these ambitious rebuildings in the years between 1550 and 1650 still survive. One example is the thatched "cottage" (in reality a substantial yeoman house) at Shottery, near Stratford-on-Avon, where Shakespeare's wife, Anne Hathaway, was born; others may be seen in any tourists' "olde English village," but they all are definitely not medieval.

Where, then, is one to look for a medieval peasant dwelling? Many a modern village house must conceal under its floor the remains of a predecessor going back to the Middle Ages, but the owners cannot be expected to encourage a search for such archaeological traces. Similarly, the postmedieval structures of many churches probably overlie earlier foundations, but again these are not available for examination except when a substantial program of repair or extension is undertaken. Fortunately for the archaeologist another major economic change in England, spread over the century that culminated in the yeomen's great rebuilding, has preserved for potential examination medieval villages by the hundreds.

That change is what historians call the desertion of the villages: a literal failure of about one English village in 10 to remain economically viable beyond the late 15th century. This was a time when the English cloth industry's demand for wool was on the rise. One consequence was enclosure, that is, the conversion of arable land into pasture for sheep raising, and enclosure certainly added to the small number of villages emptied by the plagues of the 14th century. Hugh Latimer, bishop of London under Henry VIII, noted: "Where there have been many householders and inhabitants, there is now but a shepherd and his dog."

Today some 3,000 deserted medieval villages are known throughout England; in certain counties in the Midlands as many as one village in six was deserted.

DESERTED MEDIEVAL VILLAGE in Yorkshire, Wharram Percy, is shown in the plan on the opposite page. Major "open" excavations are the site of an early manor house (*a, center*) built late in the 12th century and soon abandoned, a peasant "long house" and its surrounding farmyard (*b*) and a church and its adjacent vicarage (*c*). The church, which once served a five-village parish, was used until the 1870's even though four of the five villages, Wharram Percy included, had long been deserted. It stood in ruins when the author and his associates began to excavate the site in 1950. Not all excavations at the site, which include the unearthing of a mill dam, are shown here, but among the other major finds were a second manor-house site (*d, top center*), a deposit of Saxon potsherds at least 1,100 years old (*e*) and traces of a boundary ditch (*f*), evidence of a Romano-British farm. In the ditch was a second-century Roman coin.



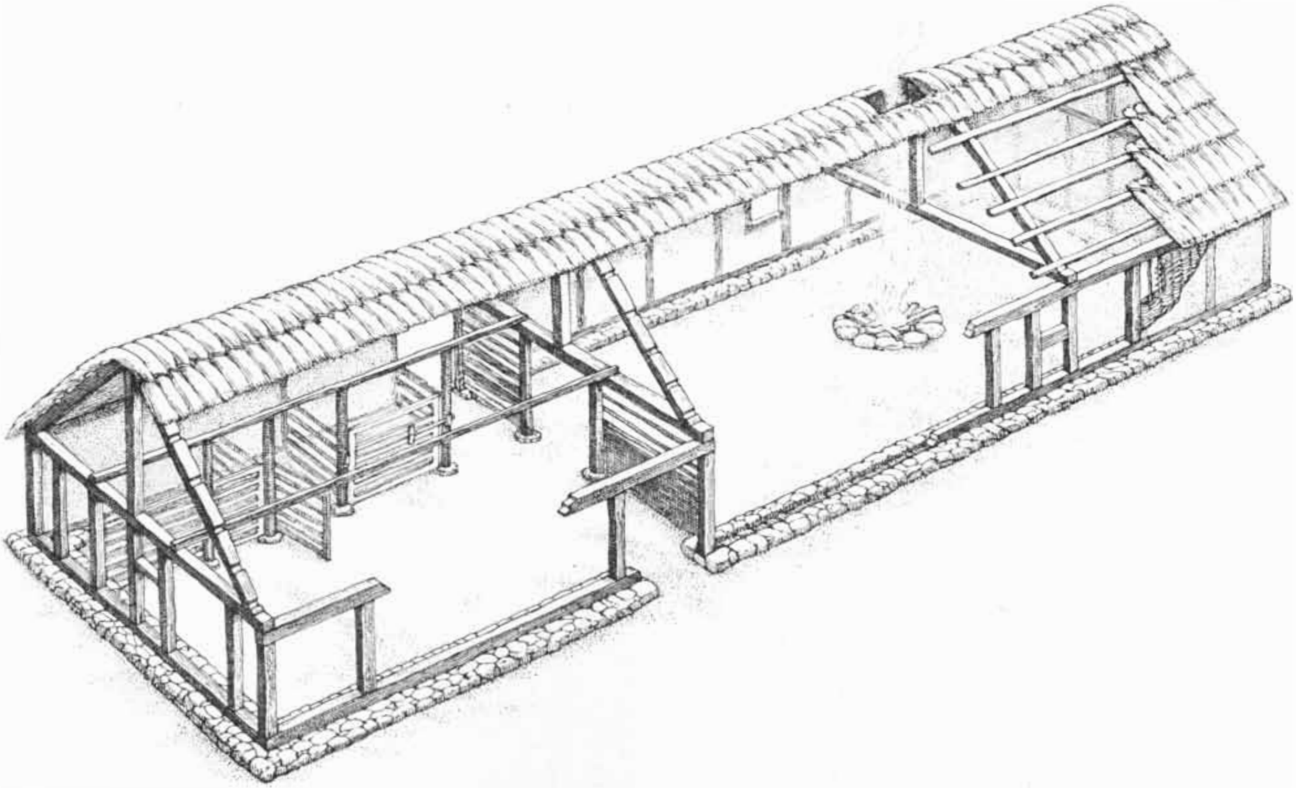
EXPANSION AND CONTRACTION of the church at Wharram Percy reflect the increase and decrease of the rural population of England during the Middle Ages. At first a modest structure of wood (*a*), perhaps built as long ago as the eighth century, it was rebuilt in stone (*b*) and then twice enlarged (*c, d*) before a tower was added at its west end (*e*), perhaps by the Normans. When the Percys became lords of the manor, the apse was remodeled (*f*) and a south aisle was added. Two further enlargements (*g, h*) in the 13th century brought the church to its maximum dimensions (*i*). After that, demolition (*j, k*) reduced it to about its Saxon maximum by the 1600's. No significant alterations occurred during the structure's final centuries of use.

Here I shall describe the study of one such village, Wharram Percy in the Yorkshire Wolds, a site my colleagues and I have investigated in a series of summer excavations over the past 26 years.

Scattered across the English countryside are more than 400 isolated and ruined churches that since the mid-17th century have been recognized as markers of the deserted villages they once served. Compared with a ruined church, however, buried house foundations are unobtrusive, and only a few unmarked village sites were recognized before 1939. The village of Boarstall in Buckinghamshire provides an example. A plan of Boarstall drawn in the mid-15th century still exists. The houses were made of timber, since building stone was not easily available in that area. The house timbers have long since rotted, and there were no buried masonry walls to leave an imprint on the turf. Walking the ground today, one sees only shallow earthworks where ditches once bounded the village farmyards and depressions where the old roads were worn deep by medieval traffic. Yet at some seasons of the year aerial photographs of Boarstall reveal crop markings and soil discolorations exactly where the houses stand in the 15th-century plan.

My own interest in excavating a deserted medieval village site developed in 1946 from work, both in archives and in the field, that I was doing in Warwickshire. (It was closely related to a pioneer study of neighboring Leicestershire that W. G. Hoskins had begun just before World War II.) Even as late as 1946 the gulf that separated conventional documentary historians from those who used archaeological techniques was wide enough to allow much skepticism about the actual extent of English depopulation in late medieval times. My own amateur efforts were intended to do no more than demonstrate that there were indeed buildings of this period concealed below the uneven surface of farm fields. I also hoped to find objects that would confirm a time late in the 15th century as the date of the last activity within any settlement I might unearth.

It was at Stretton Baskerville, a named but uninhabited place on the border between Warwickshire and Leicestershire, that I first gathered a group of weekend volunteers in the summer of 1947. A church of which no trace remained had been recorded as a ruin here in the mid-17th century. What had further attracted my attention was a document from the preceding century: in 1518 local jurors maintained that the village served by the church had been destroyed a generation earlier by the eviction of the tenantry. Tax records showed that Stretton Baskerville was still a flourishing farm community as late as the 1440's, lending color to the



LONG HOUSE, the common peasant dwelling during the Middle Ages in England, is shown in a conjectural reconstruction. The plan is based on several long-house foundations unearthed at Wharram Percy. As was customary, the oblong shelter was divided into an area with pens for the farm animals (*left*) and a somewhat larger space, with a hearth, reserved for the family's use. At Wharram Percy the building foundations were rough chalk blocks taken from quarry pits dug near the building sites. The timber house frame, the wattle-and-daub walls and the thatched roof supported by stringers are conjectural; only the chalk foundations were unearthed. The opposing doors near the animals' end of the house, however, are indicated by gaps in the foundations situated some distance from the hearth.

EXPOSED FOUNDATIONS of a long house uncovered by the open excavation of area *b* show the chalk blocks that upheld the timbers of the house frame. The hearth area lies just forward of the third surveyor's rod; door gaps appear between the third and fourth rods.

EARLY BURIALS in the churchyard included three with stone grave slabs still in place, covering the skeletons of two adults and a child. Longitudinal ribs are carved on two slabs; the carving on the third is an "expanded" cross. The motifs are typical of a period earlier than the Norman conquest of England. Adults' graves may be those of thegns, Saxon lords of the manor.

allegation of late-15th-century eviction.

A simple trench cut along the line of a rectangular earthwork that first weekend revealed walls that confirmed the existence of the ruined church. Before I could look for house sites or our group could unearth evidence bearing on the date of the desertion, however, I had moved from the Midlands to Yorkshire. Once established there I began to examine the local landscape as I had learned to do in the Midlands, seeking in particular ruined churches and places marked on the map as having been church sites in deserted parishes.

A pleasant weekend's walking excursion in June, 1948, took me to one such isolated parish church lying in a deep-cut valley of the Wolds, a chalk plateau that covers much of eastern Yorkshire. It was well documented that the church had once served five villages. Of these villages only one survived: Thixendale, some three miles away. When the Thixendale villagers built a church of their own in the 1870's, the former parish church fell into virtual disuse.

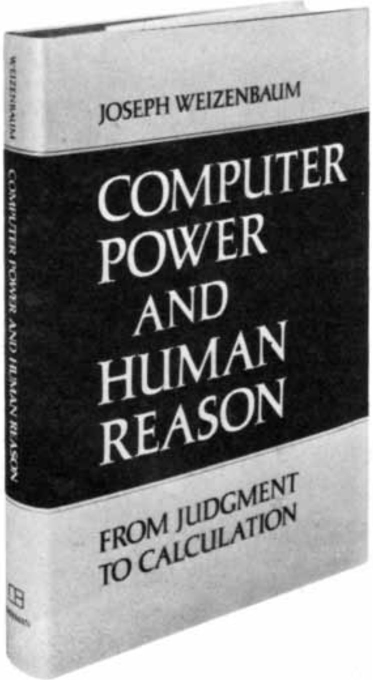
All four of the abandoned villages in the parish had left some traces on the ground, since even the meanest buildings had had foundations of locally quarried chalk. The village traces in the grass fields by the church, however, were sufficiently prominent to have given rise to a series of folklore explanations of the destruction there. Known as Wharram Percy because the lands thereabout had been sold in the 1180's to the noble family of the Percys by Henry II, the deserted village covers some 30 acres [*see illustration on page 116*]. According to one's choice of local legend, to recite them in chronological order, the village was destroyed by William the Conqueror, by the Black Death of 1349, by Henry VIII, by Oliver Cromwell, by the London plague of 1665 or by local brigands at some unspecified date.

In 1949 I happened to mention some of these erroneous attributions on a radio program marking the 600th anniversary of the Black Death. The broadcast was heard by a schoolmaster in a small Wolds village who wrote me offering to intercede with the owner of the Wharram Percy site, Lord Middleton of Birdsall, if I was interested in bringing in an excavation party. It was this chain of events that brought me to Wharram with a few students and friends early in the summer of 1950, intending to do no more than uncover a few house walls. As it happened, it was the beginning of a commitment that so far has occupied most of every July for a quarter of a century.

"History," the great French medievalist Marc Bloch has written, "has all the excitement of an unfinished excavation." At Wharram the further the archaeological work has progressed, the more complex the history of the set-

"A landmark in the integration of scientific and humanistic thinking"

–Lewis Mumford



1976, 300 pages, 13 illustrations, ISBN 0-7167-0464-1, $9.95

Computer Power and Human Reason is a distinguished computer scientist's searching probe of the limits of computer power and of scientific rationality itself. It is, above all, an eloquent defense of the uniqueness and the sanctity of the human spirit.

"Weizenbaum's deep social and ethical concerns, backed by his competence in exact science, give his work an appeal to a much wider audience than his immediate academic colleagues. . . . It will still be read half a century from now, as a landmark in the integration of scientific and humanistic thinking." —Lewis Mumford

"I found the book to be a very impressive and important one. Weizenbaum writes from a deep understanding of the promise and the limits of computer science, and he gives a very thoughtful and illuminating appraisal of what can realistically be expected, and a very cogent and telling critical commentary on some widespread illusions. . . . I know of no comparable work, and, given the considerable importance of the issues, I hope . . . for the widest possible distribution of this book." —Noam Chomsky

"I found it to be most stimulating, full of new and original ideas. I am convinced that this book will be a great success and will be read by a wide public. It deals with the societal problems with computers and their applications. These are problems which every thinking person today is interested in. Professor Weizenbaum deals with these problems in a profound way and also in a way that can be followed by those who are not experts in the computer sciences. I am convinced that this book will be widely read." —Victor F. Weisskopf

COMPUTER POWER AND HUMAN REASON

FROM JUDGMENT TO CALCULATION

JOSEPH WEIZENBAUM

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

W. H. FREEMAN AND COMPANY
660 Market Street, San Francisco, CA 94104
58 Kings Road, Reading, England RG1 3AA


RAYOVAC
RW 44
+
MADE IN U.S.A. PAT. 3655450 ETAL

The best ideas in miniaturized batteries come from ESB.

The idea: invent a watch battery that will last longer than all the others. Then make it thin enough and small enough to give watch manufacturers the greatest possible flexibility in their designs.

In other words, find a way to squeeze more power into less space.

The result: our Ray-O-Vac® divalent silver oxide button cell.

Centimeter for centimeter, it's the most powerful watch battery in the world. Bar none. Its capacity is a full 35% greater than its closest competitor.

But greater capacity, remember, was only part of our idea. We also wanted to make our super-powerful little watch battery the thinnest and the smallest in the world. And we did.

