SCIENTIFIC AMERICAN

ECOLOGY OF THE DUNG BEETLE

$2.00

November 1979

Mirada by Dodge. Engineered from scratch to be fuel efficient.

If you own a Thunderbird, Grand Prix, Monte Carlo, Cutlass Supreme, Regal

Best equipped personal car in the mid-size class.
Automatic Transmission
Power Steering
Power Brakes
AM Radio
Tinted Glass
Bucket Seats
White side-wall radials
Working Gauges
Brushed-Metal Instrument Panel

17 EPA EST. MPG. **25** HWY. MPG.

Use this estimate for comparisons. Your mileage may vary depending on your speed, weather and trip length. Your actual highway mileage will probably be less. California estimates **16** EPA estimated mpg. 24 EPA estimated highway.

Leather interior available.

But not at the expense of styling, luxury or performance.

or Cougar, you owe it to yourself to test drive The New 1980 Dodge Mirada.

Console with floor shift available.

1980
DODGE
MIRADA

Dodge

Discover more.

Let your mature taste lead you from the finest premium Scotch on up to the most expensive 12-year-old Scotch in the world. The Glenlivet. The ultimate in Scotch.

Most premium Scotch is blended and depends on several whiskies for taste and smoothness.

Yet the unblended character of The Glenlivet is noble enough to stand alone. In this distinctive Scotch whisky you'll experience an exceptional smoothness and full-bodied richness, unmatched in all other premium Scotch.

Try the taste beyond premium Scotch.

The Glenlivet
Unchanged since 1824.

THE TASTE BEYOND PREMIUM SCOTCH

Established 1845 **SCIENTIFIC AMERICAN** November 1979 Volume 241 Number 5

ARTICLES

50 **LAND-BASED INTERCONTINENTAL BALLISTIC MISSILES, by Bernard T. Feld and Kosta Tsipis** The vulnerability of this component of the "triad" is exaggerated.

62 **THE CAUSES OF DIABETES, by Abner Louis Notkins**
One form appears to result from the interaction of genetic makeup and environmental factors.

74 **SHAPE-MEMORY ALLOYS, by L. McDonald Schetky**
Formed into a shape at one temperature, such an alloy "remembers" the shape at another one.

98 **THE NEUTRAL THEORY OF MOLECULAR EVOLUTION, by Motoo Kimura**
It contends that most evolutionary change is caused not by selection but by random genetic drift.

130 **PRIMEVAL GALAXIES, by David L. Meier and Rashid L. Sunyaev**
The characteristics of the galaxies that formed after the "big bang" suggest they can be observed.

146 **THE ECOLOGY OF THE AFRICAN DUNG BEETLE, by Bernd Heinrich and George A. Bartholomew** For this kind of insect disposing of the dung of large mammals is a way of life.

158 **THE QUANTUM THEORY AND REALITY, by Bernard d'Espagnat**
Quantum mechanics curiously disagrees with the doctrine that the world is independent of mind.

182 **A NEOLITHIC AND IRON AGE SITE ON A HILLTOP IN SOUTHERN ENGLAND, by P. W. Dixon** A pre-Roman hill fort was superposed on works built 2,000 years earlier.

DEPARTMENTS

8 LETTERS
12 50 AND 100 YEARS AGO
17 THE AUTHORS
20 MATHEMATICAL GAMES
36 BOOKS
84 SCIENCE AND THE CITIZEN
192 THE AMATEUR SCIENTIST
206 BIBLIOGRAPHY

SCIENTIFIC AMERICAN (ISSN 0036-8733), PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $18 PER YEAR, U.S., ITS POSSESSIONS AND CANADA; $22 PER YEAR, ALL OTHER COUNTRIES. POSTMASTER: SEND ADDRESS CHANGES TO SCIENTIFIC AMERICAN, 415 MADISON AVENUE, NEW YORK, N.Y. 10017.

INFORMATION MANAGEMENT. DOES IT THREATEN THE WORK ETHIC?

The American work ethic has always assumed that if you worked hard, your efforts would be rewarded.

We wouldn't argue with that. But we would like to amend it a little.

The way we see it, working hard isn't nearly as important as working well.

Information Management isn't a way of sidestepping hard work. It's a way of making it more productive.

For example, there's no virtue in having to spend hours typing, retyping and re-retyping documents. So we make electronic typing systems that let you type, revise and retrieve information in a lot less time. Which gives you a lot more time to create and perfect it.

To make information easier to work with, we make a full range of copiers, duplicators and computer printers. Ones that not only reproduce, but can reduce, collate or even print in color.

To make information more accessible, we have Telecopier transceivers and communicating typewriters that transmit information crosstown or cross-country in minutes.

We even offer computer services that let you manage information without having to manage a computer.

So you see, Xerox information management systems are actually the epitome of the work ethic.

They work harder. So you can work better.

XEROX

THE COVER

The painting on the cover focuses on a ball of elephant dung that has been shaped by a male dung beetle of the species *Kheper platynotus,* which is one of more than 2,000 species of dung beetle active in East Africa (see "The Ecology of the African Dung Beetle," by Bernd Heinrich and George A. Bartholomew, page 146). The male is the insect at the lower left; he has begun to roll the completed ball a distance of several meters to a place where he will bury it. He pushes with his hind legs, standing on his head. Atop the ball is a female of the species; she will ride along passively and sink into the ground with the ball as the male digs a hole for it. When the ball has been buried, the beetles feed on it and then mate. The female lays one egg, and the remainder of the ball serves as food for the larva. Dung beetles compete so fiercely for the droppings of various mammals that dung is a scarce resource, notwithstanding the large herds of mammals manufacturing it. The competitiveness is suggested by the presence of two other beetles, which might try to steal the ball from its owners.

THE ILLUSTRATIONS

Cover painting by Tom Prentiss

Page	Source	Page	Source
10	Edward Bell	98–124	Adolph E. Brotman
25–33	Ilil Arbel	131–133	Gabor Kiss
50	Charles Moore, Black Star	134	Gabor Kiss; Kitt Peak National Observatory (*bottom right*)
52–61	George V. Kelvin	136–144	Gabor Kiss
63	A. Bennett Jenson and Kozaburo Hayashi, National Institute of Dental Research (*top*); Takashi Onodera, National Institute of Dental Research (*bottom*)	147–149	Bernd Heinrich, University of California at Berkeley
		151–156	Tom Prentiss
64–66	Patricia J. Wynne	159–177	Jerome Kuhl
67	Joseph R. Williamson, Washington University School of Medicine	182	P. W. Dixon, University of Nottingham
68–72	Patricia J. Wynne	184–187	Alan D. Iselin
73	A. Bennett Jenson	188–190	P. W. Dixon
74–75	Goodyear Aerospace Corporation	194	Stuart Travis
76–82	Dan Todd	197–202	Michael Goodman

UN MOMENT DE MARTELL

80 PROOF IMPORTED BY THE JOS GARNEAU Co , NEW YORK, N.Y.

DINNER IS OVER, BUT THE NIGHT IS STILL YOUNG.

WHAT BETTER TIME TO EXPERIENCE THE WORLD S FINEST COGNAC—MARTELL.

AFTER ALL, MARTELL COGNAC HAS HELPED MAKE EVENINGS MEMORABLE SINCE 1715.

MARTELL V.S.P., V.S.O.P., AND CORDON BLEU.

MARTELL COGNAC

LETTERS

Sirs:

Your report ["Science and the Citizen," SCIENTIFIC AMERICAN, August] on that masterly medical Munchausen, William McIlroy, does not tell the full story. After the fascinating detective story by Drs. Pallis and Bamji appeared in *British Medical Journal* in April many further accounts of McIlroy's past exploits were published in the correspondence columns of the journal. The final word came from Dr. Kennedy Cruickshank and his colleagues in Birmingham, who relate that our hero is alive and "as well as can be expected and in semiretirement [in an old people's home], no longer enacting his more complete performances." His comment on his biography was: "It doesn't do me credit."

McIlroy undoubtedly did cost the British taxpayer a good deal of money, though perhaps not as much as your account suggests, but one cannot regard his activities as being totally without value. The needs of a patient suffering from the Munchausen syndrome, though quirky, are very real and are satisfied, up to a point, by the kind of management McIlroy got. In retrospect the overinvestigation and overtreatment seem to have been wasteful of money

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 40 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

Great protection for anyone moving up.

People on the way up often need something more than temporary life insurance protection. And New York Life has just the answer: our whole life policy.

It's permanent insurance that protects you for as long as you live. You never have to renew or convert it. Year by year, too, it builds an ever-larger cash value that gives your policy remarkable flexibility. In an emergency, for example, you can borrow against the cash value. Or use it later towards retirement.

Why not talk over permanent protection with your New York Life Agent soon. It's great protection for anyone moving up.

New York Life. For all of your life.

New York Life Insurance Company, 51 Madison Avenue, New York, New York 10010. Life, Group and Health Insurance, Annuities, Pension Plans.

and skills, but who is to discern the genuine Munchausen prospectively? Underinvestigation and undertreatment of patients with truly obscure illnesses are much more serious. One wonders how McIlroy's counterparts fare in countries with a less egalitarian medical service.

D. B. STEWART, M.D.

Professor of Zoology
Brandon University
Brandon, Manitoba

Sirs:

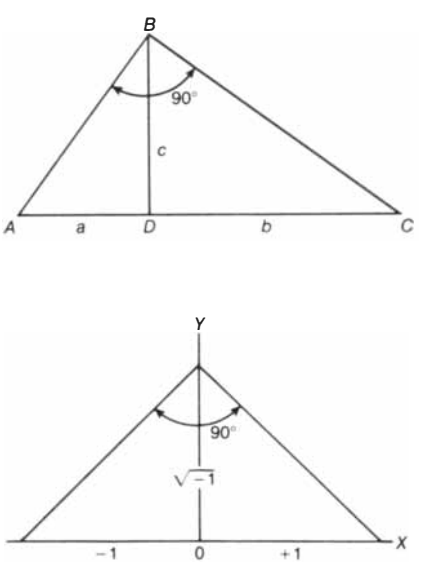
Martin Gardner's article "The Imaginableness of the Imaginary Numbers" ["Mathematical Games," SCIENTIFIC AMERICAN, August] is most interesting, but I suggest a further simplification. The following application of plane geometry removes much of the mystery of imaginary numbers.

Consider the right triangle *ABC,* with *BD* perpendicular to *AC* [*see upper illustration below*]. A proposition in geometry states that *c* is the geometric mean of *a* and *b,* that is, *c* equals $\sqrt{ab}$. This is provable from the similar triangles *ABD* and *BCD* wherein c/a equals b/c and hence *c* equals $\sqrt{ab}$.

Construct a right triangle [*see lower illustration below*] with *D* at the origin in conventional *X,Y* coordinates, with *b* equal to $+1$ and *a* equal to -1. The altitude of this triangle will then be the geometric mean of $+1$ and -1, namely the square root of -1. Hence it follows that imaginary numbers inherently plot at right angles to the *X* axis, on or parallel to the conventional *Y* axis.

WARREN A. SPOFFORD

Tyler, Tex.



The mystery of imaginary numbers simplified

Bone Fone™

A new concept in sound technology may revolutionize the way we listen to stereo music.

The Bone Fone surrounds your entire body with a sound almost impossible to imagine.

You're standing in an open field. Suddenly there's music from all directions. Your bones resonate as if you're listening to beautiful stereo music in front of a powerful home stereo system.

But there's no radio in sight and nobody else hears what you do. It's an unbelievable experience that will send chills through your body when you first hear it.

AROUND YOU

And nobody will know you're listening to a stereo. The entire sound system is actually draped around you like a scarf and can be hidden under a jacket or worn over clothes.

The Bone Fone is actually an AM/FM stereo multiplex radio with its speakers located near your ears. When you tune in a stereo station, you get the same stereo separation you'd expect from earphones but without the bulk and inconvenience. And you also get something you won't expect.

INNER EAR BONES

The sound will also resonate through your bones–all the way to the sensitive bones of your inner ear. It's like feeling the vibrations of a powerful stereo system or sitting in the first row listening to a symphony orchestra–it's breathtaking.

Now you can listen to beautiful stereo music everywhere–not just in your living room. Imagine walking your dog to beautiful stereo music or roller skating to a strong disco beat.

You can ride a bicycle or motorcycle, jog and even do headstands–the Bone Fone stays on no matter what the activity. The Bone Fone stereo brings beautiful music and convenience to every indoor and outdoor activity without disturbing those around you and without anything covering your ear.

SKI INVENTION

The Bone Fone was invented by an engineer who liked to ski. Every time he took a long lift ride, he noticed other skiers carrying transistor radios and cassette players and wondered if there was a better way to keep your hands free and listen to stereo music.

So he invented the Bone Fone stereo. When he put it around his neck, he couldn't believe his ears. He was not only hearing the music and stereo separation, but the sound was resonating through his bones giving him the sensation of standing in front of a powerful stereo system.

AWARDED PATENT

The inventor took his invention to a friend who also tried it on. His friend couldn't believe what he heard and at first thought someone was playing a trick on him.

The inventor was awarded a patent for his idea and brought it to JS&A. We took the idea and our engineers produced a very sensitive yet powerful AM/FM multiplex radio called the Bone Fone.

The entire battery-powered system is self-contained and uses four integrated circuits and two ceramic filters for high station selectivity. The Bone Fone weighs only 15 ounces, so when worn over your shoulders, the weight is not even a factor.

BUILT TO TAKE IT

The Bone Fone was built to take abuse. The large 70 millimeter speakers are protected in flexible water and crush resistant cases. The case that houses the radio itself is made of rugged ABS plastic with a special reinforcement system. We knew that the Bone Fone stereo may take a great deal of abuse so we designed it with the quality needed to withstand the worst treatment.

The Bone Fone stereo is covered with a sleeve made of Lycra Spandex–the same material used to make expensive swim suits, so it's easily washable. You simply remove the sleeve, dip it in soapy water, rinse and let the sleeve dry. It's just that easy. The entire system is also protected against damage from moisture and sweat making it ideal for jogging or bicycling.

The sleeve comes in brilliant Bone Fone blue–a color designed especially for the system. An optional set of four sleeves in orange, red, green and black is also available for $10. You can design your own sleeve using the pattern supplied free with the optional kit.

YOUR OWN SPACE

Several people could be in a car, each tuned to his own program or bring the Bone Fone to a ball game for the play by play. Cyclists, joggers, roller skaters, sports fans, golfers, housewives, executives–everybody can find a use for the Bone Fone. It's the perfect gift.

Why not order one on our free trial program and let your entire family try it out? Use it outdoors, while you drive, at ball games or while you golf, jog or walk the dog. But most important–compare the Bone Fone with your expensive home stereo system. Only then will you fully appreciate the major breakthrough this product represents.

GET ONE SOON

To order your Bone Fone, simply send your check or money order for **$69.95** plus $2.50 postage and handling to the address shown below. (Illinois residents add 5% sales tax.) Credit card buyers may call our toll-free number below. Add $10 if you wish to also receive the accessory pack of four additional sleeves.

We'll send you the entire Bone Fone stereo complete with four AA cell batteries, instructions, and 90-day limited warranty including our prompt service-by-mail address.