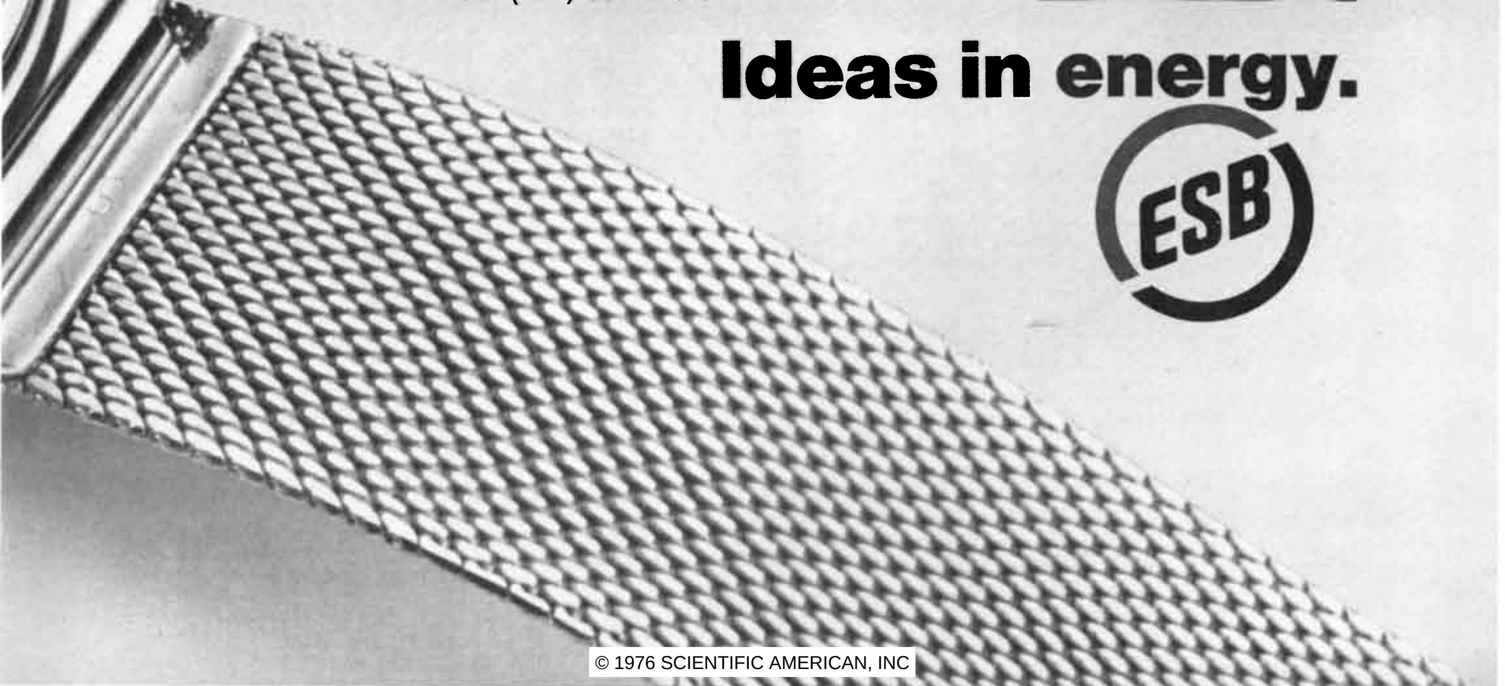
Today ESB is the undisputed leader in watch battery technology. Not because we're one of the biggest battery manufacturers in the world. Or one of the oldest. (Both of which we are.) But because our divalent silver oxide cells — like the one pictured here — consistently outperform the competition.

ESB makes miniaturized energy units for all sorts of products. From cameras to hearing aids to mini-computers.

If you'd like to learn more about our ideas in energy, write: Communications Department, ESB Incorporated, 5 Penn Center Plaza, Philadelphia, PA 19103. Or call: (215) 564-4030.

Ideas in energy.

ESB



Number One Rated!

CORVUS 500

TK Enterprises Reporting to Consumers

With MOSTEK single chip technology, the new Corvus 500 is the first non-Hewlett-Packard calculator with Reverse Polish Notation, 10 addressable memories, 4 level roll down stack to be introduced. If you compare the Corvus 500 feature by feature with the HP-45, you will find striking similarities. There are also some important differences:

	Corvus 500	HP-45
RPN (Reverse Polish Notation)	Yes	Yes
Memory Store and Recall 10 Registers	Yes	Yes
4 Level Stack, Rotate Stack	Yes	Yes
10 MEMORY EXCHANGE WITH X	Yes	No
Log, LN	Yes	Yes
Trig (Sine, Cosine, Tangent, INV)	Yes	Yes
HYPERBOLIC (SINH, COSINH, TANH, INV)	Yes	No
HYPERBOLIC RECTANGULAR	Yes	No
y^x, e^x, 10^x, $\sqrt{x}$, $1/x$, $x!$, $x \leftrightarrow y$, π, CHS	Yes	Yes
$\sqrt[x]{y}$ through INVERSE	Yes	No
GRADIANS	No	Yes
DEGREE-RADIAN CONVERSION	Yes	No
Degree-Radian Mode Selection	Yes	Yes
DEC-DEG-MIN-SEC	No	Yes
Polar to Rectangular Conversion	Yes	Yes
Recall Last ×	Yes	Yes
Scientific Notation, Fixed and Floating	Yes	Yes
Fixed Decimal Point Option (0-9)	Yes	Yes
DIGIT ACCURACY	12	10
DISPLAY OF DIGITS	12	10
%, Δ %	Yes	Yes
GROSS PROFIT MARGIN %	Yes	No
Mean and Standard Deviation	Yes	Yes
Σ+, Σ−	Yes	Yes
Product—Memories	Yes	Yes
C.F. DIRECT CONVERSION	Yes	No
F.C. DIRECT CONVERSION	Yes	No
LIT-GAL, DIRECT CONVERSION	Yes	No
KIL-LBS, DIRECT CONVERSION	Yes	No
GAL-LIT, DIRECT CONVERSION	Yes	No
LBS-KIL, DIRECT CONVERSION	Yes	No
CM-INCH DIRECT CONVERSION	Yes	No
INCH-CM DIRECT CONVERSION	Yes	No

As you can see, the Corvus 500 IS a lot more calculator for $69.95.

We have listed some of the many features, but let's amplify on some highlights:

1. RPN (Reverse Polish Notation) "COMPUTER LOGIC" and 4 LEVEL STACK. Your problem is solved the way it is written, left to right sequence, eliminating restructuring, unnecessary keystrokes, and the handicap of having to write down intermediate solutions. And all information is at your disposal—just roll the stack (R↓) to any intermediate information desired. You arrive at your solution faster, more simply and therefore more accurately.

Perhaps at this point we should address ourselves to the controversy between algebraic entry and RPN. One question we must ask is why proponents of algebraic entry *always* use an example of sum of products and *never* an example of product of sums:

(2+3) × (4+5)=

Algebraic: 2+3=MS 5+4=×MR=
TOTAL 12 keystrokes (SR51, add 2 more keystrokes)

RPN: 2 Enter 3+4 Enter 5+×
TOTAL 9 keystrokes

2. THE CORVUS 500 and HP-45 HAVE 10 ADDRESSABLE MEMORY REGISTERS, 4 LEVEL OPERATIONAL STACK, and a "LAST X" REGISTER (10th Mem. Reg.) With 10 addressable memories, you have access to more entries, or intermediate solutions; less remembering, or writing down, YOU have to do. And less chance for error.

The stack design also permits X and Y register exchange, and roll-down to any entry to the display for review or other operation.

The "last x" register permits error correction or multiple operations when a function is performed, the last input argument of the calculation is automatically stored in the "last x" register, which can be quickly recalled to correct an error, or to perform another operation using the same number.

3. DIRECT HYPERBOLIC and HYPERBOLIC RECTANGULAR to POLAR, and INVERSE. For those of you electronic and computer science engineers who require access to this specialized application, the Corvus 500 solves "your" problems.

4. A WORD ABOUT CORVUS 500 12 DIGIT DISPLAY AND ACCURACY. Finally you have displayed 12 digit accuracy in business format and 10 + 12 in scientific notation. LED is manufactured by Hewlett-Packard.

FOR THE FIRST TIME you can raise the number 10 to 199th power or calculate Factorial (x!) of up to 120. Unbelievable!

5. The CORVUS 500 CONVERTS METRIC/U.S. and INVERSE UNITS DIRECTLY. These preprogrammed functions act not just as constants, eliminating unnecessary keystrokes.

WHAT ABOUT CONSTRUCTION? With so many features, the next most obvious question must be in regard to the quality of the unit itself. We are proud to report the Corvus 500 to be double injected molded, with "tactile" feedback keyboard. The compact, contoured case is 5½" long by 3" wide by 1¼" high, and weighs just 8 oz.

The COMPLETE CORVUS 500 for $69.95 includes:

- **Rechargeable and replaceable Nickel Cadmium batteries. Optional 3 AA batteries.**
- **Adaptor/Charger**
- **Owner's handbook**
- **Soft carrying case**

The Corvus 500 is warranted by the manufacturer against defects in materials and workmanship for one year from date of delivery.

For those of you who have the HP-21 or 45 or any other advanced calculator on order, aren't you glad you still have the opportunity to take advantage of the release of the Corvus 500 for $69.95? Hurry! Order yours today.

AN INVITATION:* *TK Enterprises is proud that we are able to introduce the MOSTEK chip Corvus 500 to you, our discerning readers. (Incidentally, we have received orders for Corvus 500 from almost every major company in the USA, in every state in the Union (including HP & TI people). Many are already taking advantage of this incomparable price performance ratio instrument. We invite you to experience the new Corvus 500 by mailing in the coupon today, or call our order desk* **(800) 421-0367; in CA. (213) 370-5795.** *From the tremendous initial response, please allow 3 days for delivery. We are sure you will agree that the CORVUS 500 is the calculator of the year, and the one you have been waiting for. However, you may return the remarkable Corvus 500 within 7 days without obligation, for any reason whatsoever. So, please mail or phone today . . . before the news of Corvus 500's demand exceeds supply!

To order your Corvus 500 Scientific Calculator, phone: (800) 421-0367; (213) 370-5795 or mail coupon today!

tk enterprises
BETTER BUSINESS PRODUCTS

16611 Hawthorne Blvd.
Lawndale, Ca. 90260
(213) 370-5795
(800) 421-0367

Yes . . . I'd like to try the Corvus 500 for 7 days. Please ship me the Corvus 500 Advanced Scientific Calculator, complete with accessories, at $69.95, plus $2.50 for shipping and insurance. I understand that if I'm not completely satisfied, I may return the unit and accessories within 7 days for full refund (or, at my option, receive full credit toward the purchase of any other calculator carried by T.K. Enterprises).

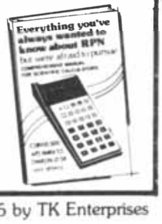
NAME ____________
ADDRESS ____________
COMPANY ____________
PHONE ____________

☐ **Cash Payment:** Check or money order enclosed. (Please be sure to add any applicable state and local taxes.) To speed-up delivery send money order.
☐ **Bill My Credit Card:** ☐ BankAmericard
☐ Master Charge Bank No. ☐☐☐☐
Account No. ☐☐☐☐ ☐☐☐☐ ☐☐☐ ☐☐☐
Expiration Date ____________
Signature: ____________

SA 10-76

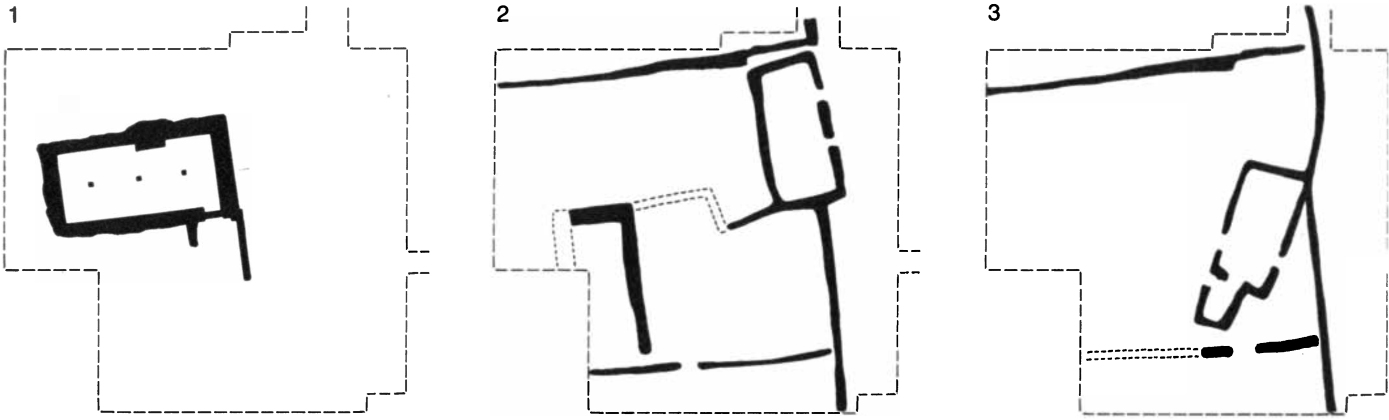
FREE 120-PG

APPLICATIONS MANUAL
(Retail Value $7.50)
IS FREE WITH PURCHASE OF CORVUS CALCULATOR.
One of the most complete and easy to read manuals on Reverse Polish Notation calculators
Also available at $7.50 to previous TK customers of the CORVUS 500 calculator.



tlement has proved to be. Perhaps this series of discoveries, which was quite unexpected in 1950, is responsible for the continuing endurance of the organizers of the work and the unflagging enthusiasm of the largely volunteer work force. The project has even become part of the history of English archaeology. The reason is that, although it was not the first medieval site where peasants' houses were unearthed, it was the first site where what were then the pioneer techniques of "open excavation," as distinct from trenching, were employed. It is no accident that the quarter century of work there has also seen the founding of the Society for Medieval Archaeology, of the journal *Medieval Archaeology* and of the Medieval Village Research Group. The research group actually came into existence at Wharram in Au-

EXCAVATED CHURCH FLOOR is shown in a photograph looking westward toward the base of a tower (*top center*) added in the 11th century to a 10th-century extension of an earlier Saxon structure. Of the four earlier stone walls three are visible here, their chalk rubble foundations partly concealed, as at the left, by blocks of sandstone masonry. A number of graves, dug both during and after the Middle Ages, have disturbed the church floor, but a few postholes (*foreground*) serve to outline the original Saxon wood structure.



SUCCESSION OF BUILDINGS at area *a* shows conversion of a 900-square-meter plot from the site of an elegant 12th-century manor house (*1*) into a farm site. In the 13th century (*2*) the plot held two long houses roughly parallel to a street. In the 14th and 15th centu-

gust, 1952. Its recent annual reports typically chronicle a dozen or more excavations in progress at English villages, usually single-house sites threatened with destruction by land development. Activity is also reported from Scotland, Wales, Ireland and the Continent, and in England alone fieldwork has now identified nearly 3,000 medieval sites.

Wharram, however, would scarcely have achieved its reputation solely as a proving ground for excavation techniques. The progress there, as in settlement studies elsewhere in England, is the result of an alliance between archaeology and advanced techniques of research using documentary sources of information on local history at the parish and village level. Aerial photography as a tool of record and investigation has been no less important.

If much of this account is written in the first person singular, it is because of my conviction that the autobiographical approach can both bring out the elements of chance that all research workers must thank for their successes and properly emphasize the organizers' good fortune in having so many diverse and friendly helpers over the years in the English tradition of amateur participation in local historical research, documentary or archaeological. And the annual congregation of volunteers (numbering over the 25 years more than 2,000, both from Britain and from abroad) would not have persisted so long if it were not for the quiet beauty of the site, the intellectual excitement of exploring the unknown and the social stimulation of a collaboration that has included summer students from American colleges, from English schools and from urban back streets and has ranged from members of the peerage to volunteers from a local youth prison.

From 1950 to 1952 my volunteers and I managed to visit Wharram for no more than two or three weekends each year. Part of our time at the site was devoted to trenching in order to expose the wall footings of four houses located on or near village streets. We also attempted to excavate sections of the elaborate late-medieval manor house at the north end of the village and of a deeply worn roadway that led in the direction of New Malton, the market town for the northern Wolds since the middle of the 12th century.

In June, 1952, we were visited by John G. Hurst, who had just received his degree from the University of Cambridge. There he had studied Paleolithic archaeology under Grahame Clark, and at Clark's suggestion he had just joined the Ministry of Works. That department of the government was then responsible for, among other things, the study and protection of national antiquities, and Hurst intended to conduct archaeological investigations of just such deserted medieval villages as Wharram. At the time of his visit we were trenching along the wall of a house some 50 meters west of the ruined church; he suggested that instead of trenching we excavate a section across the wall. We weekenders were grateful for professional advice, the more so because the sectioning shortly revealed a second wall, lower than the first and running in a different direction, then a posthole lower than the second wall and finally, below all three, a large quarry pit dug into the underlying chalk. It was a graphic demonstration that our site held evidence of several successive periods.