When you receive your unit, use it for two weeks. Take it with you to work, or wear it in your car. Take walks with it, ride your bicycle or roller skate with it. Let your friends try it out. If after our two-week free trial, you do not feel that the Bone Fone is the incredible stereo experience we've described, return it for a prompt and courteous refund, including your $2.50 postage and handling. You can't lose and you'll be the first to discover the greatest new space-age audio product of the year.

Discover the freedom, enjoyment, and quality of the first major breakthrough in portable entertainment since the transistor radio. Order a Bone Fone stereo at no obligation, today. *Pending FCC approval

JS&A® PRODUCTS THAT THINK®

Dept. SA One JS&A Plaza
Northbrook, Ill. 60062 (312) 564-7000
Call TOLL-FREE 800 323-6400
In Illinois Call ..... (312) 564-7000

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

NOVEMBER, 1929: "President Hoover, with characteristic directness and honesty, has demanded, and Congress is about to begin, an investigation into the activities of W. B. Shearer at the naval-limitation conference in Geneva. Shearer maintains he was employed by large American shipbuilders as a 'big navy' propagandist. The shipbuilders say he was employed as an 'observer.' Whatever the outcome of the investigation, it is obvious that much propagandizing was going on, and it is just as certain that it is still going on. It is hoped that the investigation not only will put to rout the lobbyists but also will indicate to Great Britain our desire to lay all our cards on the table. It is reported that Prime Minister MacDonald contemplates a trip to this country to talk over the matter with Mr. Hoover, but whether or not he comes he has already shown the willingness of the Labor government to meet us halfway in drawing up a naval-limitations agreement."

"The age we live in is the age of mechanical energy, but how long will the stores of energy last? We derive nearly all our energy from coal. Only a fraction comes or can come from water power, for there is not nearly enough water power in all the world to turn the wheels of industry. There is enough coal to last us for about 1,000 years at the present rate of consumption, and although nobody living need worry about the future source of energy when the coal is gone, the question is as fascinating to scientific people as any question one could raise. There is a special fascination in the problem of the earth's internal energy. Practical recovery of such power is already taking place in at least two localities, chief of which are Sonoma County in California and Tuscany in Italy. The California steam wells have not yet been exploited commercially but the Italian wells have. The natural steam of the Italian thermal region is employed to generate about 10,000 horsepower of electricity, which is transmitted to Florence and other cities over the power-line network of northern Italy."

"If 10 economists were asked to give the cause of the great bull market in securities that has raged almost unabated for the past five years, we would no doubt obtain various answers. The layman would be quick to say that it was due to prosperity. That would be a true but superficial answer. Any theory advanced, if at all scholarly, would have to include an analysis of the influence of research and invention upon the trend of security prices. The busy tickers of Wall Street today reflect not only good management and good prospects but also the possibilities of new products from research laboratories. The modern investor, if he is discreet and cautious in the purchase of securities, not only scrutinizes the board of directors but also extends his investigation to the research department of the corporation whose securities he anticipates buying. Today the Steinmetzes and the De Forests are as important in the management of corporate affairs as the Raskobs, the Du Ponts and the Fishers. The real wealth of the modern corporation and the potentialities for enhancing its earnings are created in the laboratory."

"A great deal of water has flowed under the Eads bridge since January 16, 1920, and a great deal of the money of Anheuser-Busch was dammed up at the source when the Eighteenth Amendment went into effect. The St. Louis showplace is a most conspicuous example of an industrial 'comeback' after an unforeseen disaster. Since prohibition this brewery concern, like the lesser ones, has found it rather hard sledding to keep the big establishment going. They went into a number of ventures that were flat failures, but now daylight shines, after a long night, on corn syrup, nonalcoholic cereal beverages, malt food tonic, glucose, ginger ale and particularly baker's yeast."

NOVEMBER, 1879: "The question of the transmission of power by means of electricity is not entirely a new one. Recently, however, new and important results have been attained. MM. Chrétien and Félix, two well-known French engineers, have used a combination of Gramme electric machines for unloading beet boats and loading wagons at the sugar works of Sermaize, in Marne. The governing Gramme machine occupies a fixed position in the sugar works and is actuated by a steam engine. The electricity thereby developed is led by conducting wires to another Gramme machine, which by means of a belt is made to turn a large wheel that carries an endless chain of buckets. The chain runs down into the boat, where six workmen load the buckets with the beets that are to be hoisted to the wagons. The use of this apparatus is found to effect a saving of about 40 per cent over manual labor. It is urged in favor of such apparatus that by its means much water power now neglected may be turned to profitable advantage in actuating the machines for producing and transmitting the electricity."

"In a recent lecture before the Lowell Institute, Professor John Trowbridge of Harvard University observed that physics is a term recently substituted for natural philosophy as being more comprehensive. He then spoke of the laws of physics and said that definite thinking on one subject leads naturally to definite thinking on another. The philosophy of physics enters upon all subjects. It is the investigation of the physical laws of the universe and is the result of the investigation of the truth by means of evidence. Conclusions are reached by processes of induction and deduction. The more nearly science approaches a deductive form, the more perfect it becomes. According to Professor Trowbridge, by the possession of good means of deduction a man might be saved half a lifetime spent in experimenting."

"The first of the great cotton mills of Lowell, Mass., commenced operations in 1823. Today, as a result of the great changes in machinery, from 60 to 64 per cent less labor is required for a given amount of product than was required in 1860. Three-fourths of all the labor in the mills today is done by women, and every year the work is more and more coming into their hands. In 1860 female operatives, working 11 hours a day (66 hours a week), received $3.26, whereas last year the female operatives, working 10 hours a day (60 hours a week), received $4.34. Today the operatives are mostly all foreigners, some English and French, but mainly Irish."

"The coming of an English gentleman, with a craze against vaccination as a preventive of smallpox, has been made the occasion of an attempt to stir up opposition to the practices of our American physicians and boards of health in this connection. By parading a portentous array of figures to show that vaccination does not prevent smallpox and does entail a vast amount of disease through blood contamination, not a little feeling is aroused, particularly among the ignorant. The anti-vaccination spirit prevailing in English and other European circles, embracing no inconsiderable body of the more intelligent classes, is urged as a reasonable ground for similar opposition here. In view of the fact that by the general adoption of correct vaccination smallpox, until recently one of the worst of human scourges, has been so thoroughly brought under subjection in New York that with 1,100,000 inhabitants there were last year but 14 cases of the disease, it is manifestly as unwise as it is absurd for our newspapers to lend themselves to propagation of the anti-vaccination nonsense."

Hardly your everyday liqueur.

Some things in life are too precious for every day. Lochan Ora is one. An imported golden liqueur with a unique taste all its own, it's to be lingered over—sipped slowly, savored fully. You'll be proud to share it—but not with just anybody.

Lochan Ora. Imported Liqueur.

70 PROOF • IMPORTED BY GENERAL WINE & SPIRITS CO., N.Y., N.Y.

A RABBIT DIESEL COULD GIVE YOU CHEAPER LONG DISTANCE RATES THAN YOUR TELEPHONE.

Actually based on fuel costs compared to direct dial weekday calls.

We'll expect to hear from the phone company about this ad. The question is, will they call us?

Or will they drive up in a Rabbit Diesel and try to save some money?

A Rabbit Diesel is so inexpensive to drive, it can take you from Boston to Hyannis on a mere $1.92 of fuel. A 10-minute phone call, on the other hand, will set you back about $3.09.

There are literally thousands of other examples, too.

Like Phoenix to Tucson: $2.95 in a Rabbit Diesel, $3.16 for the phone call.

Or Columbus to Cincinnati: $2.70 in a Rabbit Diesel, $2.81 for the phone call.

Aside from beating a telephone, a Rabbit Diesel also beats any other car you can buy. According to the 1980 EPA Guide, it gets the highest mileage in America: an incredible 40 estimated MPG and 52 MPG highway. (Use the "estimated MPG" for comparisons. Actual mileage varies with speed, weather and trip length. And your highway mileage will probably be less.)

No wonder the editors of Motor Trend Magazine said: "The Rabbit Diesel is a flat amazing car. It does many things so well, you're inclined to think that all autos should be built this way."

Why not call your Volkswagen dealer to arrange a test drive? It could be the last phone call you ever need to make.

VOLKSWAGEN DOES IT AGAIN

What kind of man owns his own computer?

Rather revolutionary, the whole idea of owning your own computer? Not if you're a diplomat, printer, scientist, inventor . . . or a kite designer, too. Today there's Apple Computer. It's designed to be a *personal* computer. To uncomplicate your life. And make you more effective.

It's a wise man who owns an Apple.

If your time means money, Apple can help you make more of it. In an age of specialists, the most successful specialists stay away from uncreative drudgery. That's where Apple comes in.

Apple is a real computer, right to the core. So just like big computers, it manages data, crunches numbers, keeps records, processes your information and prints reports. You concentrate on what you do best. And let Apple do the rest. Apple makes that easy with three programming languages—including Pascal—that let you be your own software expert.

Apple, the computer worth not waiting for.

Time waiting for access to your company's big mainframe is time wasted. What you need in your department—on *your* desk—is a computer that answers only to you . . . Apple Computer. It's less expensive than timesharing. More dependable than distributed processing. Far more flexible than centralized EDP. And, at less than $2500 (as shown), downright affordable.

Visit your local computer store.

You can join the personal computer revolution by visiting the Apple dealer in your neighborhood. We'll give you his name when you call our toll free number **(800) 538-9696**. In California, **(800) 662-9238**.

Apple Computer, 10260 Bandley Drive, Cupertino, CA 95014.

apple computer

THE AUTHORS

BERNARD T. FELD and KOSTA TSIPIS ("Land-based Intercontinental Ballistic Missiles") are physicists at the Massachusetts Institute of Technology who share a strong interest in the technical aspects of arms control. Feld is head of the Division of Nuclear and High-Energy Physics at M.I.T. and editor-in-chief of *The Bulletin of the Atomic Scientists.* Born in Brooklyn, he entered the College of the City of New York at age 15 and went on to do graduate work at Columbia University. He suspended his graduate studies temporarily to work under Leo Szilard and Enrico Fermi toward the first nuclear chain reaction. From 1942 to 1944 he was a group leader in the "Metallurgical Laboratory" at the University of Chicago, where the chain reaction was achieved. The following year he received his Ph.D. from Columbia. Feld joined the M.I.T. faculty in 1945 and was made full professor in 1957. In addition to his work in atomic and nuclear physics he has been involved in many studies of arms control and disarmament. Tsipis is associate director of the Program in Science and Technology for International Security in the M.I.T. department of physics. He did his undergraduate work at Rutgers University and went on to obtain his Ph.D. in high-energy physics from Columbia in 1966, joining the M.I.T. faculty that year. In 1973 Tsipis became a researcher at the Stockholm International Peace Research Institute (SIPRI) and a fellow of the M.I.T. Center for International Studies.

ABNER LOUIS NOTKINS ("The Causes of Diabetes") is chief of the Laboratory of Oral Medicine at the National Institute of Dental Research. He did his undergraduate work at Yale College, received his M.D. from the New York University School of Medicine in 1958 and did his internship and residency in internal medicine at the Johns Hopkins Hospital. On completing his medical training he joined the National Institutes of Health as a research associate at the National Cancer Institute. He moved to the Institute of Dental Research in 1973. Notkins' research has focused on the immunological processes involved in persistent and recurrent viral infections and more recently on the factors controlling the latency of the herpes simplex virus and the role of viruses in causing diabetes mellitus.

L. McDONALD SCHETKY ("Shape-Memory Alloys") is technical director for metallurgy of the International Copper Research Association in New York. He was educated at the Rensselaer Polytechnic Institute, where he obtained his bachelor's degree in chemical engineering, his master's in metallurgical engineering and his Ph.D. in physical metallurgy. From 1953 to 1956 he was on the research staff of the Massachusetts Institute of Technology and served as director of materials research at the M.I.T. Instrumentation Laboratory. In 1956 he founded the Alloyd Corporation, a research and manufacturing company specializing in materials research. In his spare time Schetky operates a vineyard in Rhode Island that should produce its first vintage this year.

MOTOO KIMURA ("The Neutral Theory of Molecular Evolution") is chairman of the department of population genetics at the National Institute of Genetics in Japan. He did his undergraduate work in botany at Kyoto University, and in 1949 he joined the research staff of the newly founded Institute of Genetics. In 1953 he came to the U.S. as a graduate student, first at Iowa State College and then at the University of Wisconsin, where he received his Ph.D. in 1956 in the laboratory of James F. Crow. He then returned to Japan to continue his research at the Institute of Genetics. In 1976 Kimura was awarded the Order of Culture by the Emperor, the highest cultural recognition in Japan. In his spare time Kimura breeds *Paphiopedilum* orchids.

DAVID L. MEIER and RASHID A. SUNYAEV ("Primeval Galaxies") are respectively American and Russian astronomers. Meier is a postdoctoral fellow in theoretical astrophysics at the California Institute of Technology. He obtained his bachelor's and master's degrees in physics at the University of Missouri at Rolla and his master's and Ph.D. degrees in astronomy from the University of Texas at Austin, the last in 1977. Meier recently completed a NATO Postdoctoral Fellowship at the Institute of Astronomy of the University of Cambridge. Sunyaev is head of the Elementary Processes in Astrophysics group at the Institute for Cosmic Research in Moscow. He was graduated from the Moscow Physical Technical Institute in 1966, received his Ph.D. in 1968 and was awarded the title of Doctor of Science in 1973, all in astrophysics. Sunyaev has done work in cosmology, X-ray astrophysics and plasma physics. Meier and Sunyaev have never met but were introduced by Beatrice M. Tinsley of Yale University through correspondence. They worked first with her on a scientific paper and then together on the present article, communicating only by mail and telephone.

BERND HEINRICH and GEORGE A. BARTHOLOMEW ("The Ecology

of the African Dung Beetle") study the behavior, physiology and ecology of insects. Heinrich is professor of entomology at the University of California at Berkeley. He did his undergraduate work in zoology at the University of Maine and obtained his Ph.D. from the University of California at Los Angeles in 1970, working in Bartholomew's laboratory. His work has focused on insect thermoregulation and energetics and the role of physiology and behavior in providing insights into the mechanisms of ecology and evolution. Heinrich recently published a book titled *Bumblebee Economics* (Harvard University Press). Bartholomew is professor of zoology at U.C.L.A. He received his bachelor's and master's degrees at Berkeley and his Ph.D. from Harvard University in 1947. He has been at U.C.L.A. ever since.

BERNARD D'ESPAGNAT ("The Quantum Theory and Reality") is professor of physics at the University of Paris and director of the Laboratory of Theoretical Physics and Elementary Particles at Orsay. The son of Georges d'Espagnat, a postimpressionist painter, he was educated at the École Polytechnique and at the Sorbonne, where he obtained his doctorate in physical sciences in 1950. He then worked for a year as a research assistant under Enrico Fermi at the University of Chicago, and in 1953 he received a fellowship at the Niels Bohr Institute in Copenhagen. The following year he was invited by Felix Bloch, who was then director of the European Organization for Nuclear Research (CERN), to create in Geneva the nucleus of what was to become the theoretical division of the organization. There he investigated the theory of strange particles and contributed to the method of applying group theory to the study of such objects. He joined the faculty of the Sorbonne in 1969. D'Espagnat has written three books on the conceptual foundations of quantum mechanics. The latest, titled *À la recherche du réel* (*In Search of Reality*) is soon to be published in French.