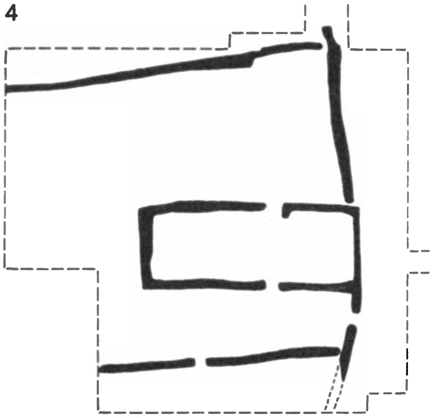
Hurst and I decided to join forces, and since 1953 he has directed the archaeological work at Wharram. For the next two decades the excavation continued on a shoestring, and our very presence at Wharram, although it was sponsored by Lord Middleton, depended on the goodwill of his tenant farmer, Mr. Wilfred Midgley. It was not until 1972 that the Middleton family trust generously placed the site under the guardianship of the Department of the Environment, and the Church Commissioners of England similarly relinquished their rights to the church and the churchyard. These actions made available government financing for a limited amount of work each year in preparation for the day when the site will be open as a public park.

To turn now to the progress of our work. All our efforts in the early years of excavation were concentrated on a house site 150 meters north of the church. The site lay conveniently near our headquarters, an empty gamekeeper's cottage of modern construction. The elements of chance that had combined to initiate the work at Wharram were equally kind to us in our choice of this house site among the 30 or so available. Not only did the excavation immediately reveal an unanticipated complexity of repairing and rebuilding (attested to since then at other Wharram house sites and at all other village excavations in England) but also, as a bonus, it uncovered below the earliest farmyard surface the foundations of a two-story manor house built in the 1180's, when the Percys were acquiring the estate.

Beginning at this house site, Hurst has shunned the traditional archaeological techniques of trenching and sectioning in favor of clearing a considerable area and then removing the soil layer by layer over the entire area, recording the level of each object as it is uncovered. For example, as soon as a stone becomes visible its position is recorded and a drawing is made, regardless of whether further excavation shows that it is merely rubble from a fallen wall or is bound by mortar to a surviving stretch of wall.

The area technique proved invaluable. For example, it is readily apparent that trenching and sectioning would have given us no coherent picture of the complex rebuildings at a typical house site [*see illustration above*]. By clearing an area approximately 30 by 30 meters, however, it was possible to trace the entire sequence of occupation. Simplifying somewhat for the sake of brevity, on top was a narrow-walled (perhaps half-timbered) cottage, oriented so that one gable end pointed toward a street to the southeast. The house had been built in the 1400's. The next house down, built

4



ries occupancy was reduced to one house and rebuilding changed the location from *3* to *4*.

in the 1300's, was oriented almost at right angles to its successor. In the 1200's two houses, parallel to each other and located at opposite corners of the farmyard, had occupied the site. Finally, late in the 12th century (documentary research suggests sometime between 1186 and 1188) the Percys' well-built stone manor house had been erected in the north corner of the area.

The thick walls of the manor house, built of facings of worked stone with rubble fill between them, and the house doorways and fireplace, all of carved sandstone, served to distinguish the Percys' abode from the flimsy foundations of locally quarried chalk on which the wood frames of the peasant cottages had been set. The manor-house sandstone is of the same quality as the sandstone that was used at the same period to rebuild the parish church. For the peasants' simpler needs crude quarries had been cut into the chalk close to each house and rough building blocks had been removed. The farmyard was pitted with a succession of these; each quarry was worked to a depth of about five feet. The resulting excavation was then backfilled with rubble and trash after a fresh pit had been opened up.

With allowances for conjecture it is possible to reconstruct the overall appearance of these medieval peasant cottages. They were oblong; the area for family occupation was at one end and the other end was set aside for animal stalls. The family end is always indicated by a cooking hearth on the floor; at the division between the human end and the animal end there are usually doors in opposite walls [*see top illustration on page 119*]. Cottages of this "long house" type have been unearthed at all the other medieval village sites so far excavated in England, from remote Dartmoor in Devonshire to the central lowlands; the oblong plan is the same whether the structures were built of wood, stone or a mixture of the two. No exact continuity has yet been proved, but the plan may be descended from Saxon houses, which were built of wood set on a foundation of timbers laid in trenches. Examples of this Saxon house plan have recently been discovered at Charlton in Hampshire and at Catholme in Staffordshire. It is remarkable that the long-house plan, although varying in dimensions, was so widespread over such a span of time, apparently unaffected by local differences in farming practices or major economic changes.

One such change during the Middle Ages was the rise and fall in population. The high-water mark that was reached in the 1200's, with its attendant pressure on available land, may be reflected at Wharram by the takeover of the Percy manor-house site (which had fallen into disuse after the construction of a second manor house at the north end of the village) and the building of the two small long houses there. The decline in population during and after the 1300's, witnessed by the number of empty houses and untended farmyards recorded in surveys conducted early in the following century, is clearly reflected at Wharram in the sequence of building and demolition we uncovered at the site of the parish church [*see illustration on page 118*]. Indeed, not only the decline but also the preceding expansion both of Wharram's population and of its landlords' wealth in the centuries before 1300 are reflected in the architectural history of the church.

The church had begun, at some still uncertain period in pre-Norman-Conquest times, as a small Anglo-Saxon structure built of wood. It was rebuilt in stone and later enlarged by additions at each end at least twice before the Normans replaced the Saxons as the masters of Wharram. Late in the 12th century, when the Percys came, they had the apse rebuilt in a semicircular plan and enlarged the church by adding a south aisle. We know that the Percys were involved in the remodeling because, although subsequent remodeling destroyed the late-12th-century apse, one piece of its masonry remained in place; on it is a mark that is the same as the mason's mark on many of the dressed stones at the manor house.

Early in the 13th century a north aisle was built, and before the end of the century a chapel was added to the east end of the new aisle. No further additions were required in either the 1300's or the 1400's, a fact we take as reflecting a drop in the population of the village. The alterations to the church about 1500 reflect this population trend even more dramatically: both the north and the south aisle were demolished and the building was reduced approximately to its pre-Conquest dimensions.

Nearly 500 burials have been uncovered in the two areas of the churchyard so far investigated; this is a very large number for the medieval period and is exceeded only in some English monastic and urban cemeteries. Among the earliest of the graves are three that were found with their stone covering slabs intact. Two of the slabs bore raised central ribs; the third was ornamented with a rude cross [*see illustration on page 120*]. It is not easy to assign a date to carving of this kind, but the burials, like the earlier of the churches on the site, are at the levels occupied in Wharram's Anglo-Saxon period. The slabs may mark the graves of the manor's Saxon lords or other persons of pre-Conquest prominence.

How old is Wharram? We may have a definitive answer after a few more seasons, but the work of the past few years has provided exciting hints of unanticipated antiquity. Moreover, the closing hours of the 1975 season revealed two important items of evidence. In addition to the putative Saxon graves our work at the church had turned up four Saxon coins of the eighth and ninth centuries, although unfortunately they were found in an unstratified context. At the end of July, 1975, however, we were examining an area at the northern boundary of the site, near the ruins of the manor house that had, we assumed, replaced the Percys', when we unearthed 100 Middle Saxon potsherds and several hundred fragments of bone, together with a silver coin struck for Eadberht of Northumbria (737–758). Along with these articles was a single potsherd of what is known as Tating ware, an expensive kind of pottery that was imported to Saxon England from the Rhineland. Tating ware is usually assigned to the second half of the eighth century or the first quarter of the ninth. Thus between the sherd and the coin it seems safe to carry the history of Wharram back at least to late eighth-century Saxon times.

The presence of imported Rhineland ware suggests some prosperous household nearby. It may therefore be that the manor at the northern boundary of the site does not merely represent a northward shift after the Percys' manor house had been abandoned. We may have here two early manorial sites, one near the parish church and the other near the northern extreme of the later village. In fact, the Domesday Book, that celebrated inventory compiled for William the Conqueror in 1086, records two separate manors at Wharram. If that was the case, what at first seemed to be a single medieval village could have been created by the coalescence of two almost adjacent communities.

So far our work along the main street frontage at Wharram has failed to produce either the Anglo-Saxon manor house or any Anglo-Saxon peasant dwelling. A new campaign is already in progress, however, at the site of the former vicarage to the north of the churchyard. If the fashion of leaving older buildings undisturbed prevailed at the vicarage as it did at the church itself, there is some chance that either the building site itself or the landholding associated with the vicarage may conceal older occupation areas.

Concerning the hints of greater antiquity, the excavation of small areas in boundary banks at the limits of the built-up area of Wharram has been under way since 1970. The work has shown that the site's premedieval features are closely aligned with the subsequent medieval ditches, boundaries and roads. Not only the banks but also Wharram's internal divisions, separating farm from farm and the Percy manor from the village, all prove to match older features. It is as if the first Anglo-Saxon settlers had found the Romano-British and even the earlier Iron Age plan still prominent enough to be adopted as their own. In 1973 we came on concrete evidence to support our suspicions: an aerial photograph revealed the presence of a substantial farmstead just beyond the boundary of the site.

Hurst's diggers ran a trial trench in the area. The cut revealed a boundary ditch 15 feet wide and eight feet deep and at the same time unearthed a coin of the period of the Roman emperor Trajan (A.D. 98–117). Thus we now can add to the Wharram inventory a fortified farm of the early Romano-British period and extend the proved antiquity of habitations at the site back to the first century. Both field-walking and air photography in the parched summer of 1976 have revealed the parent Roman villa from which this farmstead derived.

Hence what began as the exploration of a deserted medieval village has begun to push the frontier of knowledge not only into what almost everyone except Anglo-Saxon specialists has been content to call the Dark Ages but also into the even darker times before the first incoming Saxons encountered the Romanized Britons. Only a fraction of Wharram's 30 acres has been investigated, but the premedieval features we have encountered are a guarantee that much is still to be learned from this challengingly complex site.

WELL-PRESERVED BURIAL of an adult male was among some 500 burials uncovered in the churchyard at Wharram Percy. This is a priest's grave, as the presence of a pewter communion chalice indicates. He may have been a victim of the first outbreak of plague in 1349.

How You Can Save An Astonishing $4.95 To $200 On These Famous Webster Dictionaries, Bibles, Cookbooks, and Encyclopedias

Noah Webster (1758-1843) Father of Lexicography and Inspiration for the founding of our company.

These books are normally sold by high-commission door-to-door salesmen. In by-passing this costly method — and because we mail direct-to-you, from the publisher — we are able to offer you a complete selection of Webster's famous Reference Books at truly astonishing savings. These are unprecedented selling costs — discounts of up to 67% off the door-to-door prices — Dictionaries, Cookbooks, Bibles, and Encyclopedias — all reduced so incredibly low, just by avoiding the middleman.

Think of it! Webster's famous Reference Books are among the best of their kind in the world, regardless of price. They are compiled by scholars and experts alike, who are top-rated in their fields. Originally published for direct sales forces to sell door-to-door, now you can get them at these "direct-from-publisher" prices during this special mail-order offer.

All of this and more — Your purchase is backed by Webster's Ironclad Guarantee. It is sent to you on a 14-day Free Home Trial Basis (no salesman will call). Thumb through your selections as often as you wish. You must agree that they are the most thorough and comprehensive reference books of their kind, regardless of price, or return them for a prompt and courteous refund. No questions asked.

Physically, these references are deluxe editions, designed to become a permanent part of your library. Durably bound, and hard covered, these books are as attractive as they are concise. The texts are all set in easy-to-read type for those with visual problems. In short, no finer way ever devised to help you become the well-read, totally-informed, well-rounded and complete person everyone wants to be.

The selections include:

• The Giant, Library Size Authentic Webster Dictionary

This is the abridged, library size, 1,454-page, eight-pound, 158,000-definition Living Webster Encyclopedic Dictionary of the English Language — with a 300-page supplement, including a Thesaurus of Synonyms and Antonyms, Prefixes and Suffixes, Dictionary of Mythology, Foreign Words and Phrases, Popular Quotations, etc. Considered to be a library in itself. Latest 1976 edition, updated to include the newest colloquialisms.

$39.95 Now Only $19.95

• The Huge, Unabridged Thumb-Indexed, Webster Dictionary

This is the gigantic, eleven-pound, 2,336-page, 400,000-definition, 492-cubic inch, Webster's New Twentieth Century Dictionary, Unabridged Edition. Complete with 175-page supplement including 16-full page maps and full-page illustrations in full color. Latest 1976 edition, including a complete section on modern contemporary, and colloquial phrases.

$59.95 Now Only $29.95

• The Christian Life Edition of the Holy Bible

"The Most Beautiful Bible in 500 Years" is available in both King James Translation for Protestants and the Authorized text for Catholics. Designed to be passed from generation to generation as an heirloom, this pulpit-size masterpiece has 84 pages of full-color paintings, sculptures, frescoes, and sketches of Michelangelo's. Also included are personal and family record sections listing births, baptisms, confirmations, marriages, etc. Please specify Catholic or Protestant version.

$59.95 Now Only $29.95

• The New World Encyclopedia of Cooking

More than 4,000 of the finest classic and contemporary recipes. Complete with menus and recipes that are all kitchen tested by the Culinary Arts Institute. 784 extra large pages (8 3-10 x 10 7-8). More than 1,250 illustrations, 16-full color photographs. Washable leather-grain cover for lifelong wear.

$19.95 Now Only $11.95

• The New Home & Auto Do-It-Yourself Encyclopedia

A to Z, shop-tested, step-by-step, money-saving, do-it-yourself projects for home, auto, and appliances. This book can save you thousands of dollars each year. Even a housewife can learn to be an expert handywoman. Easy to follow fix-it instructions, more than 1,000 photos, drawings and charts. 512 pages that make you the expert.

$14.95 Now Only $10.00

• The 4-Volume Home Medical Encyclopedia

All about body diseases, ailments and other common health problems. More than 2,500 articles about the body, mental and emotional disorders. Medical terms are defined in laymen's words. The 4 volumes include more than 450 photographs of physical disorders, drawings of the body, charts, tables, and diagrams. When the Doctor's not around, the perfect mother's companion.

$29.95 Now Only $19.95

Special Premium Offer

(With the purchase of $19.95 or more)

• The Universal Edition of The New Standard Encyclopedia

The New Standard Encyclopedia is a joy to peruse, with over 9,200 pages, 6,000,000 words, 17,900 articles, 12,000 illustrations (4,287 in color), 1000 location and relief maps, and 128 multi-colored political maps. Top-rated by the most authoritative guide of encyclopedias, this massive compendium of man's knowledge, is an exciting glimpse into the accomplishments of man. Updated to 1976, this huge, 14-volume set keeps you as well informed on any subject as the experts who compiled it. So fascinating that once you become immersed in reading it, you'll find it hard to stop.

$299.95 Now Only $99.95

To repeat, these prices are unprecedented. If you ever wanted to own a truly fine Reference Book, now is the time to buy it.

To order, simply fill in the coupon below, clip and mail to: The Webster Dictionary Co., Inc., Suite 500, 625 N. Michigan Ave., Chicago, Ill. 60611. You can't lose. Clip out this message as a reminder. Supplies are limited. Do it Now — this offer may never be repeated.

The Webster Dictionary Company, Inc. DEPT-112-D
Reference Book Division
Suite 500, 625 North Michigan Avenue
Chicago, Illinois 60611

Please send me the volume(s) checked below.