P. W. DIXON ("A Neolithic and Iron Age Site on a Hilltop in Southern England") is lecturer in medieval archaeology at the University of Nottingham and honorary curator of the University Museum. He studied classics at New College, Oxford, and went on to receive his Ph.D. in archaeology with a thesis on late-medieval fortified houses and society in northern England and Scotland. He has directed the excavations described in his article since their beginning in 1969 and has also participated in excavations at the Tudor palaces of Greenwich and Richmond and at several medieval houses. His other interests include playing traditional English musical instruments, particularly the Northumbrian pipes.

With scores of electronic games on the market, selecting the one for you shouldn't become a guessing game. Choosing the best game can be easier when you know more about us.

THESE GAMES THINK.

We're Chafitz, the company that has become the leader in developing games that do more than buzz, sing or blink lights at you. Our games are designed to think and play as if they were human. Boris, our chess computer, has already become the standard throughout the world. Our new chess computer game, Sargon 2.5, is so powerful that it defeated a 5 million dollar chess computing Goliath.

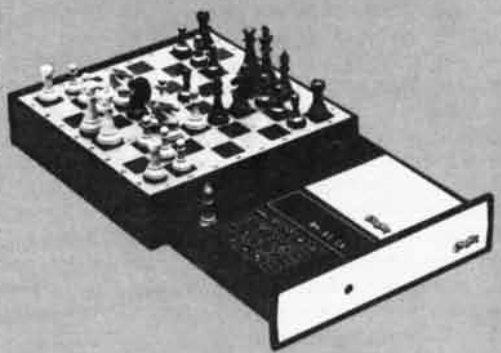
Chafitz's games are designed to provide challenge and excitement for all skill levels, beginners as well as Masters. Some of our games even talk to their human opponent with messages that advise you of a blunder or caution you of impending defeat.

Chafitz is committed to producing the most advanced games possible and that is why we're constantly pioneering new programming breakthroughs.

A program developed for our backgammon game, Aristotle, was pitted against the current world backgammon champion. Aristotle clearly demonstrated to the world that Chafitz is Number One in computer games by defeating the champion 7 to 1 in a 7 point match. This is the first time ever that a world champion was defeated by a computer.

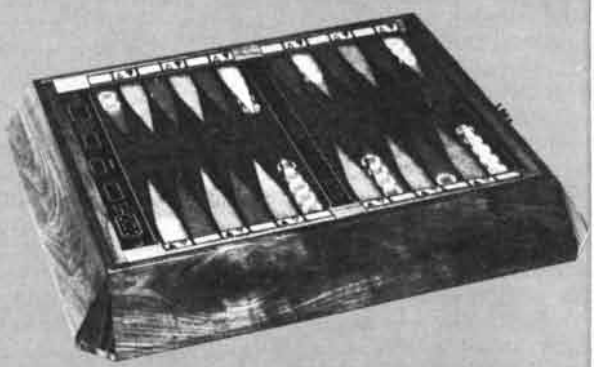
When you're ready to buy an electronic game, remember our name, Chafitz. We're the people who make the games that think.

AUTO-RESPONSE CHESS BOARD. The ultimate in computer chess. Moves and responses are electronically indicated when player moves pieces on the board. Never a button to press.



MODULAR GAME SYSTEM. An exclusive Chafitz concept designed to never become outdated. Interchangeable modules allow continuous updating as new programs are developed. Modules will soon be available for a variety of board games. Compact, battery operated and portable.

BORIS DIPLOMAT. The world's first battery operated portable chess computer. Designed for all levels of play. It is compact enough to slip into a briefcase.



ARISTOTLE. This is the most powerful computer game ever developed for consumer use. Has the features and flexibility to provide true world class competition.

SEE THESE CHESS AND BACKGAMMON GAMES at major retailers and specialty shops worldwide. For further information and the name of the store near you, call or write: Chafitz, Inc., Dept. 778, 856 Rockville Pike, Rockville, Md. 20852. (301) 340-0200. In Canada call (416) 683-4555.

Chafitz

© 1979

MATHEMATICAL GAMES

The random number omega bids fair to hold the mysteries of the universe

by Martin Gardner

A fascinating unpublished paper titled "On Random and Hard-to-Describe Numbers" has been sent to this department by Charles H. Bennett, a mathematical physicist at the Thomas J. Watson Research Center of the International Business Machines Corporation. It begins by recalling certain paradoxes involving integers, in particular a paradox that seems to stand in the way of calling any specific integer a random one. The resolution of the paradox makes randomness a property that most integers have but cannot be proved to have, a subject that was treated in these pages by Gregory J. Chaitin, also of the Watson Research Center [see "Randomness and Mathematical Proof," by Gregory J. Chaitin; SCIENTIFIC AMERICAN, May, 1975]. Bennett goes on to consider the senses in which irrational numbers such as pi can be said to be random. And he discusses at length the irrational number Ω, recently discovered by Chaitin, which is so random that in the long run no gambling scheme for placing bets on its successive digits could do better than break even.

The number Ω has other unusual properties. To begin with, it can be defined precisely but it cannot be computed. Most remarkable, if the first few thousand digits of the number were known, they would, at least in principle, provide a way to answer most of the interesting open questions in mathematics, in particular those propositions that if false could be refuted in a finite number of steps. The following discussion of these issues is taken from Bennett's paper, which begins with a simple variant of Berry's paradox. Named after G. G. Berry, the Oxford librarian who discovered it, the paradox was first published by Bertrand Russell and Alfred North Whitehead in *Principia Mathematica*. Bennett describes it as follows.

"The number 'one million, one hundred one thousand, one hundred twenty one' is unusual in that it is, or appears to be, the number named by the expression: 'The first number not nameable in under ten words.' This expression has only nine words, however, and so there is an inconsistency in regarding it as a name for 1,101,121 or any other number. Berry's paradox shows that the concept of nameability is inherently ambiguous and too powerful to be used without restriction. In a note in *The American Mathematical Monthly* [Vol. 52, No. 4, page 211; April, 1945] Edwin F. Beckenbach pointed out that a similar paradox arises when one attempts to classify numbers as either interesting or dull: There can be no dull numbers, because if there were, the first of them would be interesting on account of its dullness.

"Berry's paradox can be avoided and even tamed, however, if the definition of nameability is restricted, that is, if an integer is considered to be named when it has been calculated as the output of a computer program. To standardize the notion of computation in this definition a simple idealized computer known as a universal Turing machine is introduced. [For a discussion of Turing machines see this department for June, 1971.] This machine will accept a program in the form of a sequence of 0's and 1's on an input tape and write the results of its computation, again in the form of a sequence of binary digits, on an output tape at the end of the computation. A third tape is used during the computation to store intermediate results. (The use of separate tapes for input, output and memory is a conceptual convenience rather than a necessity; in the earliest Turing machines the same tape served all three purposes.) As is well known, a universal Turing machine can do anything the most powerful digital computer can do, although considerably slower. More generally, it can perform even the most complicated manipulation of numerical or symbolic information as long as the manipulation can be expressed as a finite sequence of simple steps in which each step follows from the preceding one in a purely mechanical manner without the intervention of judgment or chance.

"An integer x can now be named by specifying a binary sequence p that, when it is given to the Turing machine as input, causes the machine to calculate x as its sole output and then halt. There can be no doubt that the program p does indeed describe x. Hence the universal Turing machine provides an unambiguous but flexible language in which a number can be described according to any of the ways it might be effectively calculated. (For example, the number 523 might be described as the 99th prime number, as $(13 \times 41) - 10$ or, more directly, as the binary sequence 1000001011.) Every integer is nameable in this language, because even an integer with no distinguishing properties can always be described by simply giving its binary sequence.

"Returning to the question of interesting and dull numbers, an interesting number can now be defined, without paradox, as one computable by a program with considerably fewer bits, or binary digits, than the number itself. This short description would reflect some special feature by which the number could be distinguished from the general run of numbers. By this definition, then, $2^{65,536} + 1$, the first million digits of pi and (17!)! are interesting numbers. [The exclamation point is the factorial sign: $n!$ equals $1 \times 2 \times 3 \times \ldots n$.] Conversely, a dull, or random, number is one that cannot be significantly compressed, that is, one whose shortest description has about as many bits as the number itself. This algorithmic definition of randomness as incompressibility, which is reviewed in the article by Chaitin mentioned above, was developed in the 1960's by several mathematicians, including Ray J. Solomonoff and Chaitin in the U.S. and A. N. Kolmogorov in the U.S.S.R. Just as most integers are random according to the intuitive meaning of the word, most integers are incompressible or nearly so, because there are far too few short programs to go around. In other words, even if no programs were wasted (say by computing the same result as other programs), only a small fraction of the n-bit integers could be provided with programs even a few bits shorter than themselves.

"Using this definition of randomness, Chaitin demonstrated the surprising fact that although most integers are random, only finitely many of them can be proved random within a given consistent axiomatic system. A form of Gödel's famous incompleteness theorem, this result implies in particular that in a system whose axioms and rules of inference can be described in n bits it is not possible to prove the randomness of any integer much longer than n bits. Chaitin's proof of this assertion makes use of a computerized version of the Berry paradox: Suppose in a proof system describable in a small number of bits the randomness of some integer with a large number of bits can be proved. One could then design a small Turing-machine program based on the proof system that would yield the large random integer as

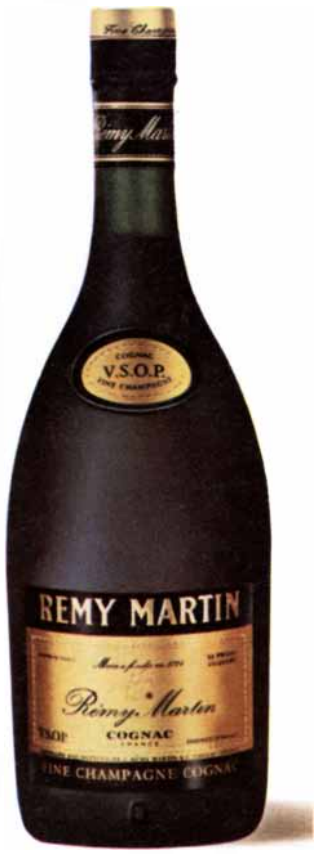
Symbols of the world's most precious elements

^{1}Au

^{2}Ag

^{3}Pt

^{4}Rm



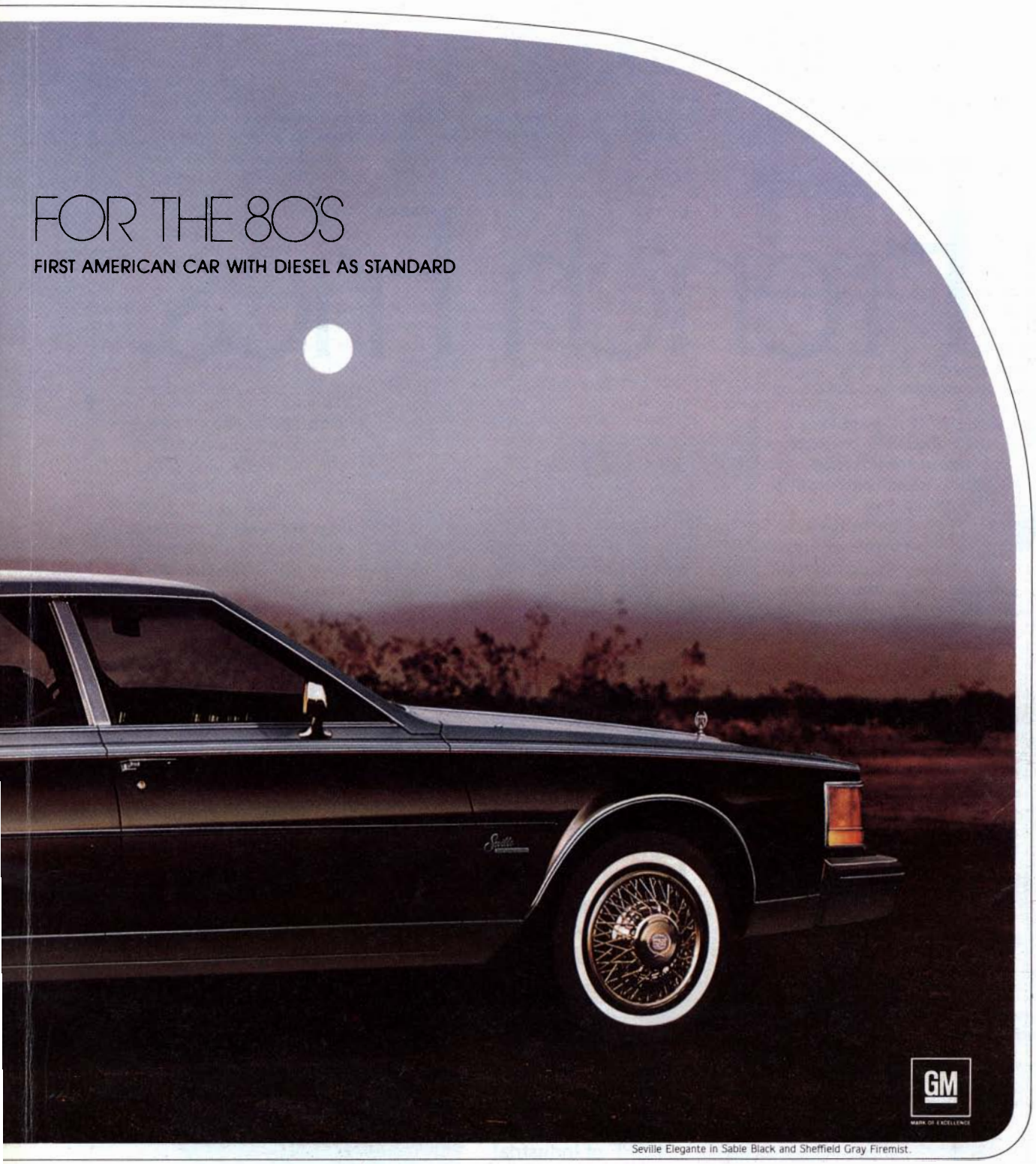
1. Gold. 2. Silver. 3. Platinum. 4. Remy Martin V.S.O.P.
Fine Champagne Cognac

80 Proof Renfield Importers, Ltd., New York

SO DISTINCTIVE OF LINE, SO ADVANCED IN TECHNOLOGY.

Here from Cadillac is one of the most advanced production automobiles ever built anywhere. And with its daring styling, one of the most distinctive.

As the first American car with a Diesel engine as standard, Seville for the 80's blends the classic beauty of yesterday with the technology of tomorrow. Almost every feature you migh desire in a world-class automobile is standard. From front wheel drive for traction and increased roominess to fou wheel independent suspension. From four-wheel disc brakes to electronic level control.



Seville Elegante in Sable Black and Sheffield Gray Firemist.

IT SETS A NEW AMERICAN STANDARD FOR THE WORLD.

Result: Seville for the 80's has considerably more roominess than its predecessor. And considerably more usable trunk space.

Only Seville offers this unprecedented choice in engines: a standard 5.7 liter Diesel V8 or Cadillac's gas engine with new Digital Electronic Fuel Injection and on-board computer diagnostics to help simplify servicing. In California: EFI gas V8 is available ... Diesel engine not available at time of printing. Sevilles are equipped with GM-built engines produced by various divisions. See your Cadillac dealer for details. And to buy or lease a new American standard for the world.

Introducing: Solar French Fries.

American fingers pop some 260,000 pounds of frozen french fries into American mouths every day.

Before they're frozen, all those fries have to be cooked in boiling oil.

Idaho Potatoes. Oregon Sunshine And A TRW System.

Keeping that oil hot uses a lot of energy, so a company called TRW and the Department of Energy decided to put the sun to work—french frying potatoes at a potato processing plant in Oregon.

Now with the help of a TRW-designed solar energy system, this plant is cooking up the country's first solar french fries.

Perched on the roof of the plant are solar collectors. Their job is to soak up enough Oregon sunshine to generate 425°F of industrial steam—which is what it takes to heat the 900 gallons of cooking oil needed to fry 20,000 pounds of potatoes every hour.

TRW engineers had to design a system that uses water as the heat transfer liquid, yet still provides the high efficiency the process must have.

What we came up with is a solar steam-generating system that's both efficient and safe for food processing.

Everybody always knew the sun could make the sidewalks hot enough to fry eggs on.

Now it's french frying potatoes—by the ton!

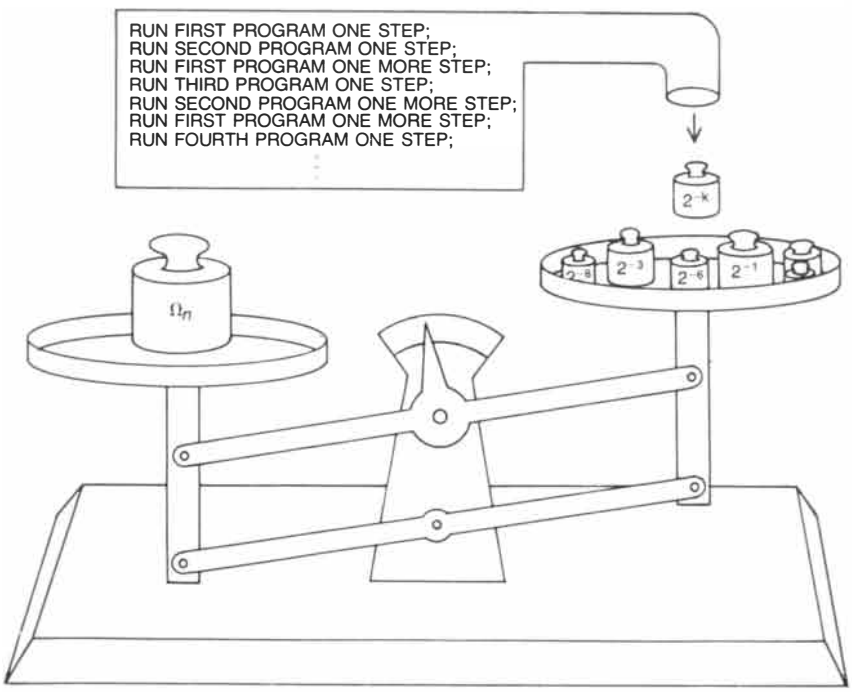
A COMPANY CALLED TRW

output. If the large integer were truly random, however, it could not be the output of any small program, and so a contradiction has been reached.

"More precisely, the Turing-machine program would incorporate a fixed supervisory routine, or subprogram, with a length of, say, c bits. This routine would utilize the n bits of axioms and rules of inference to begin systematically generating all possible proofs that could be derived from the axioms in order of increasing number of deductive steps: first all one-step proofs, then all two-step proofs and so on. After each proof was generated the routine would check to see if it was a proof that some integer of considerably more than $n + c$ bits is random. If such a proof were found, the supervisory routine would print out the large random integer specified in the proof and then call a halt to the entire computation. The total length of the Turing-machine program (the supervisory routine plus the axioms and the rules of inference) would be $n + c$ bits, however. In other words, a program $n + c$ bits long would have generated as its output a specific integer that by the algorithmic definition of randomness could not be produced by any program as small as $n + c$ bits. The only escape from this contradiction is to conclude either that the axiomatic system is inconsistent (that is, untrue statements can be proved within it) or that the systematic generation of proofs must continue indefinitely, without uncovering a proof of randomness for any integer much larger than $n + c$ bits. The original Berry paradox appeared as a nuisance, casting doubt on the seemingly meaningful notion of a random integer. The computerized Berry paradox surrounds this notion with a necessary hedge of unprovability, allowing it to be defined in a noncontradictory manner.

"Long before the notion of a random integer was taken seriously Émile Borel, Richard von Mises and other mathematicians sought to define and find examples of random real numbers or, equivalently, random infinite sequences of decimal or binary digits. It has been conjectured that irrational numbers such as pi, e and $\sqrt{2}$, which occur naturally in mathematics, are random in the sense that each digit, and indeed each block of digits of fixed length, appears with equal frequency in their decimal expansion. Any sequence of digits with this property is said to be normal. It is not hard to show that no rational number is normal no matter what base it is expressed in, and that almost all the irrational numbers must be normal in every base. So far none of the individual classic irrational numbers has been proved normal, however, although statistical evidence generally supports the conjecture that they are.

"On the other hand, it is easy to construct 'artificial' irrational numbers that can be proved normal in spite of the fact that their digits follow a trivial and transparent pattern. The most famous of these numbers was invented by D. G. Champernowne in the early 1930's: 0.12345678910111213141516171819-2021222324.... This preposterous number, which consists of the decimal integers in increasing order, has been proved not only irrational and normal (in base 10) but also transcendental.



Using the first n *bits of* Ω *to solve the halting problem for all programs of* n *bits or fewer*

(The transcendental numbers are those numbers that are not the roots of ordinary algebraic equations.) In the initial portions of Champernowne's number there are significant departures from normality, but the differences in frequency tend toward zero as the number of digits being considered is increased. It is apparently not known whether the number is normal in all other bases.

"Although the sequence of digits in the decimal expansion of pi may be random in the sense of being normal, the sequence is not unpredictable. In other words, a good gambler, betting on the successive digits of pi, would eventually infer the rule for calculating the number and thereafter win every bet. The same is true of Champernowne's number. Is there a sequence so random that no computable gambling strategy, betting on each successive digit at fair odds, could do better than break even? Any number that is random in this strong sense would necessarily be normal in every base. It is a fundamental result of probability theory that in fact almost all real numbers are random in this way, but finding a specific instance of such a number is not easy. Moreover, there is a sense in which no specifically definable real number can be random, since there are uncountably many real numbers (that is, the set of real numbers is too large to be matched up one-for-one with the positive integers) but only countably many definitions. In other words, the mere fact that a real number is definable makes it atypical. In this case, however, the problem is only to find a number that cannot be shown to be atypical by constructive, or computational, means. In particular the number must not be computable from its definition, because if it were, a perfect betting strategy could be devised.

"Chaitin's irrational number Ω is, among its other remarkable properties, random in this strong sense. To understand why that is true, however, it is necessary to deal briefly with the classic unsolvable problem of computability theory known as the halting problem: the task of distinguishing computer programs that come to a spontaneous halt from those that run on indefinitely. Leaving aside gross programming errors, which can cause a program to halt or not to halt for trivial reasons, a program halts if it succeeds in doing what it set out to do, when it has computed, say, the 99th prime number or the first million digits of pi. Conversely, a program will run on indefinitely if the task it is embarked on is an unending one, such as computing all prime numbers or searching for a planar map that cannot be colored with only four colors so that no two adjacent regions are the same color.

"At first the halting problem might seem solvable. After all, the fact that a program halts can always be demonstrated by simply running the program

Cordless Wonder

For $89.95 the Mura cordless telephone sounds like a bargain. But wait until you hear about its many disadvantages.

The Mura cordless telephone represents a major breakthrough in telephone technology.

It's about time. For years you've seen ads for cordless telephones selling for between three and four hundred dollars.

Now through some very clever planning and a sprinkle of new technology, Mura Corporation has come up with a cordless telephone that sells for $89.95. However, it has major disadvantages that could totally discourage you from buying the system—but more on that later.

ONLY IN AMERICA

The Mura weighs only 12 ounces and measures 1½"x 2¾"x 6½". The system includes a base unit that plugs into your telephone jack. You carry your cordless telephone with you and when your phone rings, you press a button and answer. And you can talk to anyone as long as you remain within 400 feet of the base unit.

But wait. We mentioned that the phone had major disadvantages. And it does. But first, let's outline some of its major advantages.

Convenience You don't need an extension telephone. With the Pocket Phone you have an extension phone that you can take with you—in the bath, in the den, in the garden, or to your neighbors.

Intercom You can use the base unit to page the person holding the cordless telephone. For example, if you're in your office and someone outside has the unit, you can press a button on the base unit and buzz the portable phone—just like on an intercom. Simply by talking on the phone plugged into your base unit, you can talk with someone on the remote phone. It's ideal for home or factory use.

Price The cost of the Mura remote telephone is only $89.95. Compare this price not only with the cost of other $300 remote telephones but with conventional phones as well, and you can appreciate what a major breakthrough the Mura system represents. But there's more.

You can plug any conventional phone into the base unit and carry on a three-way conversation. You can answer a call at the base unit and signal the remote unit to pick up the line. You can cut out the remote phone from the base unit if you want to keep a conversation private.

TALK OF VALUE

You can carry the cordless telephone with you with its antenna collapsed and the battery on standby. When a call beeps your unit, you simply extend the antenna, turn the power on, and start to talk.

The unit is FCC approved for connection directly into your telephone line. If you don't have a four-pronged jack or a modular connector, simply call your telephone company. They'll promptly install a jack for you and the cost will be around $15 or less depending on your location.

NOW THE CATCH

We mentioned that there was a catch—a few major disadvantages that you, as the consumer, should know about before you consider purchasing this product. Here they are:

Forget About Dialing The new Mura Pocket Phone can't dial out. It only receives calls. To many people, this doesn't matter because 90% of remote phones are used to receive calls and not to place them. By eliminating the dial, Mura has cleverly saved consumers hundreds of dollars.

Forget About Steel Walls The Mura unit won't penetrate them. This means that if you want to use your phone in a factory with metal walls, your unit won't work. But for most factories and practically all homes, the unit is ideal.

Forget About Snooping The unit has only a 400 foot range. At first this might seem awfully short, but nobody can snoop in on your conversations if that person is beyond this range, and 400 feet is more than enough for most applications. Most cordless telephones operate in the 27 megahertz range—the same frequency area used for citizen band radios.

The base unit for the Mura can also be used as a personal paging system or intercom.

The Mura uses the 49 megahertz range. This frequency has clearer reception with practically no interference.

The above are the disadvantages. For 90% of you, they don't mean a thing. For those 10% of you who need a dial, we would recommend the more expensive cordless telephones.

But for those of you who will accept its disadvantages, you'll be in store for the greatest idea in telephone convenience since the cordless telephone was first introduced. In fact, rather than install an extension phone, why not consider the Mura instead?

TRY IT FIRST

We suggest you try the Mura Cordless telephone system in your own home, office or factory. Use it for 30 days. Take the phone to your next door neighbor's house or with you to the bathroom while you take a shower or bath. Take it with you on your patio or balcony, or bring it in your garden as you work. Use it in your factory as an intercom or in your office as a remote telephone.

After you've given it a thorough test, then decide if you want to keep it. If not, no problem. Simply return your system for a prompt and courteous refund including your $3.50 postage and handling. You can't lose.

HERE'S THE WAY

To order your unit for a 30-day test, simply send your check for **$89.95** plus $3.50 postage and handling to JS&A Group, Inc., One JS&A Plaza, Northbrook, Illinois 60062. (Illinois residents please add 5% sales tax.) Credit card buyers, call our toll-free number below. We'll send your base unit, cordless telephone, rechargeable batteries, recharger, complete instructions, our 90-day limited warranty, and the address of the closest Mura Service Center or service-by-mail station.

Your unit is backed by Mura Corporation, a 17-year old company famous for their microphones, headsets, and other audio products. JS&A is America's largest single source of space-age products—further assurance that your modest investment is well-protected.

Very often when a product's disadvantages aren't made clear to the consumer, that product ends up being a disappointment. By explaining the major disadvantages of the Mura cordless telephone, not only are we avoiding a possible disappointment, we're proving just how great a product it really is. Order a Mura cordless telephone at no obligation today.

JS&A PRODUCTS THAT THINK®

Dept. SA One JS&A Plaza
Northbrook, Ill. 60062 (312) 564-7000
Call TOLL-FREE 800 323-6400
In Illinois Call ..... (312) 564-7000

long enough. Moreover, there are many programs that can easily be proved to halt or not to halt without ever running them. (For example, the famous four-color theorem, which states that five colors are never required for the map-coloring task mentioned above, was finally proved in 1976, and that proof guarantees that the map-coloring program will not halt.) The difficulty comes, then, not in solving particular cases of the halting problem but in solving the problem in general. A. M. Turing, the British mathematician who invented the Turing machine, showed that there is no general prescription for deciding how long to run an arbitrary program so that its halting or nonhalting will be revealed. He also showed that there is no consistent system of axioms strong enough to decide the halting of all programs without running them. The unsolvability of the halting problem can be derived from and indeed is equivalent to the more recent result that randomness, in the Chaitin-Kolmogorov sense of incompressibility, is a property that most integers have but cannot be proved to have.

"Now imagine that the Turing machine, instead of being given a definite program at the beginning of its computation, is fed a random sequence of bits. That can be accomplished by flipping a coin whenever the machine requests another bit from its input tape and feeding in 1 or 0, depending on whether the coin comes up heads or tails. This procedure raises a curious question: When the procedure is begun, what is the probability that the machine will eventually halt?

"The answer is Chaitin's number Ω. Because the value of Ω depends on which universal Turing machine is being used, it is not a single universal constant like pi. For a given machine, however, Ω is a well-defined irrational number between zero and one, with a natural interpretation as that machine's halting probability on a random program. A randomly chosen program is very likely to tell the computer to do something impossible or pointless, so that the machine either stops immediately in an error state or loops endlessly through a small sequence of instructions. For most computers the former behavior predominates, and so because the halting probability is close to 1 the decimal expansion of Ω begins with several consecutive 9's. It can be shown, however, that the digit sequence of Ω soon becomes quite patternless, ultimately defeating any computable gambling strategy, as well as being random in the Chaitin-Kolmogorov sense of incompressibility.

"The most remarkable property of Ω is not its randomness, however. After all, it shares that property with the great majority of real numbers. Rather it is the fact that if the first few thousand digits of Ω were known, they would, at least in principle, suffice to decide most of the interesting open questions in mathematics. This property, as well as Ω's immunity to gambling, is due to the compact way Ω encodes solutions to the halting problem.

"The most famous unsolved problem in mathematics is probably Fermat's 'last theorem,' which states that the equation $x^n + y^n = z^n$ has no solution in positive integers when n exceeds 2. Pierre de Fermat made this assertion in a handwritten note in the margin of a book on number theory, adding that he had discovered a truly remarkable proof of it that the margin was not large enough to hold. Fermat died without exhibiting his proof, and three centuries of effort by other mathematicians have yielded neither a proof nor a refutation.