- ☐ Living Webster Dictionary @ $19.95 = ______
- ☐ 20th Century Webster Dictionary @ $29.95 = ______
- ☐ Holy Bible (P or C) @ $29.95 = ______
- ☐ New World Cookbook @ $11.95 = ______
- ☐ Home & Auto Encyclopedia @ $10.00 = ______
- ☐ Home Medical Encyclopedia @ $19.95 = ______
- ☐ Premium Offer Encyclopedia @ $99.95 = ______

(Check One) Total $ ______

- ☐ I enclose payment
- ☐ Bill to:

BankAmericard No. ______

American Express No. ______

Master Charge No. ______ Inter Bank No. ______

Expiry Date ______

Name ______

Address ______

City ______

State ______ Zip ______

Please add $2 per selection ($10 for premium encyclopedia) for freight and crating.

Ironclad Money Back Guarantee


LANCIA BETA SEDAN
LANCIA BETA COUPE
LANCIA BETA HPE
LANCIA BETA SCORPION

MATHEMATICAL GAMES

Combinatorial problems, some old, some new and all newly attacked by computer

by Martin Gardner

The "combinatorial revolution" in mathematics is still exploding as books and articles on combinatorics continue to proliferate. The computer is certainly contributing to the revolution by its power to analyze combinatorial problems that are too complex to be approached any other way. Another stimulating force is the increasing application of combinatorial theory to science and technology, particularly in particle physics and molecular biology. On a large scale the universe may be a collection of continuums to be handled by calculus, but on the microlevel it is a jumble of discrete elements with mysterious jumps and curious combinatorial properties. In some modern theories even time and space are quantized.

Hundreds of intriguing combinatorial puzzles, some old, some new, are now receiving the attention of serious mathematicians. We shall first take a look at an amusing prize puzzle devised by Sam Loyd that has recently been "cooked," or invalidated, by a computer program and then follow with several combinatorial problems that have no general solution in sight but present challenging tasks on their lower levels.

The greatest puzzle book ever published in the U.S. is a large volume, bound in pale green cloth, that has on its cover *Sam Loyd's Cyclopedia of 5000 Puzzles, Tricks and Conundrums, with Answers.* The spine bears a price of $5, but today one is lucky to find a copy for less than $25. Almost nothing is known about its publishing history. Although it is profusely illustrated, none of its several artists is identified. Copies bear the imprint of either the Morningside Press or the Lamb Publishing Company, both of New York, but no one knows which edition came first.

All copies are dated 1914. Because this was three years after Loyd's death it was widely assumed that his son, who took his father's name, had selected the puzzles from older newspapers and magazines and had patched them hastily into one massive volume. (The book is peppered with omissions, mistakes and printer's errors.) It was later discovered that the elder Loyd had edited and published a quarterly called *Our Puzzle Magazine,* which first appeared in June, 1907. Copies of it are now exceedingly rare. The *Cyclopedia* is simply a reprinting of this magazine, made from its unaltered page plates.

Loyd's "Back from the Klondike" puzzle, shown in the top illustration on the next page, appeared in the second issue (October, 1907) of his magazine and is reprinted on page 106 of the *Cyclopedia.* It is not known whether Loyd called it a "prize puzzle" because a prize for its solution was offered to readers of the magazine or whether it had been a contest puzzle in some earlier incarnation.

"Start from that heart in the center," Loyd says, "and go three steps in a straight line in any one of the eight directions, north, south, east or west, or on the bias, as the ladies say, northeast, northwest, southeast or southwest. When you have gone three steps in a straight line, you will reach a square with a number on it, which indicates the second day's journey, as many steps as it tells, in a straight line in any one of the eight directions. From this new point when reached, march on again according to the number indicated, and continue on, following the requirements of the numbers reached, until you come upon a square with a number which will carry you just one step beyond the border, when you are supposed to be out of the woods and can holler all you want, as you will have solved the puzzle."

Loyd's sneaky solution is shown by the colored line. The path simply shuttles up and down along a main diagonal until a number is reached that makes possible what Loyd, in his answer, calls a "bold strike via S.E. to liberty!" The solution, Loyd claims, is unique. "Those who failed to master it readily discovered that one false step at any stage of the game throws one into the whirlpool from which there is no egress."

But the old maestro was wrong! Early this year Penelope J. Greene, a graduate student in sociology at the University of Washington, R. Duncan Mitchell, a graduate student in economics there, and Horace A. Greene, a science teacher at Roosevelt University in Chicago, made a combined attack on the problem with a Fortran program written by Mitchell. In a few seconds the program found Loyd's solution along with an explosion of hundreds of others. Indeed, one can leave the center 3 in any of the eight directions and escape from the woods. The start of one alternate route is shown in the bottom illustration on the next page. All these other routes lead eventually to 4, a key link on the diagonal, from which the path is completed as it is in Loyd's solution.

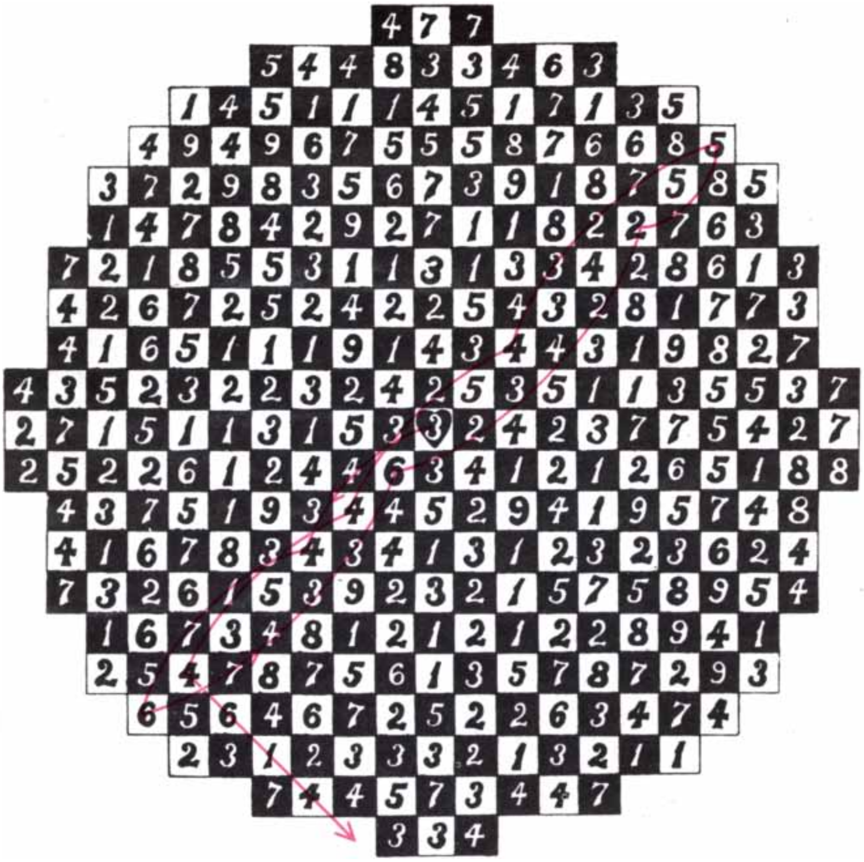
The three programmers noticed something unusual about the alternate answers. One cell and one only, not a link in Loyd's solution, is common to all the alternative routes. They called it the "troublesome 2" [*circled in illustration*]. Could this have been an artist's error? It seems likely, because when the programmers substituted other digits for 2, the computer found only one nonzero digit that eliminates all alternative paths.

This is the pleasant task—it is really quite easy—that I offer readers before next month's answers: Correct Loyd's Klondike puzzle by changing the troublesome 2 to the only nonzero digit that restores uniqueness to Loyd's solution. In calling Loyd's solution unique, we discount wasted "side trips," where one leaves a cell only to return to it again before proceeding or where one leaves cell *A* for a side trip to *B* when one can go directly from *A* to *B*.

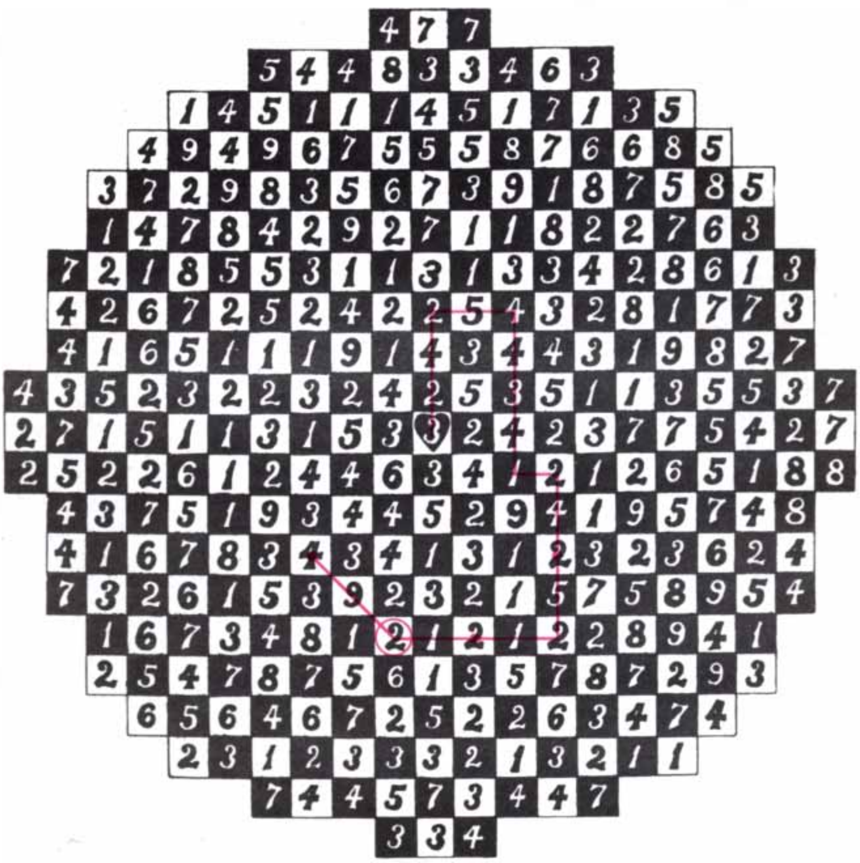
The origin of Chinese checkers, a board game popular around the world since about 1920, is obscure. Apparently it was first marketed under this trade name in the U.S. It has no connection with China and almost no resemblance to checkers, but the name has persisted.

The standard board is shown in the illustration on page 133. The holes are numbered for convenience in recording moves, and colored borders have been added to outline all star boards of smaller size. If two people play, 10 marbles of one color are placed in the triangular "yard" formed by holes 1 through 10 and 10 marbles of another color go in holes 112 through 121. Rules of play are similar to those of Halma (see this department for October, 1971), a popular British game that probably suggested the variant represented by Chinese checkers.

Two kinds of move are allowed. A "step" is a move in any of six directions to an adjacent vacant hole. A "hop" is a move over an adjacent marble, again in any of the six directions, to a vacant hole



Sam Loyd's solution to his Klondike puzzle



An alternate path through the "troublesome 2"

on the other side. The hop is like a checker jump except that the jumped piece is not removed and can be either one's own piece or an opponent's. A piece can continue hopping as long as possible or the chain can be stopped at any point. Hops are never compulsory. The combining of steps and hops is forbidden. The winner is the first player to occupy all the holes of his opponent's starting yard.

The game is obviously unplayable on the degenerate smallest "star" of one hole. On the next larger star each player has one marble and the game is a trivial win for the first player to move. The increase in complexity as the size of the board is increased to three marbles on each side is enormous. I believe it is not known which player has the win.

On this star the game is quite pleasant. To prevent one player from forcing a draw by keeping a marble perpetually in its home yard one more rule must be added: If a counter can leave its yard by hopping an enemy counter, it must do so. Once it is out of its yard it may not return, although it is allowed to pass *through* the yard in a chain of hops.

Let us consider on each star board the following solitaire problem. Marbles of one color are placed in starting positions on the bottom yard. What is the minimum number of moves (counting a chain of hops as a single move) necessary to transfer all of them to the yard at the top?

The problem is meaningless on the smallest star and trivial on the 13-hole star. For the star with a yard of three marbles 11 moves will do it. Example: 92–81, 81–71, 105–82–61, 61–52, 93–82, 82–61–42, 42–30, 71–61, 61–42–17, 52–41, 41–29.

For the six-marble star the best solution I have been able to get is 18 moves. Starting with 113–93 there are many ways to construct in nine moves a vertical "ladder" of marbles on holes 9, 30, 52, 71, 93 and 114. Nine more moves, reversing the previous play, put the six marbles in the top yard.

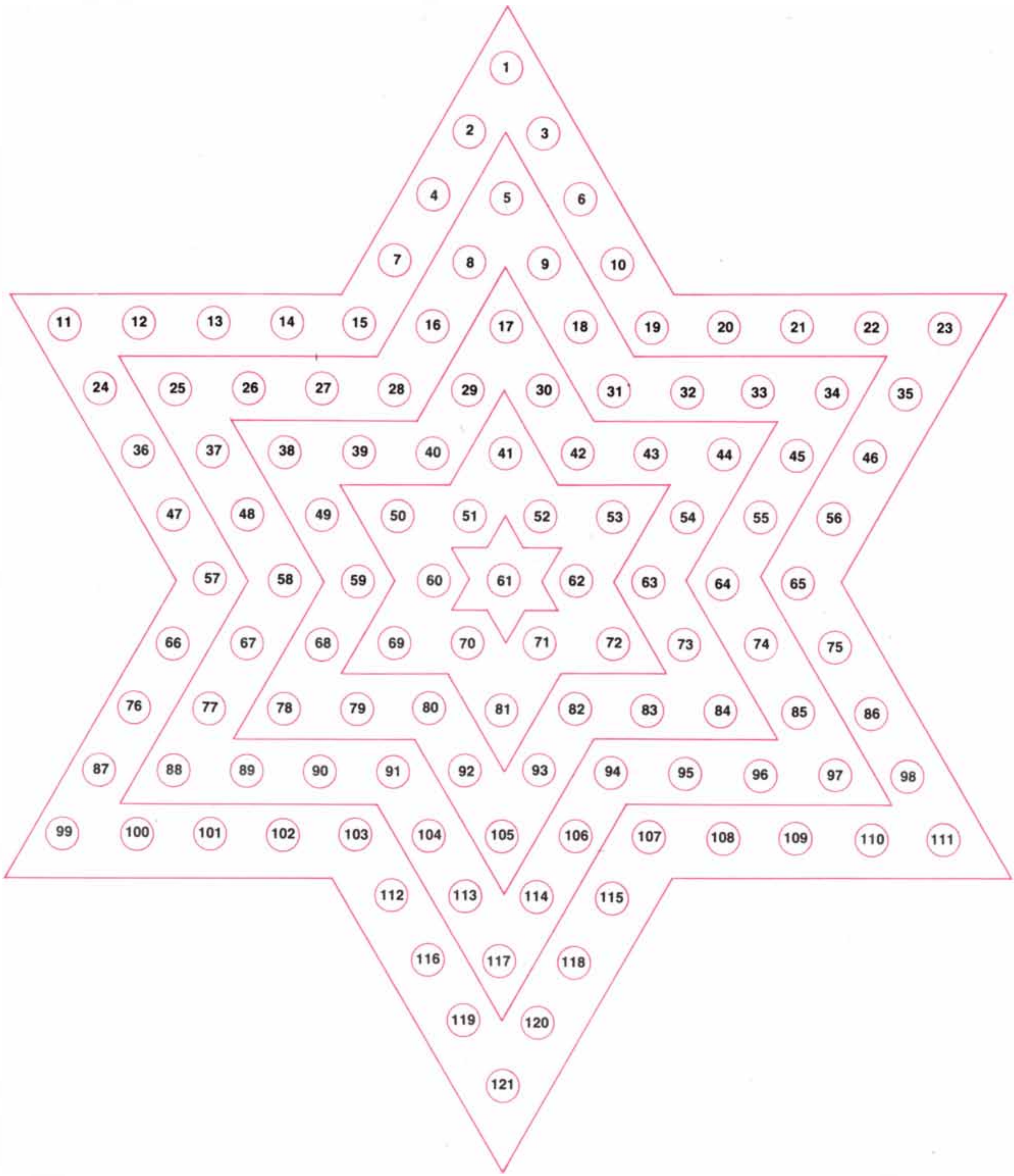
Now for our main problem: What is the minimum number of moves needed to transfer 10 marbles from one yard to the opposite yard on the standard Chinese-checkers board? I first heard of this problem in 1961 from Octave Levenspiel, professor of chemical engineering at Oregon State University. He wrote that the best he could do was 31 moves but that his mother, "the family puzzle solver, chess and bridge champ," had once solved it in 28 but had failed to record the moves. *Ibidem,* a Canadian periodical on magic (now defunct), published in August, 1969, a "proof" that 29 is the minimum. Then in 1971 Levenspiel, trying to reconstruct his mother's 28-move solution, hit on a spectacular solution in 27! Harry O. Davis of Portland, Ore., believes he has a proof that

27 is the minimum, but the proposition has not yet been confirmed.