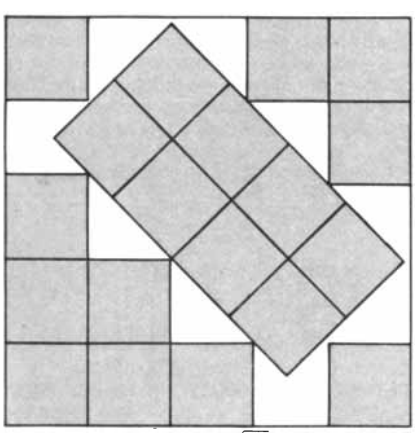
"Fermat's last theorem, like most of the famous unproved conjectures in mathematics, is an assertion of nonexistence and therefore could be refuted by a single finite counterexample, namely a set of integers x, y, z and n that solve the equation. Such finitely refutable conjectures are equivalent to the assertion that some computer program that searches systematically for the allegedly nonexistent object will never halt. Another famous finitely refutable proposition is Goldbach's conjecture, which asserts that every even number is the sum of two primes.

"Another kind of finitely refutable conjecture that has had an important place in the history of mathematics is the assertion that some proposition is independent of a given set of axioms, that is, the proposition can be neither proved nor refuted. The most famous propositions of this type are the parallel postulate, which states that through a given point there is exactly one line parallel to a given line, and the continuum hypothesis, which states that there is no infinite number between aleph-null (the number of positive integers) and aleph-one (the number of real numbers). In the 19th century the parallel postulate was shown to be independent of the other axioms of Euclidean geometry, and in this century the continuum hypothesis was shown to be independent of the axioms of set theory. The independence of a proposition P from a given set of axioms is equivalent to the nonhalting of a program that systematically generates proofs from the axioms, searching for a proof or refutation of P.

"Not all famous conjectures are finitely refutable. For example, no finite amount of direct evidence can decide whether pi is normal, whether there are infinitely many twin primes (consecutive odd primes such as 11 and 13 or 857 and 859) or whether the $P \neq NP$ conjecture in complexity theory is true. [The $P \neq NP$ conjecture asserts that there are mathematical problems for which the validity of guessed solutions can be verified quickly but for which solutions cannot be found quickly.] Such conjectures are not equivalent to halting problems, but there is good reason to believe most of them could be settled indirectly, by proving stronger, finitely refutable conjectures. For example, many twin primes are known, and empirical evidence indicates that the spacing between them grows rather slowly. Therefore the twin-prime conjecture may be viewed as an unnecessarily weak form of a stronger and still probably true assertion about the spacing of twin primes, say that there is always at least one pair of twin primes between 10^n and 10^{n+1}. This stronger assertion is equivalent to the nonhalting of a program that looks for an excessively large gap (greater than a factor of 10) in the distribution of twin primes. (It is important to note that some mathematical questions cannot be reduced to halting problems, for example, some questions about Ω itself. These irreducible questions tend, however, to be rather artificial and self-referent.)

"Interesting conjectures, like interesting numbers, tend to be concisely describable. It is hard to imagine a mathematically interesting, finitely refutable conjecture so verbose that it could not be encoded in the halting problem for a small program, one a few thousand or tens of thousands of bits long. Thus the answers to all interesting conjectures of this kind, including those that have yet to be formulated, would in principle be available if there were some kind of 'oracle' capable of solving the halting problem for all programs shorter than a few thousand bits. The number of programs involved would still be enormous; for example, there are about $2^{1,000}$ programs shorter than 1,000 bits. Hence it would seem that any oracle able to answer all these questions correctly would have to either be very smart or possess an enormous amount of stored information. In fact, because of the compact way in which Ω encodes the halting problem, its first few thousand bits serve as just such an oracle.

"How can solutions to specific halting problems be recovered from Ω? Since Ω is defined as the overall halting prob-



$n = 19,\ k = 7/2 + \sqrt{2} = 4.914+$

Solution to last month's packing problem

Every Panasonic gives you a picture that's life-like. This Panasonic also gives you a picture that's life-size.

Panasonic brings you life in your choice of sizes. From our little 7" (meas diag) portable color TV with its brilliant color and realism right up to our biggest—the Panasonic Color Video Projection System—a super-real 5 ft. (meas diag) of picture so life-like you can't help feeling you're part of it!

Panasonic Projection TV, with a picture area nearly 10 times that of an ordinary 19" diagonal TV, transforms your living room into a screening room. A sports arena. A concert hall. You can even connect one of our Omnivision VI™ VHS™ six hour video recorders to it and replay great films. Favorite sports events. Or your own video tape masterpieces, wherever you made them!

Yet the Panasonic Projection TV folds into a compact one-piece unit that's just 27 inches deep. And it comes with Panasonic remote control tuning, so you can relax in your favorite chair for an evening of brilliant television viewing.

To keep the viewing brilliant, this Projection TV, like many other Panasonic TV's, has two electronic color control systems. VIR, which "reads" color information transmitted by the TV station. And, when a VIR signal is not available, the Panasonic ColorPilot™ system takes over to maintain accurate, true-to-life color from program to program, from channel to channel—automatically.

Whether you choose a 7", 10", 12", 13", 25" or 5-ft. Panasonic Color TV (all meas diag) you'll get a picture so life-like you'll feel like you're part of it. If you choose the 5-foot Projection TV, you'll get a picture so life-size, you may begin to feel it's part of you! Cabinetry is simulated woodgrain. TV picture simulated.

Panasonic®
just slightly ahead of our time.

Money is not our most valuable asset.

When your problem is more than just a question of money, come to Deutsche Bank, where precision and attention to detail are qualities that guarantee perfection in all money matters.

We act as a universal bank, worldwide, whether your problem is something as normal as arranging credits for import and export financing or as sophisticated as the management of international bond issues.

Come to Deutsche Bank, where precision is not only a long tradition. It's our most valuable asset.

Deutsche Bank
A century of universal banking

Telex TRT 177747.
Tokyo; Representative Offices: Beirut, Bogotá, Caracas,
Subsidiaries Hong Kong, London, Luxembourg, Singapore.

GRAND AWARD. THE ONLY 12 YEAR OLD CANADIAN WHISKY.

12 years. That's how long this Canadian whisky takes.

12 years to allow a blend of the finest whiskies to marry perfectly. To achieve an unrivaled smoothness. And an absolutely remarkable richness of taste.

Grand Award. No other Canadian whisky tastes like it.

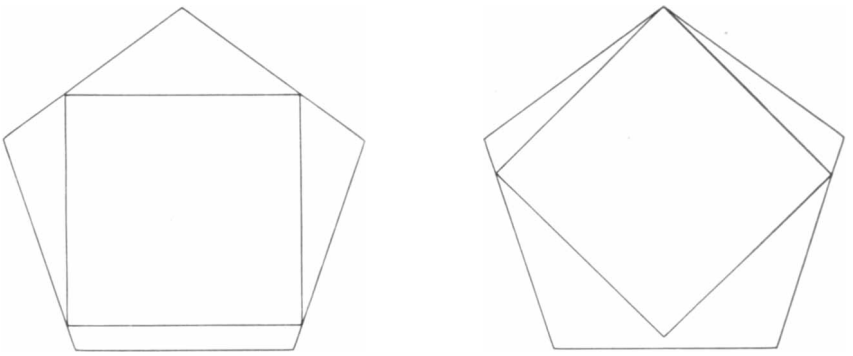
No other is presented at 90.4 proof.

And in the tradition of items of rare and exceptional quality, *only a limited amount of Grand Award is available each year.*

ability of a computer with random input, it can be regarded as the sum of the probabilities of all halting computations. Each halting program contributes to the sum its own probability of being chosen (by accident, as it were) when the input bits are supplied by coin tossing. The probability of generating any particular k-bit program in k coin tosses is $1/2^k$. Therefore if feeding this program to the Turing machine one bit at a time causes it to embark on a halting computation in which all k bits of the program but no more are actually requested and read, then that program's contribution to Ω is $1/2^k$. (Programs that call for the machine to read more or fewer bits than there are in the program are considered not to halt, ensuring that each halting computation's contribution to Ω is counted only once.)

"The illustration on page 25 shows how the first n bits of Ω can be used to solve the halting problem for all programs of n bits or fewer in length. Because Ω is an irrational number the number that consists of its first n bits, Ω_n, slightly underestimates its true value: Ω_n is less than Ω, which is less than $\Omega_n + 1/2^n$. In order to solve the halting problem for all n-bit programs, one begins a systematic but unending search for all the programs that halt, of whatever length, running first one program and then another for increasingly long times until enough halting programs have been found to account for more than Ω_n in total halting probability.

"One way to visualize this process is to consider a balance with a weight equal to Ω_n in its left pan. As is shown in the illustration the programs are run in a way that is reminiscent of the song "The Twelve Days of Christmas": first one step of the first program is run, then one step of the second program and another step of the first, then one step of the third program, another step of the second and another step of the first, and so on. Every time a program of length k is found to halt, a weight of $1/2^k$ is dropped into the right pan of the balance, because $1/2^k$ is the probability of that program being chosen and executed by a computer whose input bits are supplied by coin tossing. Eventually the balance must tip to the right, since the total weight of the programs that halt—the halting probability of Ω—is an irrational number between Ω_n and $\Omega_n + 1/2^n$. By this time a great many programs will have been found to halt, some longer than n bits and some shorter, and many programs that may halt later will not have done so yet. After the balance has tipped, however, no more programs of length n or less can halt, because if one did, that would raise Ω above its established upper bound of $\Omega_n + 1/2^n$. In other words, the subsequent halting of a program of n bits or fewer would alter one of the known digits of Ω.



Solution to last month's pentagon problem

"If this gargantuan computation were carried out with a sufficiently precise estimate of Ω, say the first 5,000 bits, then among the programs whose fate would be decided would be one whose nonhalting would verify Fermat's last theorem, and programs that would decide Goldbach's conjecture and all other simply stated, finitely refutable conjectures. In addition programs would be included whose nonhalting would almost certainly settle many conjectures that are not finitely refutable, such as those about the normality of pi, the twin primes and the $P \neq NP$ question, by proving stronger, finitely refutable statements.

"Returning to the senses in which Ω itself is random—its incompressibility and the impossibility of successfully gambling on its digits—it may seem strange that Ω can contain so much information about the halting problem and yet be computationally indistinguishable from a meaningless random sequence. Actually Ω is a totally informative message, one that appears to be random because all redundancy has been squeezed out of it, a message consisting only of information that can be obtained no other way.

"To put Ω's lack of redundancy in perspective, consider a more traditional way of encoding the halting problem in an uncomputable irrational number: let K be defined as the real number whose nth bit is 1 or 0, depending on whether or not the nth program halts. K is indeed often referred to as an oracle for the halting problem, but it is a very dilute oracle in the sense that the first 2^n bits of K contain about the same information as the first n bits of Ω: enough to solve the halting problem for all programs of n bits or fewer. K is susceptible to gambling precisely because it is dilute. For example, a sizable portion of all programs can easily be proved to halt or not to halt for trivial reasons. The corresponding bits of K are predictable, and by betting only on those bits a gambler could win consistently. Moreover, even the unpredictable bits of K are not totally unpredictable. Often two programs can be shown to be attacking the same nontrivial problem in different ways. The halting of one program will then decide the halting of the other, and a gambler who "passed" on the first program, not knowing which way to bet, could bet confidently on the second.

"The fact that Ω is incompressible and immune to gambling follows from its compact encoding of the halting problem. Because the first n bits of Ω solve the halting problem for all programs of n bits or fewer, they constitute an 'axiom' sufficient to prove the incompressibility of all incompressible integers of n bits or fewer. If Ω_n could be computed by a program significantly shorter than n bits, then a program of similar size would suffice to find and print out the first incompressible n-bit integer, which is a contradiction. In other words, since Ω_n provides enough information to compute a specific n-bit incompressible integer, it must be incompressible itself.

"Throughout history mystics and philosophers have sought a compact key to universal wisdom, a finite formula or text that would provide the answer to every question. The use of the Bible, the Koran and the I Ching for divination and the tradition of the secret books of Hermes Trismegistus and the medieval Jewish Cabala exemplify this belief or hope. Such sources of universal wisdom are traditionally protected from casual use by being difficult to find as well as difficult to understand and dangerous to use, tending to answer more questions and deeper ones than the searcher wishes to ask. The esoteric book is, like God, simple but undescribable. It is omniscient, and it transforms all who know it. The use of classical texts to foretell mundane events is considered superstition nowadays, yet in another sense science is in quest of its own Cabala, a concise set of natural laws that would explain all phenomena. In mathematics, where no set of axioms can hope to prove all true statements, the goal might be a concise axiomatization of all 'interesting' true statements.

"Ω is in many senses a Cabalistic number. It can be known of through human reason, but not known. To know it in detail one must accept its uncomput-

able sequence of digits on faith, like words of a sacred text. The number embodies an enormous amount of wisdom in a very small space inasmuch as its first few thousand digits, which could be written on a small piece of paper, contain the answers to more mathematical questions than could be written down in the entire universe—among them all interesting finitely refutable conjectures. The wisdom of Ω is useless precisely because it is universal: the only known way of extracting the solution to one halting problem, say the Fermat conjecture, from Ω is by embarking on a vast computation that would at the same time yield solutions to all other simply stated halting problems, a computation far too large to be actually carried out. Ironically, however, although Ω cannot be computed, it might be generated accidentally by a random process, such as a series of coin tosses or an avalanche that left its digits spelled out in the pattern of boulders on a mountainside. The first few digits of Ω are probably already recorded somewhere in the universe. No mortal discoverer of this treasure, however, could verify its authenticity or make practical use of it."

As this column was going to press I received a telegram from the notorious numerologist Dr. Matrix asserting that he is in possession of the first 1,101,121 bits of Ω (in principle enough to answer a good many uninteresting questions as well as all the interesting ones). He is currently soliciting bids on individual bits or consecutive blocks.

Here are some odds and ends having to do with earlier columns:

P. Howard Lyons and others found that the smallest rectangled rectangle, shown in this department for June, was not unique. The five subrectangles of the second solution to the problem are of dimensions 4 by 8, 1 by 11, 5 by 7, 2 by 6 and 3 by 9.

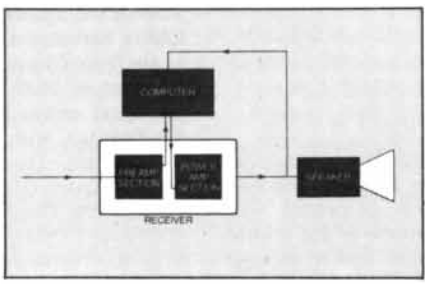
Abraham Schwartz pointed out that the joke chess problem given in June can also be solved in "a fraction of a move" by lifting either the bishop or the queen one inch above the board. Friedrich-Wilhelm Scholz was the first to send in the following novel interpretation of mating in a fraction of a move: White has advanced a pawn to QB8 and called for a queen; Black has removed the pawn but has not yet placed a queen on the square. Hence only the second half of the move is needed to complete the mate. Fred McCarthy also observed that the promoted pawn could have captured a black piece on either QB8 or QN8.

Mike Jones found a pattern of 20 nonattacking superqueens (pieces that combine the moves of a queen and a knight) for the 20-by-20 board, which leaves the 21-by-21 board as the lowest unsolved case.

How to get an honest 30Hz from a 1.25 cubic foot speaker system.

The story behind the new KLH Computer Controlled Loudspeakers.™

With the introduction of acoustic suspension more than 20 years ago, the loudspeaker industry took an impressive step forward. This technology allowed speaker manufacturers to achieve full-range frequency response in a cabinet substantially smaller than any previous full-range loudspeaker.