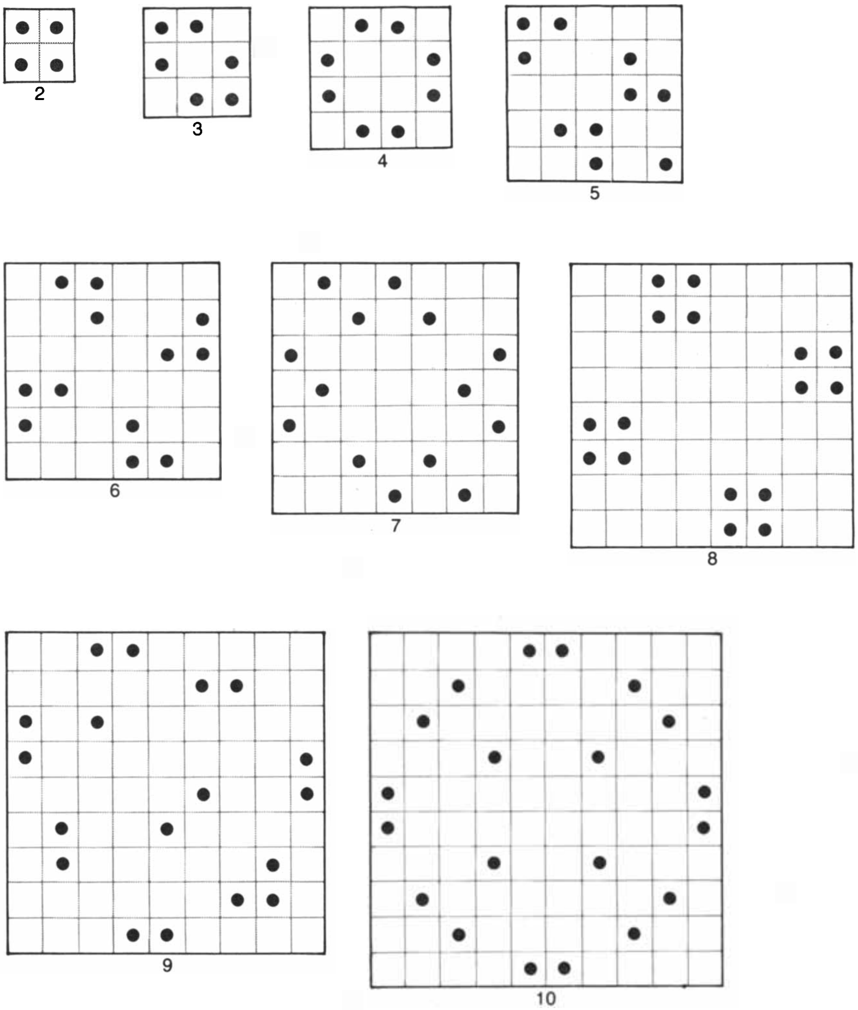
Can the reader find a 27-move solution? I shall be pleased to hear from anyone who does it in fewer moves or improves on any of the minimums given here for smaller boards. Chinese checkers is sometimes sold with a larger board that has 25 marbles in each yard. For such a board the best solution in my files, sent to me in 1974 by Min-Wen Du of Taiwan, is 35 moves.

In the London *Tribune* for November 7, 1906, Henry Ernest Dudeney, Loyd's British counterpart, published the problem of placing 16 pawns on a chessboard so that no three are in line in any direction. (The problem later appeared as No. 317 in his *Amusements in Mathematics.*) By "line" is meant *any* straight line, not just orthogonals and diagonals. The pawns represent points at the centers of cells. Since then the generalized no-three-in-line problem, as it came to be known, has been the topic of several technical papers.

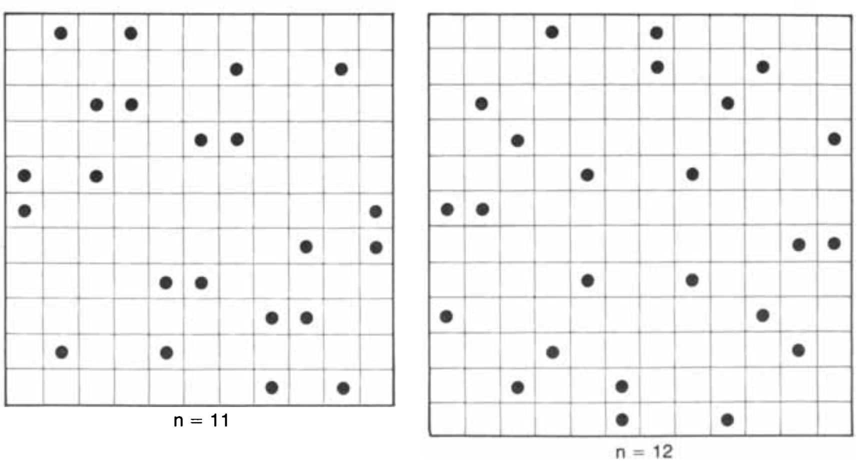
The main problem, which is still un-



A variety of boards for Chinese checkers



Some solutions to the maximum "No-Three-in-Line Problem"



Recent solutions to the problem

solved, is to answer the following question: On a square checkerboard of n cells on a side (n greater than 1) is it always possible to place $2n$ counters so that no three are in line? The old "pigeonhole" principle proves that $2n$ cannot be exceeded. No row or column can hold more than two counters, and since $2n$ counters require two in each orthogonal, one more counter is certain to put three counters in one row as well as three in one column. A less trivial argument (developed by R. R. Hall, T. H. Jackson, A. Sudbery and K. Wild in their paper "Some Advances in the No-Three-in-Line Problem" in *Journal of Combinatorial Theory,* Series A, Vol. 18, May, 1975, pages 336–341) proves that at least n counters can always be placed. For large boards these authors show that one can get quite close to $3n/2$ counters.

Michael A. Adena, Derek A. Holton and Patrick A. Kelly, in their paper "Some Thoughts on the No-Three-in-Line Problem" (in *Combinatorial Mathematics: Proceedings of the Second Australian Conference,* edited by Holton, Vol. 403 of *Lecture Notes in Mathematics,* Springer-Verlag, Berlin, 1974), reported on computer programs that found all distinct solutions for n equals 2 through 10. Rotations and reflections are excluded. The number of solutions are respectively 1, 1, 4, 5, 11, 22, 57, 51 and 156. The top illustration at the left gives an example for each n from 2 through 10. Note the startling simplicity and symmetry of the order-8 solution!

At the time these authors wrote, no solution for $n = 11$ was known. They gave a solution for $n = 12$, but it proved to be invalid. They had failed to notice two rows of three in line.

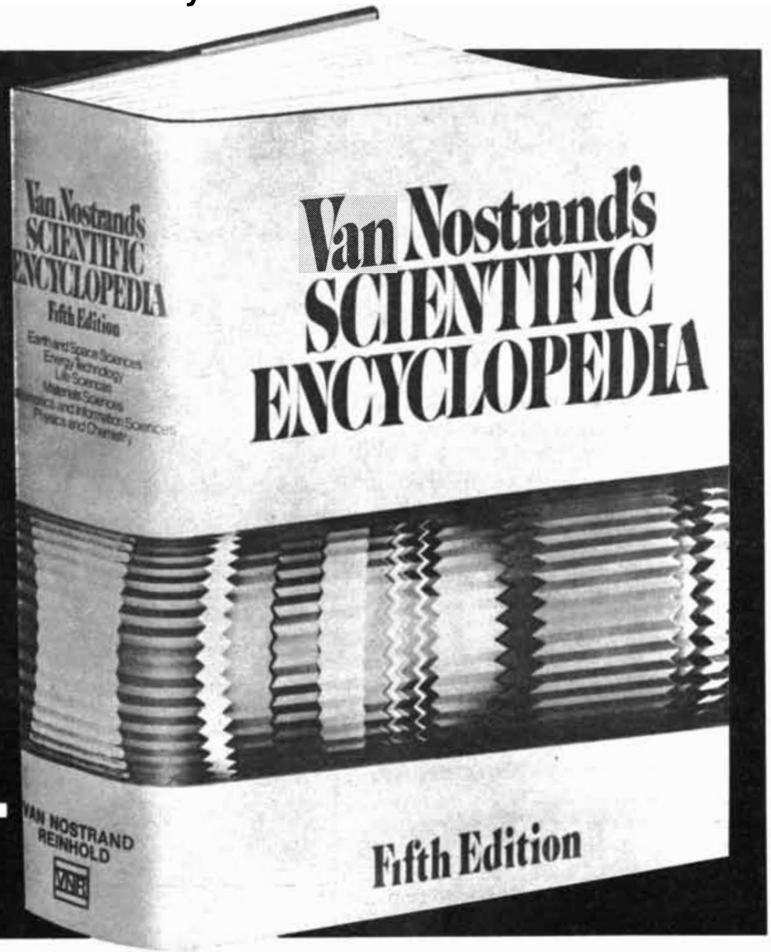
Solutions for $n = 11$ and $n = 12$ were found in 1975 by D. Craggs and R. Hughes-Jones of the University of Kent and published in *Journal of Combinatorial Theory* (Series A, Vol. 20, May, 1976, pages 363–364). They are shown in the bottom illustration at the left. Craggs and Hughes-Jones found five other solutions for $n = 11$ and three others for $n = 12$. The total number of solutions for these two values of n is not known, and no one has constructed a solution for a square of an order higher than 12. Richard K. Guy and Patrick A. Kelly, in their paper "The No-Three-in-Line Problem" (*Canadian Mathematical Bulletin,* Vol. 11, No. 4, 1968, pages 527–531), give arguments to support their conjecture that the number of orders with $2n$ solutions is finite. The smallest board on which a $2n$ solution is impossible is not known.

If we narrow the definition of "line" to an orthogonal or diagonal row, the problem is solved. The maximum cannot be greater than $2n$, and $2n$ is possible on all boards of orders higher than 1. For n greater than 3 the problem is solved by superposing any two solutions of the classic nonattacking-queens problem. (This is the problem of placing n queens so that no queen attacks another.) It is always possible to do this with a pair of solutions that put the $2n$ queens on $2n$ cells.

Instead of asking for the maximum number of counters that can be put on an order-n board, no three in line, let us ask for the minimum that can be placed such that adding one more counter on any vacant cell will produce three in

Your fastest way to find answers in every area of science... in one authoritative volume.

Examine without risk this new, up-to-date edition of the world's most popular science encyclopedia.



Van Nostrand's—the classic.

Turn confidently to the new Fifth Edition of this classic—find answers to virtually all questions on every aspect of today's science, engineering, and technology.

This authoritative reference contains thousands of definitions, explanations, overviews, and specialized discussions. It provides a complete range of scientific and technical knowledge in clear language—in just the right depth.

Newly revised cover to cover.

For four decades *Van Nostrand's Scientific Encyclopedia* has been hailed as the most comprehensive and practical reference of its kind. Now more than 80% of its text and illustrations is brand new, keeping pace with the explosive growth of technical knowledge.

This new edition contains nearly 2,400 pages, more than 2,000,000 words, over 2,500 photographs, drawings, diagrams, and charts.

Many short entries have been consolidated to yield more concentrated information. Increased cross referencing enables you to find answers quickly. Cross-referenced items are set in boldface for immediate visibility.

Examine this "library" without risk.

This reference classic, in large format, is strikingly handsome. Its authority and clarity make it an ideal gift—a fine addition to any library.

Mail the handy coupon now, to see for yourself how this convenient single volume is like a multi-volume science library at a fraction of the cost—only $67.50.

Examine the new Fifth Edition of *Van Nostrand's Scientific Encyclopedia* for 15 days—absolutely without risk. Then you, and you alone, decide how important it can be to you.

Among the hundreds of topics covered ■ astronomy and astronautics ■ biology and biochemistry ■ chemical engineering ■ communications ■ computer technology ■ electronics ■ energy technology ■ environmental science ■ geology and hydrology ■ mathematics ■ mechanical engineering ■ medicine ■ nuclear science and engineering ■ oceanography ■ physics ■ statistics

New edition gives you ■ 2,328 pages ■ over 2,500 photographs, drawings, diagrams, charts ■ more than 2.2 million words ■ 7,200 articles ■ 550 tables ■ 8,000 boldface cross references ■ over 80% new text ■ 90% new illustrations

THE EQUIVALENT OF A MULTI-VOLUME SCIENCE LIBRARY AT A FRACTION OF THE COST.

VNR NO-RISK EXAMINATION MAIL THIS COUPON NOW

VAN NOSTRAND REINHOLD
300 Pike Street, Cincinnati, Ohio 45202

Please rush me the new Fifth Edition of *Van Nostrand's Scientific Encyclopedia*. If not completely satisfied, I may return it in 15 days for full refund.

☐ **Single Payment:** Save postage, handling—enclose check/money order for $67.50 plus local tax.

☐ **Budget Plan:** Enclose check/money order for $7.50 deposit—remit balance in three monthly payments, for a total of $67.50 plus postage, handling, tax.

☐ **Master Charge** ☐ **BankAmericard**

No. ____________ Exp. ____________

Sig. ____________

Name ____________
(Please print)

Address ____________

City ____________

State ____________ Zip ____________

Offer good in U.S.A. only. Subject to credit department approval. Payment must accompany order to P.O. box.

IN CANADA: For price, terms write Van Nostrand Reinhold Ltd., 1410 Birchmount Road, Scarborough, Ontario.