Since then, breakthroughs have come and gone in the industry. But none that has significantly reduced the size of a true, full-range system.

The reason is actually quite simple. Accurate bass reproduction requires a woofer to displace a large volume of air. In a small system with a small woofer, the woofer cone must therefore travel a long way to reproduce the lower frequencies.

Although a small woofer is perfectly adequate most of the time, occasional high level, low frequency signals can drive the cone well beyond its intended excursion, causing severe overload distortion.

To avoid this, it has been necessary to attenuate lower frequencies in smaller systems. Which is why small speakers have always had compromised bass.

The KLH Analog Bass Computer.™*

To solve this problem, we developed a completely new approach — computer control.

We designed a separate component, the KLH Analog Bass Computer, as an integral part of the entire speaker system. This component sits next to the receiver or amplifier and constantly monitors its output. The computer derives an electronic analog of cone motion, and controls the woofer at the precise instant at which overload distortion would otherwise occur.

With this kind of accurate, reliable control, our designers were free to extract the optimum theoretical performance for any given cabinet size. And develop a line of loudspeakers that can deliver extended bass response in cabinets that are substantially smaller than ever before possible.

The KLH-1 is one example. From a 1.25 cubic foot cabinet, it delivers bass to 30 Hz (−3dB) at 105 dB s.p.l. with absolutely no possibility of overload distortion.

Beyond the Computer.

Since the Analog Bass Computer and the speakers must be designed as a single, integrated system, we started from scratch with the objective of optimizing our new technology.

To achieve the widest possible bandwidth with acceptable efficiency, we employed sixth-order equalized systems. Combined with the Analog Bass Computer, these systems provide a −3dB point equal to conventional acoustic suspension systems of at least four times their volume.


The new KLH-1 with Analog Bass Computer

In keeping with our objectives, we also refused to compromise other elements of the design.

For our cones, we selected polypropylene, a material first developed for use in studio monitors by BBC engineers. The movement of polypropylene reflects the electrical signal more faithfully than either paper or bextrene. The result is a remarkably clear, transparent, uncolored midrange.

For our speaker baskets, we used die-cast aluminum rather than stamped steel.

And we used massive magnet assemblies, optimized for the sixth-order design.

Three Applications.

Finally, we applied all we had learned to accomplish three distinct objectives.

Our first objective was to produce a speaker that raises the absolute level of low-frequency response in a cabinet that is still practical for the home environment. The new KLH-1 does exactly that. It delivers flat bass to 30 Hz (−3dB) from a floor standing unit just 11″ x 30½″ x 10¼″. At a price per pair of $1100** including Analog Bass Computer.

Our second objective was to provide the best possible combination of price and performance. Our solution is the KLH-2. At $660** per pair with computer, the KLH-2 can deliver flat bass to 38 Hz (−3dB) at 102 dB s.p.l. from a cabinet that measures 10¼″ x 21″ x 8½″.

Our third and final objective was to design a moderately priced speaker with performance equal to or better than anything near the cost, in a cabinet one fourth the size. This is the KLH-3. It measures 8½″ x 12½″ x 6″, delivers bass to 40 Hz (−3dB) at 95 dB s.p.l. and costs $450 ** per pair including computer.

The new line of KLH Computer Controlled Loudspeakers.

Listen to them.

For more information, call toll-free 800-225-6042 (in Mass. 1-800-532-9566). Or write KLH Research and Development Corp., 145 University Avenue, Westwood, MA 02090. In Canada: The Pringle Group, Ontario.

KLH

*Patent applied for.
**Manufacturer's suggested retail price.

BOOKS

Science in traditional Japan, the computer analysis of literary style, bioluminescence

by Philip Morrison

SCIENCE AND CULTURE IN TRADITIONAL JAPAN, A.D. 600–1854, by Masayoshi Sugimoto and David L. Swain. The MIT Press ($39.95). In 1754 three medical men sought and received official permission to dissect one of five thieves executed in Kyoto. The master physician Toyo Yamawaki stood behind the scenes; it was his students who had initiated the unprecedented request. Within a few years his small book, six pages of text and four pages of drawings, appeared. The treatise, which dealt mainly with the visceral organs and ignored the head, the muscles and the nerves, had an "extensive impact" on the physicians of Japan. There was a vigorous and sophisticated national debate. Was not the body the gift of the ancestors, not to be dissected without impiety? Even if the traditional Chinese drawings were false to reality, was it not function rather than structure that mattered? Does the study of dead organs have any therapeutic utility? Yamawaki knew still more: he had a Dutch version of a post-Vesalian anatomy published by a Paduan professor in 1633. Unable to read the text, Yamawaki was "shocked by the accuracy and precision" of the figures. "Dutch learning" offered tangible support for the unsure and concerned practitioner. There was a boom in Western medical studies, viewed through the narrow channel of Dutch books, yet Japanese practice was still linked to the subtle Chinese systems of medicine, with their wide philosophical and cosmological framework. Innovators arose. One was the remarkable Seishu Hanaoka, who removed a breast cancer under a general anesthetic of datura and aconite alkaloids in 1805, the first major surgery under anesthesia recorded anywhere in the world. (One of his remarkable step-by-step illustrations is reproduced, although not in the original color.) His school thrived for generations, in semisecrecy.

Or consider *wasan,* literally meaning Japanese mathematics. From a freshet of interest in about 1620, which saw the publication of many books in extension of the then novel Chinese work of the time, *wasan* grew over two centuries into an esoteric and "proudly impractical" specialty of certain samurai scholars. The training schools of *wasan* were called by the same name as the schools of tea ceremony or of flower arrangement. It was a discipline quite distinct from those of surveyors or astronomers, held determinedly free from applications, "a leisured art." One scholar wrote of three styles of mathematics: the extremes were *yo no yo,* "useful for some use," and *muyo no muyo,* "not useful [and] for no use"; in other words, applied and pure. *Wasan,* he held, ought properly to take the middle way: *muyo no yo,* "useful for no use," useful in method but useless for results. These mathematicians thought little of the West; they disdained Euclid because he solved no hard problems. They never received much of the pure mathematics of Europe, no analytic geometry and only a little of the calculus. It all seemed clumsy and inferior to *wasan,* whose discoverers had made an algebra of some strength, a number of studies of infinite series and a scheme for the use of stacks of tiny rectangles to calculate in the limit the area and volume of any figure. By the early 19th century the Japanese had the equivalent of tables of definite integrals of some functions. The problems they attacked were primarily the properties of very complicated figures, blind-alley issues such as the volumes contained in the intricate intersections of several cylinders. They printed introductory methods, but the highest skills remained in handwritten secret manuals. Schools formed around a succession of masters, from Soto I to Soto VIII in one example, each solving problems and setting others for his successors, in a linked chain nearly three centuries long.

The authors of this fascinating and substantial volume have not emphasized the anecdotal matter sampled above. Their survey of science in traditional Japan is broad and contextural, their emphasis on the ebb and flow of style and institution. As the valuable foreword by Nathan Sivin puts it, they take the large view of "the movement of knowledge and ideas into a society and the rhythm of their assimilation." They nonetheless concentrate on three central fields of science: astronomy and its applications (astrology, geodesy, navigation, calendrical science), mathematics and medicine. They discuss the important techniques, from shipbuilding to agriculture, wherever these bear on the sciences or simply evoke the quality of an epoch. They divide their chronicle into waves of eager importation and innovative seclusion, first taking from China between A.D. 600 and 894, then again from China between 1401 and the isolation of 1639. In that second period the West entered, the Portuguese and the Spanish, with the globe, the arquebus and the cross. From 1639 a limited flow from China and from Europe through the Dutch nourished the scholars. After 1720 books could be imported from the Netherlands and Dutch could be legally studied. This became the epoch of Dutch learning; a clear conflict arose between the neo-Confucianism of the Ming and the new work of the distant redheads.

The first Chinese wave had been sponsored by the rulers themselves, the founders of the Japanese state. They dispatched entire expeditions to China to learn the culture of the great T'ang. One such party of four vessels might carry, besides servants, a diplomatic team, a study team in two sections (one for Chinese law, science and the like, the other for Buddhist doctrine), a ship's crew (many oarsmen for the times of unfavorable wind) and a technical team of physicians, pharmacists, diviners, artists, jewelers, foundrymen and more. All told the group might number 500 or 600. Thus there arose in Japan the new capitals, successively Fujiwarakyo, Nara and Kyoto, all magnificent if smaller models of imperial Ch'ang-an. After the building of Kyoto the impetus declined, the new rulers being preoccupied with a new life in Japan. In a few centuries the sciences had atrophied. The proud imperial institutes faded into family monopolies at court. In the 17th century it was not the state but the private scholars who printed, for example, the treatises and tables, based mainly on the Chinese-Islamic system of the 14th century, that led to a much-needed reform in the Japanese lunisolar calendar, the first since its introduction from China eight centuries before.

A reader familiar with the books from Padua and Bologna, with Trinity Great Court and Franklin's kite, with the rise of logarithms and the Ether Dome will find this narrative both deep and strange. The human themes and the growth of authentic knowledge are familiar, yet the emphasis, the order, the events and the characters are all new. A couple of excellent histories of Japanese astronomy and mathematics already exist in English, but nothing as broad as

Ford Fiesta.
It received a seven-flag salute.

The car that wowed Europe is winning the hearts of America.

Ford Fiesta. The little front wheel drive car that comes from Germany. Applauded all over Europe by the experts for its engineering, design and overall performance. It was voted the most significant import of the year in 1978 by readers of Car and Driver.

Germany 1976 Car that Makes the Best Sense — *Mot*

And when you understand how beautifully Fiesta performs, you'll understand why it continues to get rave notices.

Finland 1976 Car of the Year — *Tuulilasi*

Front wheel drive.

Denmark 1976 Car of the Year — *Morgen Posten*

Through rain, ice, sleet, hail and snow, Fiesta's remarkable front wheel drive traction will help you keep your appointed rounds. And its Michelin steel-belted radials will help you come to grips with all kinds of roads.

Italy 1977 Most Successful Foreign Car — *Motor*

Best gas mileage of any German import.

Americans love Fiesta's manners. It prefers sipping to guzzling. 1979 EPA est. mpg:

28 EPA EST. MPG **39** HWY. EST. MPG

Compare this estimate to other cars. Diesels excluded. Actual mileage may differ depending on speed, weather and trip length. California estimates will be lower. Actual highway mileage will probably be lower.

Yugoslavia 1977 Car of the Year — *Automotive Writers*

A masterpiece of European engineering.

Fiesta is assembled by Ford in Germany. And its European engineering makes it feel right at home on streets and highways of America. It's quick, nimble and maneuverable.

Spain 1977 Car of the Year — *Criterium*

Ford Fiesta is sold and serviced by over 5,000 authorized Ford Dealers across America. There's even an Extended Service Plan available, providing longer protection over your car's basic warranty. So test-drive a Fiesta today. You'll discover why it's won international acclaim.

Great Britain 1978 Design Council Award

FORD FIESTA

FORD DIVISION Ford

Fiesta. Wundercar from Germany.

Fiesta 3-Door Sport

this book. There is a welcome promise too: a second volume will carry the story from the year of Commodore Perry's Black Ships through one decade of the shogunate's desperate efforts to gain the technology of cannon and steamers, then past one turbulent century up to our days, to the meson, the zoom lens and the supertanker.

LITERARY DETECTION: HOW TO PROVE AUTHORSHIP AND FRAUD IN LITERATURE AND DOCUMENTS, by A. Q. Morton. Charles Scribner's Sons ($17.50). This volume by a classical scholar who is also an experienced computer user stands convincingly and stably in both of the two cultures. One day the author was watching a line printer cranking out parts of a concordance of a set of Greek texts. He remembers the striking rhythms, now a line or two ticking out, then a long burst of lines. A glance at the printout led him to formulate a statistical hypothesis: however many times a pronoun was used by a Greek author, the fraction of the times he used it in the genitive form was about the same from text to text. The idea worked, not only for the lifelong output of the rhetorician Isocrates but also for a selection of other Greek authors. The size of the sample needed was examined in detail, with a watchful eye on the runs and correlations that tend to demand larger samples than might be required for purely random data.

Morton describes the principles and practice of the science of stylometry, first pioneered in its modern form on the works of Hippocrates by W. C. Wake in his London doctoral thesis of about 1951 B.C. (before computers). Of course the literati have long studied texts assiduously to fix authorship, often concentrating on the rare words that "show the subtleties of excellence." In the powerful if prosaic methods of this new scholarship these are set aside. Rare words are evidently not fruitful for statistics, and vocabulary reflects the subject more than the author. Once it was held that the gospel of Luke was written by a physician, since medical terms were frequent in it. The context of the words, however, was nonmedical, and their use was shared by the historian Josephus, never a physician. In the Book of Acts the same compiler, judging by his vocabulary, has become an old salt, displaying much knowledge of shipwreck. Nouns notably mark the topic and not the author.

The touchstone turns out to be what is frequent but obscure, the swift little choices that confront every author. These are habits not affected by the passage of time, change of subject matter, literary form and other testable circumstances. Take an example in English. (The book treats about equally the inflected classical or biblical Greek and the uninflected, word-order-stressing English.) "The inimitable Jane Austen of the complex style and biting wit" was once meticulously imitated. She left us an unfinished novel, with a summary, when she died in 1817. An anonymous admirer completed *Sanditon,* which was published in 1975 "by Jane Austen and Another Lady." That highly literate lady tried to imitate her coauthor, for example in the Austen habit, decried by Scottish schoolteachers, of using "and" following commas, semicolons and colons. In the habits of detail, however, she failed totally. Half-a-dozen samples from three different Austen novels agree well, with the difference levels of little significance. The unknown lady's text agrees with them poorly, with difference levels that would occur by pure chance only once in a billion. What are those habits of Austen's? They are typically the placement of small words. The use of "such" preceding "a" runs about 10 percent of all the "a"s in the three Austen samples, and it drops to just two out of 82 uses of "a" for the unknown lady. The same effect is found for the "I" following "and"; often used by Austen, it appears only once in the pastiche! It is too bad that a control is lacking here; if the unknown lady would offer some prose samples of her own, it would be fascinating to see whether her normal writing style bears the stamp hidden in her loyal tribute to Austen.

The 50,000 words of the 14 epistles attributed to St. Paul are a challenge to the stylometrist. Morton, himself a minister of the Church of Scotland, ruefully admits that the study of texts held sacred is not quiet scholarship. "Write on Shakespeare or Plato and angry letters begin to arrive. Write on Paul or Jesus and abusive and anonymous letters will surely flood in." The first hint of a statistical stylometry was offered in a letter written in 1851 by Augustus de Morgan suggesting the measurement of the distribution of word lengths as a clue to the authorship of the Pauline corpus. Indeed, W. C. Wake studied this same material, although he did so by measuring the distribution of sentence length. A painstaking study of sample size, with close attention to various anomalies of correlation and to the peculiar effects of repeated copying on initial and final pages, was required and is outlined here. In this detailed chapter the results come clear. Four of the Pauline epistles show on the average distinctly shorter sentences, and the frequencies of several simple words used as the first or second words in sentences concur in setting the epistles apart. A subtler measure, the occurrence of *de* (but) between successive occurrences of *kai* (and), was applied to a wide range of Greek authors by Morton and a colleague. That measure gives "astronomical" odds against a single author for all the epistles, not the four alone. The case gains strength not only from the high odds but also from the concordance of many distinct sets of hypotheses, a lesson in robust statistical studies. It is interesting that the radical German "higher criticism" of a century ago came to the same conclusion without statistics: only the big four epistles of St. Paul show a common origin.