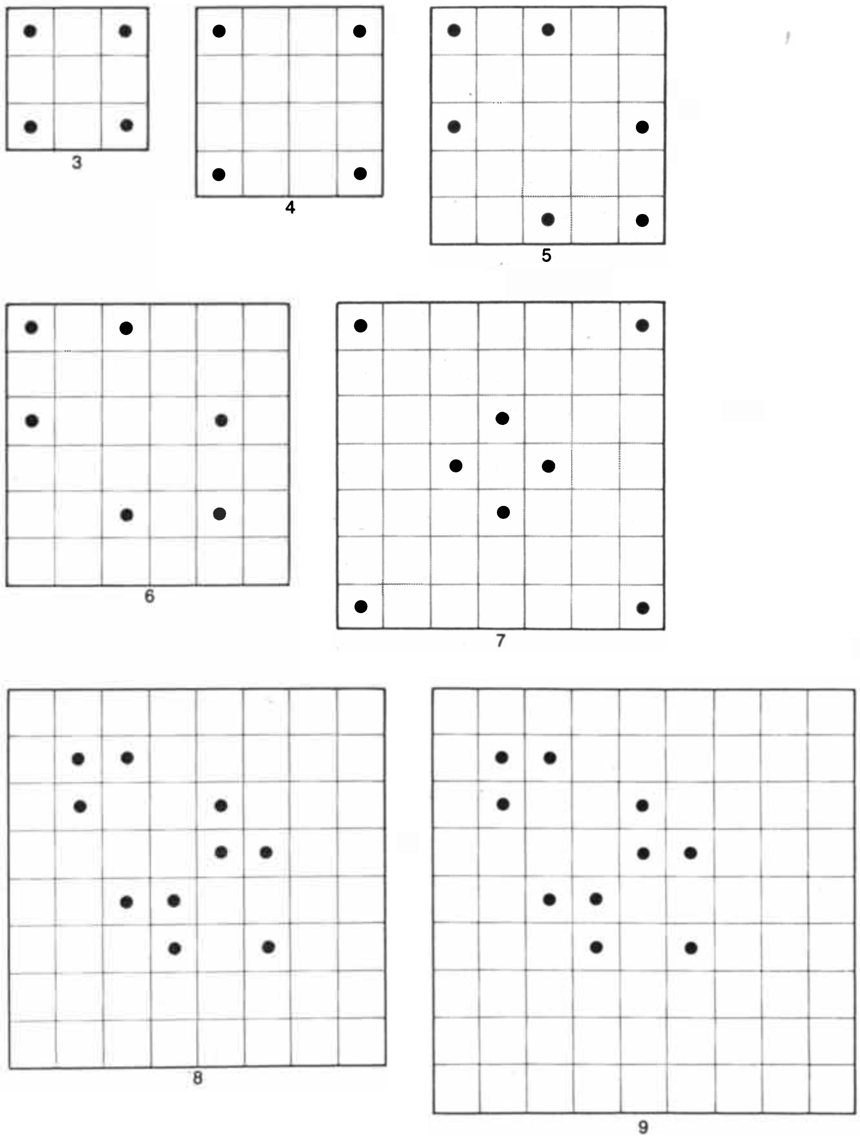
21629-7 S 1062

line. This interesting problem has not yet received serious attention from the experts.

If "line" is taken in the broadest sense (a straight line of any orientation), the problem may be very difficult. Adena, Holton and Kelly mention it briefly in their paper and give the following best results, which they obtained by hand, for n equals 3 through 10: 4, 4, 6, 6, 8, 8, 12 and 12. As far as I am aware it is not known how this tantalizing pattern continues.

The problem is also unsolved if "line" is restricted to orthogonals and diagonals. In other words, what is the smallest number of pawns you can put on a board of side n such that no pawn can be added without creating three in a row, a column or a diagonal? Another way to view the problem is to see it as a game in which two players alternately place a counter on a square board until one player loses by being forced to make three in an orthogonal or diagonal line. How short can such a game be? A puzzle asking for a solution on the order-8 board appeared in *Technology Review* (June, 1967) in Allan J. Gottlieb's excellent "Puzzle Corner." This is the only publication known to me of a problem of this type.

The illustration below shows the best results I have found for order n equals 3 through 9. I made no attempt to find all solutions for any except the smallest boards. There are dozens of solutions for $n = 8$, many with bilateral symmetry and others with twofold symmetry. (If 10 is the minimum for the order-8 square, fourfold symmetry is ruled out because 10 is not evenly divisible by 4.) Can the reader construct, before next month's answer, a single pattern of 12



Solutions to the minimum problem

counters that simultaneously provides a 12-counter solution for orders 10, 11 and 12? I cannot prove that 12 is minimal for these three values, but I strongly suspect it is.

Last February's column on brick-packing included an unsolved problem about the packing of "canonical bricks" ($1 \times 2 \times 4$) into cubes of side n, with n greater than 3. When n is even, the cube can be fully packed (in a trivial way) only if n is a multiple of 4. I gave a simple proof that if n is even and not a multiple of 4, complete packing is impossible. It is necessary to omit one brick.

When n is odd, the situation is more interesting. Each $n \times n$ "slab" must contain a unit hole; consequently the maximum number of bricks that an odd-order cube can take is $(n^3 - n)/8$. For what values of n can this maximum packing be achieved?

When I wrote, only the 5-cube had been proved impossible. The proof was later generalized by David A. Klarner to cover exactly half of all odd n, namely cubes with sides equal to plus or minus 1 (modulo 8). Many impossibility proofs for the 7-cube were found, but they did not generalize.

The problem is now solved. Robert Ammann of Lowell, Mass., first proved impossibility for all odd n except n equal to plus or minus 1 (modulo 16). He then found a maximal packing for the 15-cube that generalized to provide maximal packings for all odd n equal to plus or minus 1 (modulo 16).

Frank Barnes, working independently in England, obtained Ammann's earlier impossibility result and conjectured that maximum packings exist for the 15-cube and all others with sides plus or minus 1 (modulo 16). Told of Ammann's work, Barnes verified its soundness. In addition he found a way to show that all impossible cubes *could* be packed by omitting just one brick from the maximum number. It is hoped that all these results will eventually be published.

Neither Ammann nor Barnes has an academic post. Ammann is a computer programmer working in nonmathematical areas. Although Barnes has lectured on mathematics at the University of Michigan and the University of Reading, he has been earning a living for the past two years by piloting a hot-air balloon for advertising purposes.

Last month's problem was to find the winning response to a first-player move in a game called Cutcake. The cake is a 4×7 rectangle. If the first player breaks the cake vertically into a 4×4 square and a 4×3 rectangle, the unique winning reply is to break the 4×3 piece into two 2×3 rectangles.

BOOKS

The scientific and public life of P. M. S. Blackett, the folktale and other matters

by Philip Morrison

PATRICK MAYNARD STUART BLACKETT, BARON BLACKETT OF CHELSEA: A BIOGRAPHICAL MEMOIR, by Sir Bernard Lovell. The Royal Society, Carlton House Terrace, London SW 1 (£4 10s). Midshipman at the Battle of Jutland, first to photograph a nuclear reaction "by the Wilson method," World War II director of naval operations research, founder of the defense research organization of the new government of India, P. M. S. Blackett played diverse and important roles in the drama of science and society in our century. Sir Bernard Lovell (of Jodrell Bank), who became "one of Blackett's young men" at Manchester in 1936 (Blackett was himself not quite 40), has prepared here no mere impression but a 100-page account of that dramatic life, full, intimate and yet clearly seen. His memoir is published both as a small book by the Royal Society and in the 1975 volume (Volume 21) of the less accessible *Biographical Memoirs of Fellows of the Royal Society.*

Blackett was born in Kensington; his father was an unconventional stockbroker whose "great interests were in literature and nature." Military, missionary and Indian service were all in the family. As a 12-year-old he was put up for appointment to naval college and was interviewed by four admirals, who, playing the usual gambit of testing the candidate's alertness, asked him about the first round-trip flight across the English Channel, which had been completed the day before. Since the boy had been spending his spare time building model aircraft and crystal radio sets, he "bored the Admirals by telling them far more than they wanted to know." He was at the top of his class at Dartmouth when the war began, and he saw action at sea even before Jutland. There he served on the heavily damaged cruiser *Barham;* when he first left his gun turret, he found belowdecks "an extraordinary reek of TNT fumes, which mixed with the smell of disinfectant and blood was awful." The destroyer on which he became fire-control officer was posted to Harwich fresh from the Grand Fleet at Scapa Flow, and within 12 hours it had dashed without permission across their startled new admiral's bows at full speed—to sink a U-boat, the first sunk by the Harwich force, which P. M. S. thought "must have been very galling."

Blackett intended to leave the navy after the war, to seek work with some instrument firm (he already held a secret patent on a gunnery device), but was sent to Cambridge for six months' study as an effort to redress the curtailed education of his wartime class at naval college. He came to Cambridge in uniform. Within a few days he was happily immersed in intellectual conversation and had found his way to the flowering Cavendish Laboratory. He was out of the navy within three weeks, earned his degree a couple of years later and became a research student under Ernest Rutherford late in 1921.

Blackett remains a paragon of the experimental physicist. Like Rutherford, he had a deep grasp of the fundamental concepts behind his experiments; unlike Rutherford, he did not conceal his admiration for theory, and he embraced quantum theory itself. Indeed, when he left Rutherford to work for a year in Göttingen with James Franck, he was among the first English physicists to reopen contact with Germany. Rutherford found in that one trip two sins: leaving the Cavendish and studying the outside of the atom!

In the early 1930's Blackett wrote an unforgettable paper ("a perfect self-portrait") titled "The Craft of Experimental Physics." One passage reads: "With such varied manual and mental skills... does the experimenter go about his work in the laboratory, an amateur in each alone, but unique in commanding them all. It is the intimate relation between these activities of hand and mind, which give to the craft of the experimenter its peculiar charm. It is difficult to find in other professions such a happy mixture of both activities.... He must be enough of a theorist to know what experiments are worth doing and enough of a craftsman to be able to do them. He is only pre-eminent in being able to do both." But Blackett's world was no longer Rutherford's string-and-sealing-wax one. He analyzed in depth the practical systems he used: quick-moving pistons, stereographic cameras, good-sized magnets. He improved his apparatus not by improvisation but by systematic expert design and test. His gear was complex by the standards of its time; nowadays, in particle physics at least, that complexity has grown to require big engineering teams for experiments, and Blackett's example for the experimenter is now more a heroic ideal than a reality.

He made major changes in three provinces of physics. The first decade he spent as a pioneer examining nuclear reactions with the cloud chamber. The second decade was one of particle physics as it is revealed in cosmic rays, beginning when he and his colleague G. P. S. Occhialini, who came from Florence for three weeks and "stayed for three years," operated the first counter-controlled cloud chamber. It ended with the first strange particles, found in his Manchester cloud chamber, as modified and operated by G. D. Rochester and C. C. Butler. The third decade was taken up with paleomagnetism and continental drift, a field Blackett entered after his revival of an old theory that regarded magnetism as being generated by the simple rotation of bodies, even electrically neutral ones. He developed the best of magnetometers, with which in a virtuoso experiment he quickly proved his own ideas wrong. In so doing he invigorated the study of the magnetism of rocks over geologic time, and from the data he inferred the concept of continental drift that dominates the earth sciences today. A nuclear physicist in his fifties, he had made a fundamental contribution to another field that "merits his inclusion in the small group of scientists... who have done so much to create modern geophysics."

For more than 50 years he spoke his mind on political matters and acted on them. He was throughout "remarkably consistent in his political and social attitudes," although the times often changed around him. He described himself as a Fabian socialist; from his first entry into active naval service he worried about the class distinctions of the navy. The Depression turned him still more clearly toward the problems of wealth and poverty. He was an early participant in the scientific defense of Britain against the Third Reich, a partisan of radar. He applied to air and naval operations his own combination of personal involvement and objective analysis, crudely quantitative and mathematical where possible. Out of this work grew modern operations research. When he left the antiaircraft command in March, 1941, to help with the U-boat threat, the air battle over Britain having been won, his general complained: "They have stolen my magician."

He was no magician, but he joined thought to action. One night at headquarters of the Western approaches his attention was drawn by a wall map of estimated U-boat positions in the Atlantic. He calculated from the hours flown by the antisubmarine aircraft that they ought to be sighting four times as many U-boats as they reported. From informed estimates of how much time the U-boats spent on the surface he concluded that in three cases of the four the U-boats saw the plane and submerged before they were sighted. Not long afterward all the aircraft of the Coastal Command were painted white, to reduce the chance of their being seen against a light sky. (All other bombers were painted black, to reduce the chance of their being picked up by searchlights.) It made a distinct difference.

Lovell lays out the complex story of Blackett's unsuccessful opposition to the bomber campaign to dehouse the German worker. Blackett wrote in *Scientific American* in 1961 of the "fanatical belief in the efficacy of bombing.... The only major campaign in modern history in which the traditional military doctrine of waging war against the enemy's armed forces was abandoned for a planned attack on its civilian life was a disastrous flop. I confess to a haunting sense of personal failure."

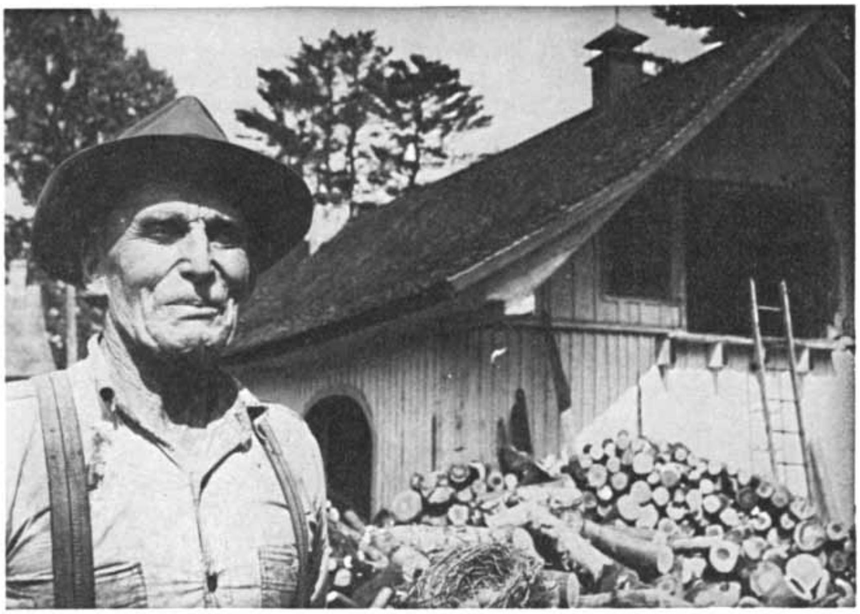
He was of course not always right. He saw early that the atomic bomb could not be a five-million-pound two-year venture, as was argued by its early proponents. At first, however, he regarded nuclear weapons as bringing no discontinuity in warfare. That cost him much in the Establishment world of the cold war; Prime Minister Attlee once rebuffed a report of Blackett's with the astonishing remark that "the author, a distinguished scientist, speaks on political and military problems on which he is a layman." Although he held the American Medal for Merit for his wartime work, "on one occasion he was arrested when an intended overflight from Mexico to Canada had to make a refuelling stop in the United States."

He was a cosmopolitan, at home in many lands. He answered a question put by Jawaharlal Nehru in 1947, estimating that the newly independent Indian armed forces could be Indianized well enough within two years to fight on their own frontiers. Their British commanders had reckoned on a generation. With that beginning he "came to love the country and its people," and he gave much time to visits and advice impinging on many facets of Indian national life. He spoke and wrote on the economic gap between the rich and the poor, which he hoped science and technology could with goodwill help to narrow. That there is much more than science involved he learned both in ministries of the Labor government in the 1960's and around the world. "Science is no magic wand" to convert a poor country into a rich one; the unities of wartime do not justify the inference that in peacetime people will do what they ought to do for the general good.

There is too much fascination in this man's life to summarize. He and his wife celebrated their 50th wedding anniversary before his death in 1974; Patrick and Pat (as Costanza, Lady Blackett, is often known) walked all that time across the world of physics and government, a striking pair, generous, tasteful, interested without pose, assured without vanity, hopeful without illusion. We remain in debt to him, a debt we can meet in his own style, seeking workable ways to replace war with peace, to narrow the gap between the rich and the poor and to understand this world on scales large and small.