More than a third of the book tries to make these ideas and the practice of statistics clear to the innumerate. The effort is engaging, and on the whole it is a successful start, although one doubts that the tyro will feel much at home with the full statistical apparatus. These include the faintly paradoxical negative binomial distribution, well suited to describe what would be a Poisson distribution if equal intervals were being compared but instead extended to treat intervals whose lengths themselves are randomly variable. There is scholarly amplification of the economics of ancient publication, where many authors sought to father their ideas on authority, since the money rewards accrued not to them but to the scroll seller. In the modern world it is the reverse: all the profits of fame and a share of fortune go to the named author, even when he has raised a ghost to produce his text. The authorship of confessions entered in legal trials is a special study in which Morton has also taken a lead. Forensic stylometry is at least an inhibition to the temptation on the part of police officers to adjust the admissions they present in court. Half the cases in the Crown courts of London are settled on the "bitterly contested evidence of two or more police officers who assert that the accused said something which he denies ever having said."

This is an unusual book by an unusual and gifted author, with an explicit appeal for those who live far from numbers and calculation. At the same time it offers the scientific reader an exemplary case of the difference between the blind use of rich but ill-examined data and the careful and ingenious effort to obtain samples apt for analysis by the powers of modern statistics.

BIOLUMINESCENCE IN ACTION, edited by Peter J. Herring. Academic Press ($51.75). We are eye-minded, diurnal primates. By daylight we enjoy the colors of nature. Bright flowers and green grass speak to us strongly, although their evolution was entirely before our own. By night most of the color has gone, but the signals offered to the sensitive eye by the intrinsic luminescence widespread among living forms, particularly in the seas, intrigue us as color does. The phenomena have been studied in fascination at least since the time of Robert Hooke and Robert

Boyle. Some 30 years ago the physiologist E. Newton Harvey reviewed a lifetime of work on bioluminescence in a scholarly volume of profound influence. In this rich review 17 investigators from the U.S. and Europe pool their specialties to make a volume less personal than Harvey's, which glows even more strongly with the light of understanding. This is not to say, however, that there are even tentative solutions yet to all the big questions of luminous life.

The how of the life glow is clearer. Photochemically, a general method of efficient light emission is known. The excitation energy can be held in a singlet spin state safe against the collisional deexcitation of the electron because that would require a magnetic spin flip, a slow process for the usual electronic interactions. If that state can flip by radiation, it becomes an efficient light source. Sometimes it transfers stored energy to another particular molecule that actually emits. In vitro cell-free systems from fireflies and ostracods have been studied for more than 20 years; their enzyme luciferin is no longer a mere functional name like vitamin but a clearly defined molecule. Even the crystalline material has in a few cases been prepared. For a long time the kinetics of firefly glow has been applied as a routine bioassay for the ubiquitous biofuel ATP. (That use is in part a tribute to the interest in good photometric measurements.) Several distinct compounds are now known that play a similar part: three of them are closely related multiring molecules, found in many marine forms, all quite distinct from the simpler three-ring firefly luciferin. In luminous bacteria the compound whose oxidation is the key step is instead a long-chain aliphatic aldehyde, six or eight links long. Other forms contain photoprotein, a large molecule probably coupled to some kind of oxidized luciferin, whose photon emission can be triggered by simpler reactions, in one case by the calcium ion alone. The detail available extends to the study of reaction rates; the whole story seems to lie well within the implied powers of modern biochemistry. These differing systems shed their light on broader issues, because their evolutionary formation is plausible by the canons of comparative biochemistry. Many species share the same mechanism at their base, an antidote to earlier views of the haphazard evolution of luminescence in many phylogenetically scattered forms.

The why of the glow is less sure. The firefly case is understood best. Here are marvelous experiments on behavior. A female firefly is held in a container out of which it can peer but from which its own glow does not leak. A phototube alone records its hidden signal, in response to the flash of males some dis-

NASA PATENT 4,052,648

NASA Hits

A new invention by America's space agency will help all Americans save energy and make some companies very wealthy.

Exxon has it. So does about a dozen other manufacturers. And if our hunches are correct, a new space-age product invented by NASA may not only save Americans millions of dollars but make fortunes for the companies that sell it.

The new NASA invention uses the latest space-age technology to save energy. Your refrigerator for example, is a major energy user. With this new device, your refrigerator compressor will run quieter, there will be considerably less heat generated from the motor, and it will run more efficiently saving at least 30% in energy.

The invention requires no installation. Just plug it into your outlet and plug your refrigerator into the device.

OVER PRICED UNIT

But there's a catch. Most manufacturers sell the device for as much as $200. Using it with your refrigerator, it will take many years before it will pay dividends. On a powerful motor, however, the device will pay for itself in a matter of months.

Manufacturers who have announced their units are selling them like hot cakes. Although you may have heard a great deal of publicity about the product, you may not have seen any advertising because most manufacturers are currently sold out.

Watch for it! We predict great success for all those associated with the product. The power-saving device invented by NASA is a big hit. It will grow in popularity and save energy and make many companies very successful.

A SMALL COMPANY

There is one small company however, that is credited with improving the device and developing it for the consumer market. Called ERI (Electronic Relays, Inc.) the company has developed several models to service specific products such as a refrigerator, a washing machine, dishwasher, swimming pool and a typewriter.

This small company actually improved the NASA invention by adding its own refinements. ERI had a great deal of experience in solid state relays which use TRIACs and integrated circuits—two important elements in the NASA invention. A TRIAC is a bidirectional thyristor which controls AC from a single control input. TRIACs also produce a great deal of heat.

ERI's experience taught them how to control the TRIAC and its heat dissipation and thus they were able to reduce the device's cost through more efficient handling of the heat problem. They were already one of the nation's largest purchasers of TRIACs—thus their costs were already low.

NATIONAL PUBLICITY

They called their product, the Power Chopper and sent a sample to a national magazine for their review. In several tests, the device out-performed even the claims made by the manufacturer and the magazine ran a glowing article on their findings.

The manufacturer felt that the product might at first be misleading. Although it does save 30% on energy and in many cases up to 60%, ERI felt most consumers would expect a 30% reduction in their total electric bill—which of course the product will not do. Consumers will only get a 30% savings on the particular appliance used with the Power Chopper.

STILL PESSIMISTIC

The manufacturer also felt that the product was primarily for the industrial market—restaurants with large banks of refrigerators. The consumer must wait over one year before the device would pay for itself. And finally, the manufacturer did not feel that the consumer would respond in great numbers to the article which ran in the July, 1979 edition of a popular magazine.

Well, the consumer did respond. So much so that the small manufacturer, with absolutely no marketing staff, was buried with mail. The president of ERI called JS&A to help him out.

TEST ONE YOURSELF

We agreed to offer the Power Chopper to the consumer market for $29.95—a major price breakthrough for the product.

Even if Exxon lowers their prices considerably, they'll never come close to the low cost of the Power Chopper. ERI's expertise with the TRIAC and JS&A's direct-to-consumer marketing, make the new NASA invention a practical power-saving accessory for every home.

The sophisticated electronics of the Power Chopper consist of a TRIAC, two integrated circuits and several solid-state devices.

We urge you to simply test just the refrigerator module. Order one from JS&A on a 30-day no-obligation trial. In the meantime, while you are waiting for your unit, feel the heat generated from the bottom of your refrigerator. Listen to the sound level of your compressor.

When the Power Chopper arrives, plug it in and notice how much quieter and cooler your refrigerator runs. See how much less time the compressor must run. The compressor not only will run more efficiently but will save energy every day you use it.

AWARD WINNER

If after 30 days you are not convinced that the Power Chopper will save you energy and money while making your refrigerator run smoother, then just unplug it and send it back for a prompt and courteous refund, including the $2.50 postage and handling. But if you've definitely noticed the difference, you'll want to purchase more units for the remainder of your motor-based appliances.

JS&A feels that ERI's technology, their improved NASA design and their low manufacturing costs will catapult them to the forefront of those introducing the new NASA invention. ERI's Power Chopper is one of the nation's major new innovative products and just recently won the Industrial Research IR-100 Award.

To order your Refrigerator Power Chopper, send **$29.95** for each unit plus $2.50 for postage and handling to JS&A Group, Inc., One JS&A Plaza, Northbrook, Illinois 60062. (Illinois residents please add 5% sales tax.) Credit card buyers may call our toll-free number below. We'll send your Refrigerator Power Chopper, one-year limited warranty and you'll be ready to save. If you wish to order additional units for other appliances at $29.95, you may, but we suggest you test the refrigerator module first and totally convince yourself.

GOVERNMENT REBATE

Purchase of the Power Chopper entitles you to a full 15% energy tax credit on your income tax return. It's like having the government give you a $4.50 rebate.

JS&A is America's largest single source of space-age products—further assurance that your purchase will be backed by service for years to come.

NASA technology was responsible for the development of the integrated circuit and many other space-age products. Their latest product could not have been developed at a better time. Start saving and order a Power Chopper at no obligation, today.

JS&A® PRODUCTS THAT THINK®

Dept.SA One JS&A Plaza
Northbrook, Ill. 60062 (312) 564-7000
Call TOLL-FREE 800 323-6400
In Illinois Call ..... (312) 564-7000

tance away in the warm evening. The experimenter, the Italian zoologist F. Papi, uses the captive signal to trigger a nearby artificial flash, which can reproduce the natural one exactly or can substitute a square pulse of adjustable length and delay. In this way Papi has studied the language of those romantic flashes. Females that are artificially warmed above the ambient temperature flash too eagerly; they lose out in the male response to cooler individuals. There is less support for plausible theories for other forms, although they are often striking. For example, it is known that most of the many luminous creatures of the shallower seas, unlike the deep-sea lantern bearers, do not have cells of their own able to glow. Instead they provide an organ adapted to shelter and nourish certain symbiotic luminous bacteria. The bacterial glow can be modulated by a device as simple as an eyelid. But the bacteria in question, whose luminous biochemistry is known in some detail, almost never glow while they are free-living. Their glow system is induced by the environment provided by their symbiont host.

It seems probable that the glow of the bacteria developed in coevolution with the symbiotic habit. There is a strong, if unproved, case that the downward lights fitted on the ventral surface of many forms of the upper and middle depths act as a form of defensive camouflage, as countershading or, more properly, counterlighting. When the light organs match the brighter sea surface overhead, as it is seen by predators looking up from below, the telltale shadow of the prey is effectively eliminated. Even more wonderful, a number of these forms sport a small luminous organ on top, positioned to appear in the visual field of the animal itself. A theorist has argued that we view here a feedback scheme. The fish that needs to match its belly brightness to the changing glow of the waters above has evolved a system designed for exactly that feat. "The anatomy of the... arrangement is so beguiling that it is hard not to accept its working just as proposed." Behavioral proof, however, is not yet at hand.

The organism whose blooming is the red tide, a dinoflagellate, provides a ubiquitous sparkle at sea and the glow of certain phosphorescent bays. The source is known even in vitro. It appears to be a tiny particulate package of glow biochemistry; the light seems to consist mainly of fast flashes, with a slow diurnal rhythm. We begin to see how the internal membranes on which the particles are held can respond by emitting light on being mechanically disturbed. But why the organisms should glow so commonly we do not understand. Are predators frightened by the spark? Do some competitors stay away from the light, which they interpret as indicating too shallow a depth? If erupting patches of light or great phosphorescent wheels kilometers across are real, as it has been reported, they might be due to a wavelike disturbance of the organisms. It is better to admit that "we presently understand surprisingly little of the causes, distribution and importance" of the background night light of the sea.

The overall impact of the book, which contains well-written summaries from the detailed biochemistry and measurement methods to lists of luminous organisms by species, is likely to last as long as Harvey's did so many years ago. These are wonderful puzzles, whose elucidation challenges the effort of many ingenious investigators.

THE CALIFORNIA WATER ATLAS, William L. Kahrl, project director and general editor. Illustrated with maps, photographs, drawings and graphs in color and in black and white. Distributed by William Kaufmann, Inc., Los Altos, Calif. 94022 ($37.50). In 1850 the basin of the Central Valley of California, 400 miles long from the fogs of Lassen Peak to the desert ridge at Tehachapi, was a "vast expanse of alkali flats, grassland prairie and marshlands." The bottomland forests and the wetlands filled with tule reeds sheltered a diverse wildlife; yellow-billed cuckoo, mink and river otter were common. A dozen mountain rivers contributed the rushing flow they themselves collected from the annual snowmelt high in the Sierras. They formed the great meandering streams of the flatlands, the Sacramento to the north and the San Joaquin to the south. The two central rivers built an inland delta, which spills their joined waters into San Francisco Bay. Irrigation is 5,000 years old in the ancient valleys of the Middle East, but in the Central Valley the entire great system belongs to this century. It waters more cropland than the Nile and the Tigris-Euphrates together. Across two pages of this atlas the California waterscape of today is mapped, a vascular system of 135 canals, aqueducts, drains, tunnels, conduits, pipelines, laterals, interties and a couple of plain ditches. The lakes and reservoirs they connect are scattered throughout the high country. The cleanest tenth or so of this life water is sent to the cities, but the bulk of the transport is intended for the wide fields, the rice of the northern valley, the cotton and grapes of the southern one and the diverse rows, orchards, pastures and hay of both valleys. The smaller irrigated regions of the state are also fully treated.

The big pages (16 by 18 inches) of this well-focused instrument admirably mix text, textual asides, full-page maps, reproductions of old maps, elaborate statistical diagrams, flow charts, glowing Landsat photographs from orbit and evocative photographs from tunnels underground to the snowpack of the Sierras. The matter of the book is about half retrospective, about half analytical and reflective of present and future. Good maps show the virgin waterscape, the snow depth, the detail of the complex arteries that bring the water to San Francisco and to Los Angeles, the hydropower by site and the applied irrigation water by location and crop, the groundwater resources (mapped here for the first time) and a couple of dozen more such handsome and useful pieces of modern cartography. The text is a compilation of materials produced by an entire group of advisers with multiple viewpoints; the essays both read smoothly and show a long-range order not usual in such reference sources. Time is by no means neglected; this is not a set of still pictures but a work with the cinematographic viewpoint one might expect from Californians.