FOLKTALES TOLD AROUND THE WORLD, edited by Richard M. Dorson. The University of Chicago Press ($17.50). "Fairy tales should be charming, delightful, beautiful, touched with an aura of magic and mystery befitting the childlike minds." Professor Dorson's ironic comment touches a nerve. Readable literary versions of the old peasant tales, which are unquestionably laden with symbolic matter of real weight in the formation of personality and the transmission of culture, are justly popular. The brothers Grimm naturally glossed and smoothed out the ruder oral utterances of their peasant informants. Every reader knows how distinct the way we talk is from any language that from the start is written down. It is clear that a true report of oral literature, which includes but goes far beyond the Märchen, the European magical fiction collected by the Grimms, cannot be found in the usual well-written collections, even those

P. M. S. Blackett at Cambridge in 1932



Florent Lemay, teller of folktales, from **Folktales Told around the World**

by some scholars. It almost demands acoustic recording.

This unique volume, compiled by the director of the Folklore Institute at Indiana University, the scholarly center of American folklore studies, seeks to redress the long-standing distortion. It is the real thing, the evidence itself. Even here, however, we stand parted from the reality of a fascinating set of about 100 tales of varying length and diverse form by the barriers of language; a translator has had to restate most of them, from the Japanese or the Xhosa or the Rawang ("just one of hundreds of Tibeto-Burman languages spoken" in the isolated valleys fronting the peaks of everlasting snow). To translate may traduce, but there is nothing else to be done. Some of the stories repeat foreign speech line by line, and more include the catch phrases or odd words that add dramatic accent (as when clever Chameleon rolls along hidden inside the dry kabo fruit, "kpikiri, kpikiri").

All these stories are reported from accurate transcriptions of fluid spoken performances, caught as they were told, not as they have been fixed in smooth text. Each has a note prepared by the collector to evoke place and time, say in 1955 around the bonfires in a Scottish berry field as the gypsy tinker families whiled away their evening. These are living tales, nearly all of them collected within the past two decades. The living tellers are here too, most of them identified by name and condition; we see photographs of a score of these talented people. Here is Mrs. Zsuzsanna Palkó, an unlettered, unworldly peasant woman looking like a dried crabapple, who saw her first full-length mirror in Budapest at the age of 74 and who has borne the official title of Master of Folklore. She is a specialist in Märchen, able to expand any text once the mood strikes her, while maintaining "the detached epic manner." For contrast consider the cheery, strong Jaime Rodríguez Huillca, "a natural group leader" at the age of 13, whose father, a *curandero* and construction laborer from Cuzco in Peru, taught Jaime the Spanish stories he retells with such effect to the schoolboys in a squatter settlement near Arequipa.

The tale needs a background, which often only the collector can supply to us who dwell far from the culture out of whose best-known beliefs the story often takes its meaning. Finally, the oral literature is much more varied in genre than the corpus of the brothers Grimm. Even they collected Sagen, although there is no translation of those volumes into English. Legends are taleworthy happenings rather than developed fiction; they are local, fragmentary, often a set of episodes unified by some landmark or noteworthy person rather than by plot. Then there are bilingual stories and dialect stories, notoriously hard to enjoy from the printed page. Brief anecdotes, jokes and humorous inventions are the kind of material whose representation in the great tape archives is most rapidly growing. All are sampled here, not easy reading, not always as memorable as Cinderella but so varied and intricate a fabric of word and thought that the reader is entangled by the whirring loom of human creativity.

Here are three abbreviated samples. In Japan they recall, out of a dim and fearful past, the times of "sixty canyon abandonment" when people who had reached 60 were by custom thrown into a mountain canyon. One father was instead hidden by his son under the porch. The lord of the country liked to set his subjects difficult tasks. This time it was to weave a rope from ashes. No one could do it. But the father explained to the son how to weave a tight rope and burn it to ashes. When the son showed it to the lord, he received high praise. Next the lord demanded a conch shell with a thread passed through it. The father again knew how: tie a rice grain to the end of the thread, give the rice to an ant and make the ant crawl through the shell. This time the lord demanded to know how his young liege man had managed the feat. When the explanation was given, all realized the "old people are very wise," and sixty canyon abandonment was abolished.

From northern Mindanao in the Philippines comes a long, moving tale in the Manobo tongue. It tells of seven swan maidens, with monkey attendants that are tempted by bananas, and how one of the swans is captured and married. Later she vanishes to a place where the moon rises and the sun sets. There finally the great chief displays his seven identical daughters to the long-seeking husband. At first he cannot pick out his wife, the mother of his own little daughter back home. Then he looks carefully at the women's fingers, and in the center of the little finger of one of them he can "recognize this tiny mark of a needle." "You can take her," said the chief, "for she is indeed your wife."

Some storytellers are not far away. Such an artist is Burt Mayotte, the mechanic. He learned a French-Canadian way of speaking English. ("The dialect is my own grandfather's.") The story he made up, and he told it "all over the Green Bay zone with Standard Oil." Paree has gone to the carnival to see the wrestlers. There they challenge all comers to stay with the Tiger, "'Twenty-fi' dollaire for ten min*oots*.' Et Maudit, Hi fin' dat place fas'.... She's look prettee dam tough dis man.... He have much hair on dee bellee, han he look like dee Tigaire for w'ich he is call.... Han I look prettee dam tough too." The wrestling goes forward, but the Tiger is genuinely fierce. "Paree, Paree, Hi cry like ba*bee*. Oh, dat hurt.... Almost Hi teenk Hi don't leave for ten min*oots*. But Hi don't give hup. Wid hall dee cries han dee knots, Hi see my chanse." Paree takes his last bit of courage and bites hard. Alas, Paree has bitten himself. (It is put more graphically in the original.)

The book is rated a mild *X*, which should make it *PG* for any parents who would like a young reader to learn something of the facts of life as men and women speak of them honestly, in the classical domain of Rabelais and Boccaccio, and over all the earth.

THE HOHOKAM: DESERT FARMERS & CRAFTSMEN (EXCAVATIONS AT SNAKETOWN, 1964–1965), by Emil W. Haury. The University of Arizona Press ($19.50). About 100 years ago, once peace with the Apaches had made an isolated settlement safe, Pima families moved out to set up new villages where they could irrigate their crops with the waters of the Gila River. By 1910 the Gila channel had become so deep that downstream agriculture without pumping was dead. The riverbed is dusty and windblown now, about where Interstate Route 10 crosses it. Before the Pima or the highway, however, an enduring settlement was already there, continuous in its culture from a few centuries B.C. until about A.D. 1450.

In their tongue the Pima called the place Snaketown, noting an abundance of rattlers. The snakes were common because rodents were too; they could nest more easily in the relatively soft ground of the mounds, a few dozen of which are scattered over a square kilometer of the flat reddish terrain, some reaching a height of three or four meters. The mounds were first excavated in the 1930's, revealing the pots, figurines, house foundations, hearths and ball court of the ancient dwellers, the Hohokam, a Pima term that can be construed as "those who have perished." The Hohokam were probably the forebears of the Pima; similar settlements dot the central Arizona desert throughout the drainage of the Gila River, above Gila Bend.

This handsome and well-written volume is the second full-scale excavation report on Snaketown. Haury is now a senior American archaeologist "returning to the scene of the crime"; he dug there in the 1930's as a junior, under the direction of the critical and heterodox Harold Gladwin. The new book is in a familiar genre; there are pages of excellent maps, photographs of sherds, drawings of pots, clay figures, tools and necklaces, tables of design motifs and counts of pot types. The land is made real in text and illustration; the giant saguaro cactus and the mesquite still bear their delicious fruit in that arid scene. A number of expert appendixes end the book: instrument surveys, the detective work of the botanists and zoologists, details of the cremations (often giving evidence of the multiple burial of a single cremated individual).

The volume is rich with method and result. Here one can mention only a few points of general interest. The people lived by the staple maize, watered by expertly made canals tapping the waters of the Gila. They were protein-short, it seems. Beans are scarcer than one would expect, and the bone remains in the mounds show that the little mule deer of the foothills, the main source of meat, were never very abundant. More directly, residues of human feces reveal a remarkable jumble of bones too fine to be those of the dog or the owl, which were sometimes roasted. They are identifiably the remains of every kind of small creature that "walked, crawled, slithered, flew, or swam." The Hohokam must have crunched up "anything that would fill the void."

Finds of designs on shell go back to the 1930's. Now we have etchings on clamshell; the resist drawing was made in some piny pitch and the shell was treated with the vinegary residue of saguaro-fruit wine gone sour. The neat intaglio patterns are evidence of a technological achievement unique to the Hohokam. (No earlier shell etching is known, although Chinese etching of bronze and European etching of iron is older.) The particular rich tidal flats on the Gulf of California where these desert folk went to collect their shells have been identified. Another indication of discovery is the apparent, if not quite demonstrated, use of stone palettes, flat rectangles of stone with a shallow central depression, to display the magical color change of white lead carbonate to red lead oxide in the crematory fire. A few globules of metallic lead in spoons hint that the Hohokam were at the threshold of learning how to smelt lead; the only other sign of metal is one necklace of tiny copper bells, but it was without doubt acquired from the west coast of Mexico.

One of the stimuli to the digging of the 1960's was the new means of dating that the physical sciences had given archaeology since the 1930's. They are all here, done by experts, often by more than one laboratory. It is fine work, but by and large the results are somewhat discordant. Just as 40 years ago we still trust the judgment of the archaeologist to recover among the counting and the sampling errors what he thinks the overall picture shows. Haury sees a long run of almost 2,500 years beside the Gila, "a cultural form or style that ensured unusual stability.... Protective of their environment, [the Hohokam] forged a social and economic system that... endured.... Few people can match that record." It is striking to perceive that this quiet life held the promise of novelty, witnessed by the shell etching and the protometallurgy. As we learn more about the processes of human change we can expect to find many cases where an epochal transformation did not quite come about.

This is no example of the new archae-

Excavated house floors at Snaketown, from **The Hohokam: Desert Farmers & Craftsmen** ology. "My historico-cultural approach will . . . be a disappointment to those who are searching for articulation of hypotheses of a processual nature with appropriate tests. These will come in due time." Meanwhile we have a first-rate book, written humanely and with a flair for holding the interest of the general reader. It is sure to engage the imagination of anyone who has known the southwestern desert.

ABYSSAL ENVIRONMENT AND ECOLOGY OF THE WORLD OCEANS, by Robert J. Menzies, Robert Y. George and Gilbert T. Rowe. John Wiley & Sons ($32). The continental slopes rise to the continental shelves out of the full ocean depths within a distance of a few hundred miles. Three-fourths of the sea floor, then, lies deeper than three kilometers. It is an area twice that of all dry land: vast plains dotted with seamounts, broken by rifts and mountain ridges and cut by a couple of dozen flat-bottomed trenches dropping to a maximum depth of 10 kilometers. It is a desert of unending night, never lighted by the sun, free even from cosmic rays and of course devoid of living green plants. The water is salt and cold, below zero Celsius near the poles and averaging some two degrees C. worldwide.

It is the bottom life at depth these three marine biologists seek out and describe. There are many line drawings and photographs of the creatures; the most striking illustrations display the authors' own catch seaward from Cape Hatteras, down to a depth of five kilometers. A shelf crab at 420 meters is succeeded by a hermit crab at 800 meters and a brittle starfish at two kilometers. Down we sound until a stalked barnacle shows up at 4.6 kilometers and a different brittle starfish appears on the Hatteras plain, the "tranquil zone" at 5.3 kilometers. Even the inexperienced reader finds these forms plausible enough; they would not amaze him if he found them cast up on Jones Beach.

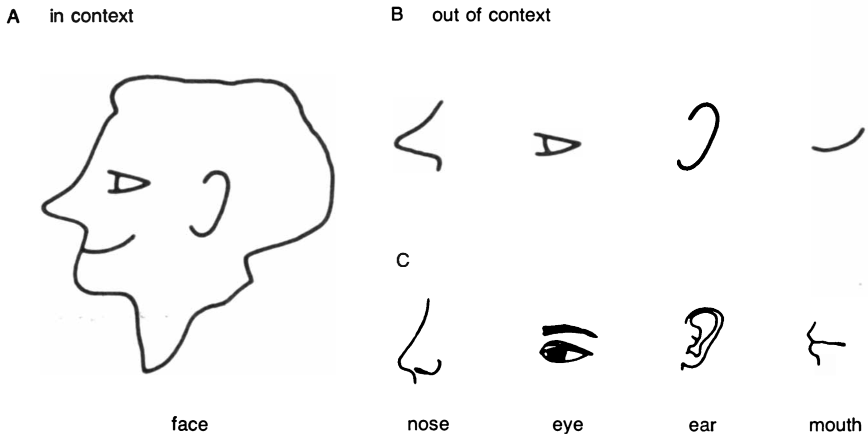
Showy animals like these, however, are by no means common. It takes many a flash before the cameras catch more than burrows and ripple marks. There are burrowers too; one sea star commonly taken in trawls has never been caught by any camera. It must dwell just out of sight, burrowed snugly in the ooze. Most of the animals make a living from the organic particles in the ooze or in the water, all derived from the organic fragments continuously settling out of the sunny world of plankton above. A few forms are predators on their neighbors; nearly all are pallid and unpigmented. They are blind as well, often having a lens but lacking, so to speak, a shutter stop and film; both iris and sensitive retinal pigment are gone. Many of the forms—the mud eaters and the scuttlers, the stalked sea pens and the sponges—are found over a wide range of depths. Species vary with depth and also with position; the authors have collected data from the Antarctic to north of Siberia, from Peru to the Bay of Biscay. They find no specific ocean-floor life, no worldwide abyssal fauna; rather, there are many abyssal groupings, each bearing more resemblance to the collection of forms found on the species-rich continental margins nearby than to their deep-living kin across the world.

Even the trenches are far from empty. Plenty of animals live there: clams, stars and scuttling isopods are more abundant in the right place 10 kilometers down than in the wide red-clay basins only half as deep. Everything points to the authors' thesis: the life of the depths is relatively new, adapted only quantitatively from the fauna of the slopes. These animals are not archaic: they are related more to Tertiary forms than to Paleozoic ones. The old blueblood families are found in the shallower waters; the depths themselves and their forms of life are not so old. The surface temperatures of the past determined the temperature of the depths: glacial ages, coldest seas. When the ocean was most uniform in temperature, the spread of forms to the depths was easiest. To put it simply in energetic terms—too simply—the effects of 1,000 atmospheres of pressure are no more than those of 10 degrees of temperature. The abyss is today a cold world, the pressure only a perturbation. Most of the present deep-sea animals originated on the polar undersea slopes, which have existed for no longer than tens of millions of years.

The ocean depths are not unchanging but are constantly evolving, and with them the deepest and most hidden of all life also evolves. A slope fauna fitted mainly for life on fine mud in dark cold water has dispatched migrants to great depths. The trenches too are constantly changing; their flat floors are layered with sand, derived from slumps and slides. ("It is therefore most probable that the trench environment is one of the least stable regions on earth, as far as the bottom fauna is concerned. The bottom fauna is being alternately buried and repopulated, and it is probable that this cycle has continued frequently since the Pleistocene glaciations.") These are pioneers, these sightless creatures of the depths, not relicts of an endless pitch-dark past. They are sparse: area for area their biomass is only about a thousandth the average of the shallow bottoms.

This diverse, argumentative, lively and not too technical monograph offers

Bottom organisms at 6,200 meters, from **Abyssal Environment and Ecology of the World Oceans**



An illustration of the part-whole context, from **Memory and Attention**

an unusual look at an emerging concept; it will be of interest to see how the authors' novel thesis fares.

MEMORY AND ATTENTION: AN INTRODUCTION TO HUMAN INFORMATION PROCESSING, by Donald A. Norman. John Wiley & Sons, Inc. ($10.95). Photographic memory—eidetic imagery is the psychologists' term—was once a major area of study in psychology. "Today it is hardly studied, and no one quite knows what to make of it." It is real: "The tests appear foolproof. Many psychologists are skeptical, but without any good reason." The talent is rare enough; the young woman who is the modern American example will understandably no longer perform as a subject, and "no other person has yet been found with a similar facility." The papers that support this rather strong claim are cited, but not much more is said. Bear in mind, however, that you too can remember thousands of visual images. You can usually pick out any illustration you have seen before, particularly if it is a color photograph, even if you have not seen the picture for years. So experiment showed in 1973, described under the title "Learning 10,000 Pictures."

The sample will display the nature of this compact, agreeable little textbook by a psychologist at San Diego, now published in its second edition. His is strictly a black-box science: input and output, with no look at that intricate circuitry inside, not even to discuss the classical studies of brain injury. It surveys the topics of its title in a series of chapters, mostly accompanied by several pages of readings from the literature of the subject, as old as William James (actually older, with Cicero and Quintilian on the mnemonic art) and as new as 1975. The approach is in part historical. The earlier work is described, and its conclusions are pressed—usually not very far—until they fail. Then an alternative set of experiments and the model they gave rise to are followed, and the tentative state of today's stance is adduced. (It is generally that both elements have some truth in them.) "These are indeed exciting times" for the workers in the field. They nonetheless see there is a long, long way to go.

One theme is the role of meaning and context. We attend to and remember what in some sense we understand. (Hence the air of suspicion that hangs over the eidetic experiment.) In the 1930's a Dutch experimenter, J. R. Stroop, showed that it was harder to name the color of lettering when the sample was the word "blue" presented in red. It seemed quaint then, but this semantic level now enters in almost everywhere. Norman uses the terms data-driven for what we process from the outside and concept-driven for what we take from internal hypotheses to reconstruct the world we perceive. Both kinds of processing operate, and both claim resources available only within limits.

The chess master seems to have an incredible memory. Presented with random configurations of pieces, however, he does no better than "regular people." The experimenters estimated that the chess master has a vocabulary of chess patterns about as large as the language vocabulary of a well-read adult. Memory must have features at a high level of abstraction; the modern workers speak of conceptual frameworks, structural networks, schemata and scripts, all terms for one or another model of the flexible and subtle way we store information, actively as well as passively. Here there are studies of the recall of stories or drawings through the analysis of error. Such memories are in fact chiefly new constructions from internal operations on structural units. Norman also takes up the large question of consciousness, but only one neurological topic is discussed: the fascinating hemispheric specialization that has generated so much pop psychology (logic and intuition, algebra and geometry, science and art). This is certainly a readable (although slightly lopsided) book, with a lively air of work in progress.

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

THE ERADICATION OF SMALLPOX

SMALLPOX. C. W. Dixon. J. & A. Churchill Ltd, 1962.

SMALLPOX. Donald A. Henderson in *Maxcy-Rosenau Preventive Medicine and Public Health: Tenth Edition,* edited by Philip E. Sartwell. Appleton-Century-Crofts, 1973.

SMALLPOX ERADICATION IN WEST AND CENTRAL AFRICA. William H. Foege, J. D. Millar and D. A. Henderson in *Bulletin of the World Health Organization,* Vol. 52, No. 2, pages 209–222; 1975.

THE PHOTOVOLTAIC GENERATION OF ELECTRICITY

DIRECT USE OF THE SUN'S ENERGY. Farrington Daniels. Yale University Press, 1964.

SATELLITE SOLAR POWER STATION. P. E. Glaser in *Solar Energy,* Vol. 12, No. 3, pages 353–361; 1969.

GROWTH OF CONTROLLED PROFILE CRYSTALS FROM THE MELT, PART III: THEORY. Bruce Chalmers, H. E. LaBelle, Jr., and A. I. Mlavsky in *Materials Research Bulletin,* Vol. 6, No. 8, pages 681–690; August, 1971.

ENERGY: THE SOLAR-HYDROGEN ALTERNATIVE. J. O'M. Bockris. John Wiley & Sons, 1975.

SOLAR CELLS. HAROLD J. HOVEL IN *Semiconductors and Semimetals: Vol. 11.* Academic Press, 1975.

NEUTRON-SCATTERING STUDIES OF THE RIBOSOME

NEUTRON DIFFRACTION. G. E. Bacon. Clarendon Press, 1962.

NEUTRON SCATTERING. B. P. Schoenborn and A. C. Nunes in *Annual Review of Biophysics and Bioengineering,* Vol. 1, pages 529–552; 1972.

NEUTRON SMALL-ANGLE SCATTERING OF BIOLOGICAL MACROMOLECULES IN SOLUTION. Heinrich B. Stuhrmann in *Journal of Applied Crystallography,* Vol. 7, Part 2, pages 173–178; April 1, 1974.

DETERMINATION OF QUATERNARY STRUCTURE BY SMALL ANGLE NEUTRON SCATTERING. D. M. Engelman and P. B. Moore in *Annual Review of Biophysics and Bioengineering,* Vol. 4, pages 219–241; 1975.

NEUTRON SCATTERING MEASUREMENTS OF SEPARATION AND SHAPE OF PROTEINS IN 30S RIBOSOMAL SUBUNIT OF *ESCHERICHIA COLI:* S2–S5, S5–S8, S3–S7. Donald M. Engelman, Peter B. Moore and Benno P. Schoenborn in *Proceedings of the National Academy of Sciences of the United States of America,* Vol. 72, No. 10, pages 3888–3892; October, 1975.

COSMIC GAMMA-RAY BURSTS

OBSERVATIONS OF GAMMA-RAY BURSTS OF COSMIC ORIGIN. Ray W. Klebesadel, Ian B. Strong and Roy A. Olson in *Astrophysical Journal Letters,* Vol. 182, No. 2, Part 2, pages L85–L88; June 1, 1973.

RECENT OBSERVATIONS OF COSMIC γ-RAY BURSTS. T. L. Cline in *Annals of the New York Academy of Sciences: Seventh Texas Symposium on Relativistic Astrophysics,* Vol. 262, pages 159–163; October 15, 1973.

A PRELIMINARY CATALOGUE OF TRANSIENT COSMIC GAMMA-RAY SOURCES OBSERVED BY THE *VELA* Satellites. Ian B. Strong, Ray W. Klebesadel and Roy A. Olson in *Astrophysical Journal Letters,* Vol. 188, No. 1, Part 2, pages L1–L3; February 15, 1974.

OBSERVATIONS OF γ-RAY BURSTS. I. B. Strong, W. D. Evans and R. W. Klebesadel in *Annals of the New York Academy of Sciences: Seventh Texas Symposium on Relativistic Astrophysics,* Vol. 262, pages 145–158; October 15, 1975.

THEORIES OF γ-RAY BURSTS. M. Ruderman in *Annals of the New York Academy of Sciences: Seventh Texas Symposium on Relativistic Astrophysics,* Vol. 262, pages 164–180; October 15, 1975.

WHITE-LIGHT HOLOGRAMS

PHOTOGRAPHY BY LASER. Emmett N. Leith and Juris Upatnieks in *Scientific American,* Vol. 212, No. 6, pages 24–35; June, 1965.

ADVANCES IN HOLOGRAPHY. Keith S. Pennington in *Scientific American,* Vol. 218, No. 2, pages 40–48; February, 1968.

LASERS AND LIGHT: READINGS FROM SCIENTIFIC AMERICAN. W. H. Freeman and Company, 1969.

HOLOGRAPHY: WITH AN INTRODUCTION TO THE OPTICS OF DIFFRACTION, INTERFERENCE, AND PHASE DIFFERENCES. H. Arthur Klein. J. B. Lippincott Company, 1970.

THE SOCIAL ORDER OF JAPANESE MACAQUES

PRIMATE SOCIETIES: GROUP TECHNIQUES OF ECOLOGICAL ADAPTATION. Hans Kummer. Aldine Publishing Company, 1971.

THE SOCIAL BEHAVIOR OF MONKEYS. Thelma Rowell. Penguin Books, 1972.

BIOLOGICAL BASES OF HUMAN SOCIAL BEHAVIOR. Robert A. Hinde. McGraw-Hill Book Company, 1974.

CONCEPTUAL AND METHODOLOGICAL PROBLEMS ASSOCIATED WITH THE STUDY OF AGGRESSIVE BEHAVIOR IN PRIMATES UNDER SEMINATURAL CONDITIONS. G. Gray Eaton in *The Neuropsychology of Aggression,* edited by Richard E. Whalen. Plenum Press, 1974.

MALE DOMINANCE AND AGGRESSION IN JAPANESE MACAQUE REPRODUCTION. G. Gray Eaton in *Reproductive Behavior,* edited by William Montagna and William A. Sadler. Plenum Press, 1974.

DUST STORMS

A COMPARISON BETWEEN LABORATORY AND ATMOSPHERIC DENSITY CURRENTS. J. E. Simpson in *Quarterly Journal of the Royal Meteorological Society,* Vol. 95, No. 406, pages 758–765; October, 1969.

EFFECTS OF THE LOWER BOUNDARY ON THE HEAD OF A GRAVITY CURRENT. John E. Simpson in *Journal of Fluid Mechanics,* Vol. 53, Part 4, pages 759–768; June 27, 1972.

AN AMERICAN HABOOB. S. B. Idso, R. S. Ingram and J. M. Pritchard in *Bulletin of the American Meteorological Society,* Vol. 53, No. 10, pages 930–935; October, 1972.

LOCAL STRUCTURE OF MATURE DUST STORMS. G. I. Barenblatt and G. S. Golitsyn in *Journal of the Atmospheric Sciences,* Vol. 31, No. 7, pages 1917–1933; October, 1974.

A DESERTED MEDIEVAL VILLAGE IN ENGLAND

DESERTED MEDIEVAL VILLAGES: STUDIES. Edited by Maurice Beresford and John G. Hurst. Lutterworth Press, 1971.

MAN MADE THE LAND: ESSAYS IN ENGLISH HISTORICAL GEOGRAPHY. Edited by Alan R. H. Baker and John B. Harley. Newton Abbot, David & Charles, 1973.

WHARRAM PERCY in *Current Archaeology,* No. 49, pages 39–49; March, 1975.

MATHEMATICAL GAMES

AMUSEMENTS IN MATHEMATICS. Henry E. Dudeney. Dover Publications, Inc., 1958.

THE BEST MATHEMATICAL PUZZLES OF SAM LOYD. Edited by Martin Gardner. Dover Publications, Inc., 1959.

MORE MATHEMATICAL PUZZLES OF SAM LOYD. Edited by Martin Gardner. Dover Publications, Inc., 1960.

"For the record, my Omega represents efficient spending. Off the record, I just love it. They sure built one for me."

When somebody's operating on a platform of economy and clear-headed thinking, their choice of automobiles ought to set a good example. And this one does. Oldsmobile's Omega combines driving comfort with fiscal responsibility.

It's a car built for the long run. Just look at the standard equipment: energy-absorbing bumpers, front disc brakes, and an exhaust system that stands up to rust. You can even go as far as 22,500 miles between tune-ups.

Like to preserve your options? Equip your Omega with available power brakes and power steering,

steel-belted radials and air conditioning. And more.

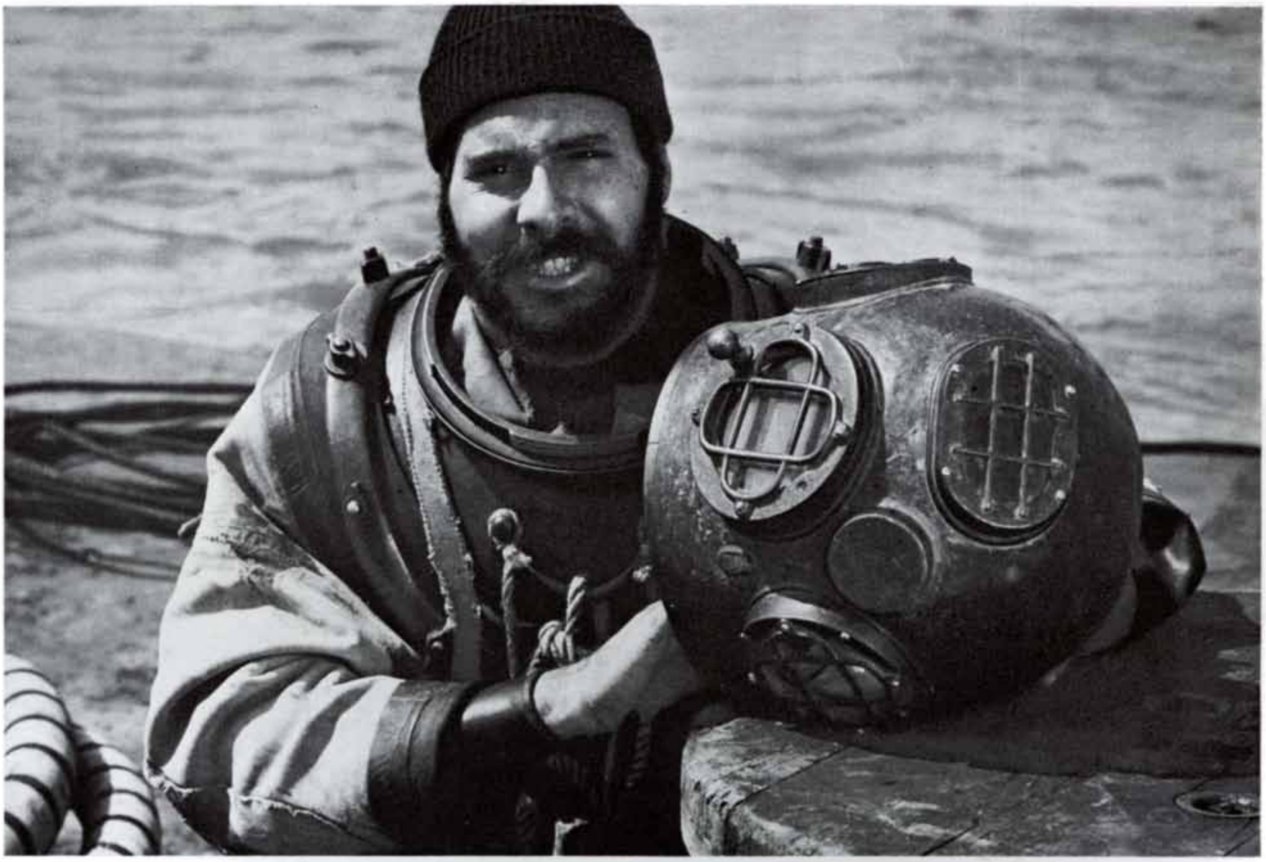
Best of all, you can save on gas and still get the looks and comfort you want. As you can see above. (After all, how many candidates can campaign in style like this and still get Omega's kind of mileage? EPA estimates are 27 mpg in the highway test, 16 in the city test, with standard 231 V-6 engine, manual transmission and rear axle. Your mileage depends on how you drive, your car's equipment and its condition. EPA ratings lower in California.)

So, if comfort, thrift, and Oldsmobile quality are the sort of combination it takes to get your vote, Omega looks like just the ticket.

Oldsmobile
Omega
Can we build one for you?

DEWAR'S® PROFILES

(Pronounced Do-ers "White Label")



BLENDED SCOTCH WHISKY • 86.8 PROOF • © SCHENLEY IMPORTS CO., N.Y., N.Y.

MICHAEL SANDLOFER

HOME: Nyack, New York

AGE: 28

PROFESSION: Commercial deep-sea diver and lecturer in underwater life cycles

HOBBIES: Marine specimen collecting, skydiving.

MOST MEMORABLE BOOK: "The Frail Ocean" by Wesley Marx

LAST ACCOMPLISHMENT: Established "Northwind," a nautical sea museum in Nyack, New York.

QUOTE: "As the final earthly frontier, the sea is the ultimate source of food and minerals. Man must learn to harness its power in a positive way, and be able to replace whatever is removed from it."

PROFILE: Inquisitive. Devoted to expanding man's knowledge of the ecological potentials of the world beneath the sea.

SCOTCH: Dewar's "White Label®"