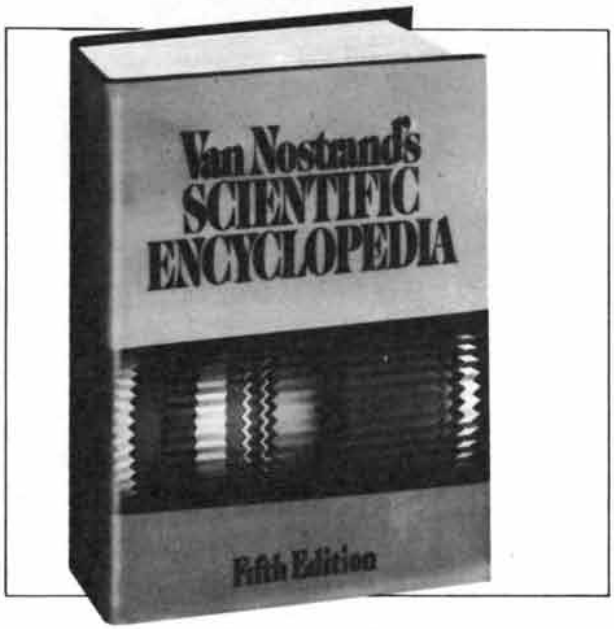
The flow of the main streams is shown as it is measured month by month and as it has been projected back to their unimpaired state, the peak stream flows are gathered graphically for a century and the twin drought years of 1976 and 1977 are well recorded. The reservoir that serves suburban Marin County then ran very low indeed, but the statewide agricultural product remained almost unchanged, partly because the south had had a normal rainfall and snowpack and partly because 10,000 new wells tapped the ground-water basins of the valley—deficit financing. Half of the crops were watered by wells, even though the low state of hydropower reserves imposed a big need for diesel fuel on the water seekers. Such variability is of course no new thing. The year 1924 had seen a similar one-year drought, the worst in the century up to 1976–77. On the other side of the ledger, in November of 1861 the rains began, and they remained heavy for nearly two months. Then a tropical storm unseasonably thawed the deep snowpack. The consequences are legendary: an inland lake 60 miles across filled the Sacramento Valley, and what is now the Los Angeles metropolitan area was inundated, the waters collapsing adobe houses and washing away carefully tended vineyards.

The book looks at the economics of water, the projections for demand growth, the plateau of new supply after so much ambitious engineering. Here is Hetch Hetchy, the valley shown wild and then flooded, and John Muir, who lost that battle for wilderness. Truth to tell, the waterfalls still flow picturesquely into the mirroring waters of the reservoir. Los Angeles took its bigger needs up to the Owens Valley, with Tommy guns as useful as theodolites,

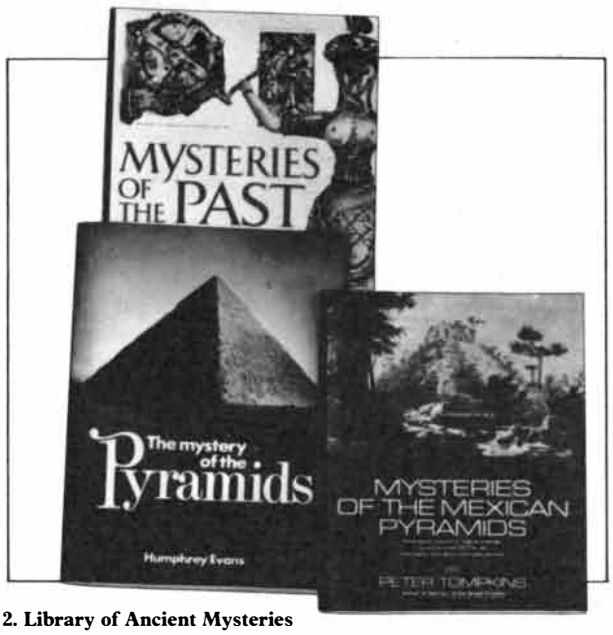
There may still be places on earth where Grand Marnier isn't offered after dinner.

Product of France. Made with fine cognac brandy. 80 proof. Carillon Importers, Ltd., New York, New York 10019. © Carillon Importers, Ltd.

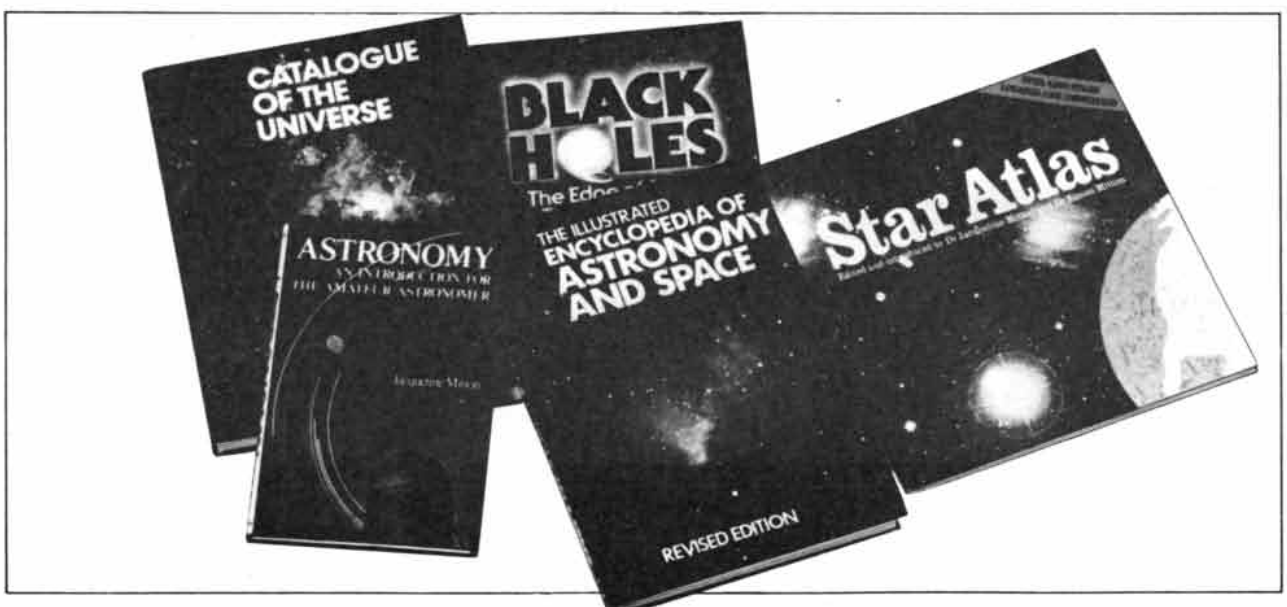
5 good reasons for joining the Natural Science Book Club



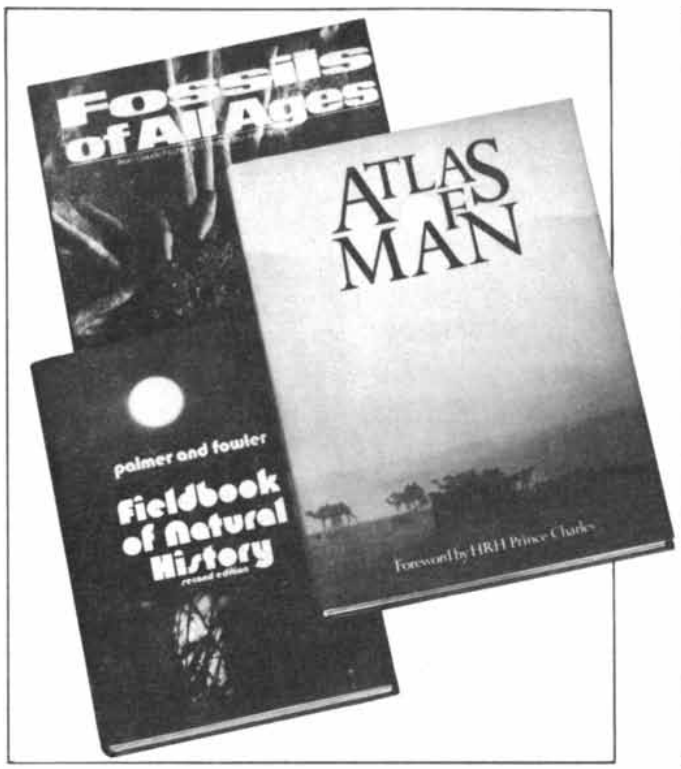
1. Van Nostrand's Scientific Encyclopedia
Edited by Douglas M. Considine. Nearly 200 experts have contributed to this thoroughly revised and greatly expanded fifth edition of the most authoritative single-volume source of scientific knowledge ever assembled. Enormous 9¼″ x 12″ volume contains 2.2 million words, 2382 pages, 2500 photographs, drawings and charts, and 500 tables. 7200 articles cover mathematics, from information sciences to physics and chemistry. "...an amazing book... for both the general and scientific reader."—*The New York Times.* Publisher's Price $67.50.



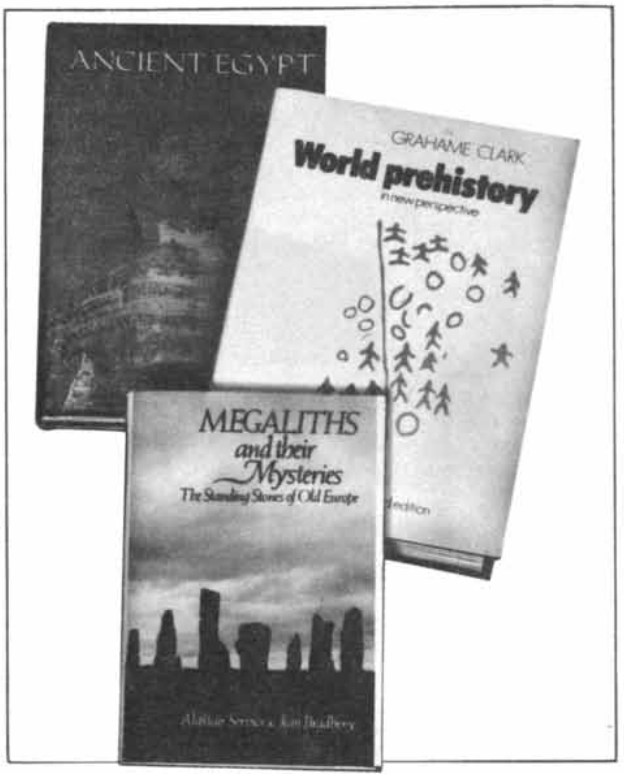
2. Library of Ancient Mysteries
Stunningly beautiful set consisting of three 8½″ x 11″ volumes overflowing with magnificent color photographs consists of ***Mysteries of the Past***—a captivating investigation of such great riddles as who raised the megaliths? what caused the sudden rise of modern man? what caused the collapse of the Maya? and many more; ***The Mystery of the Pyramids***—a thoroughly fascinating overview of Egyptian civilization and the historical, religious, cultural, and political reasons for its awesome pyramids; and Peter Tompkins's ***Mysteries of the Mexican Pyramids***—an absorbing look at the Mesoamerican pyramids and the civilizations that built them. Publishers' Price $74.90.



3. Astronomy Library
Here's a carefully selected set of five books that will provide you with a basic, well-rounded astronomy library. First, a "must" for any astronomy reader, the ***Illustrated Encyclopedia of Astronomy and Space.*** Then, Jacqueline Mitton's invaluable introduction to the fundamentals, ***Astronomy: An Introduction for the Amateur Astronomer.*** Next, a marvelously readable overview of the latest advances in the "new astronomy." Walter Sullivan's ***Black Holes***. Then, the ***Catalog of the Universe,*** which features more than 300 photographs. Finally, a useful guide that will show you how to locate and identify more than 4,000 stars that are visible to the naked eye, ***Star Atlas.*** Publishers' Price $61.75.



4. Nature Library
This sumptuously illustrated three-volume set consists of Palmer and Fowler's ***Fieldbook of Natural History***—a huge volume acclaimed as the most authoritative and comprehensive field guide to plants, animals, and minerals ever published; ***Atlas of Man***—9″ x 12″ volume filled with over 450 full-color photographs and maps which cover the culture, climate, religion, and geography of more than 400 tribes and peoples; and ***Fossils of All Ages***—9″ x 13″ word-and-picture masterpiece revealing the history of fossils, how they are formed, and the rules that dictate their shape and symmetry. Publishers' price $79.95.



5. Archaeology Library
World Prehistory in New Perspective—three-volume set consisting of Grahame Clark's enormous, global survey of *all* ancient civilizations from the first paleolithic gatherings through the Stone and Iron Ages to the Minoans, Aztecs, and Vikings; T.G.H. James's ***An Introduction to Ancient Egypt***—a lavishly illustrated look at the history, religion, and culture of ancient Egypt; and Service and Bradbery's ***Megaliths and Their Mysteries***—a detailed exploration of the astronomical, religious, and symbolic significance of the neolithic structures of Europe. Publishers' Price $62.90.

Select either **Van Nostrand's Scientific Encyclopedia** or any one of the four outstanding sets shown here for **only $2.97**

You simply agree to join for a trial period and take only 3 more books, at handsome discounts, over the next year.

Values to $79.95!

As a member, you'll be kept informed of the most interesting, most important science books being published. You'll get to choose from a dazzling array of books—from fascinating accounts of the discovery of lost civilizations to speculations on the origin of our universe, from sumptuous word-and-picture books on the world of nature to stimulating books on where man came from and where he is going.

And you get these books at handsome savings. In addition to receiving up to $79.95 worth of books for less than three dollars when you join, you keep on saving substantially—up to 30% and occasionally more. Your total savings as a trial member, for example, can *easily* be over 50% if you include your introductory offer. And these are real savings, because our books are always of equal quality to the publishers' editions, *never* "economy" editions.

Moreover, you will *immediately* become eligible for our Bonus Book Plan, with savings of at least 70% off the publishers' prices. And you can choose your Bonus Book from *any* of the titles we offer.

Why not start enjoying the many benefits of Club membership? Join the **Natural Science Book Club** today.

If the reply card has been removed, please write to **Natural Science Book Club, Dept. 4-A7E, Riverside, N.J. 08370** to obtain membership information and an application.

HP computer and measurement advances

Series 30: the new, lowest-priced member of the HP 3000 computer family.

When you need immediate access to information, the new Series 30 can provide it with unprecedented economy. An entry level, general purpose computer, the Series 30 can operate as a stand-alone system or as a station in a distributed data processing network. It can simultaneously handle transaction processing, on-line program development, batch operations, and data communications in any of five high-level programming languages.

HP's proprietary silicon-on-sapphire (SOS) integrated circuit technology makes possible the small size and low power requirements of the Series 30—it is housed in a cabinet measuring only 24 by 36 by 18 inches, and is accompanied by a disc drive and system console.

All communications are handled by the new Intelligent Network Processor (INP), an SOS-based computer in its own right. The INP offloads the communications management task from the CPU, permitting high levels of system performance concurrent with communications to other systems.

All members of the HP 3000 family use fully compatible systems software†, and application programs written on any HP 3000 system can be run on any other without reprogramming, recompiling, or relinking. Thus the Series 30 can prove its value in two significant ways: for the company purchasing its first computer, software developed for the Series 30 will remain valid when the system is expanded; for the company wishing to exchange information among a number of departments or manufacturing sites, the Series 30 is an economical station in a distributed systems network of HP 3000's. Prices start at $49,750*.

†APL/3000 is available only on the Series III.

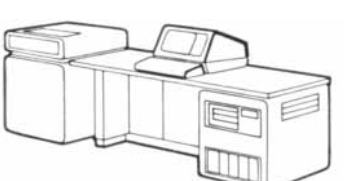
A family of compatible business systems for distributed data processing.

HP 3000 Series 30

The entry level Series 30 provides immediate information access at low cost, as a stand-alone system or as a station in a distributed data processing network.

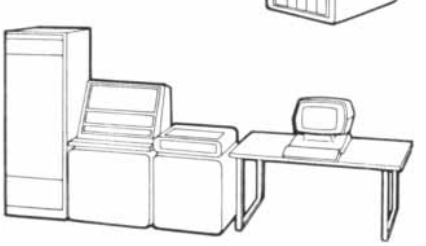
HP 3000 Series 33

The mid-range Series 33 offers an effective solution where several or possibly all of the functional areas of a business require efficient local information processing, with added capability for handling data communications lines.

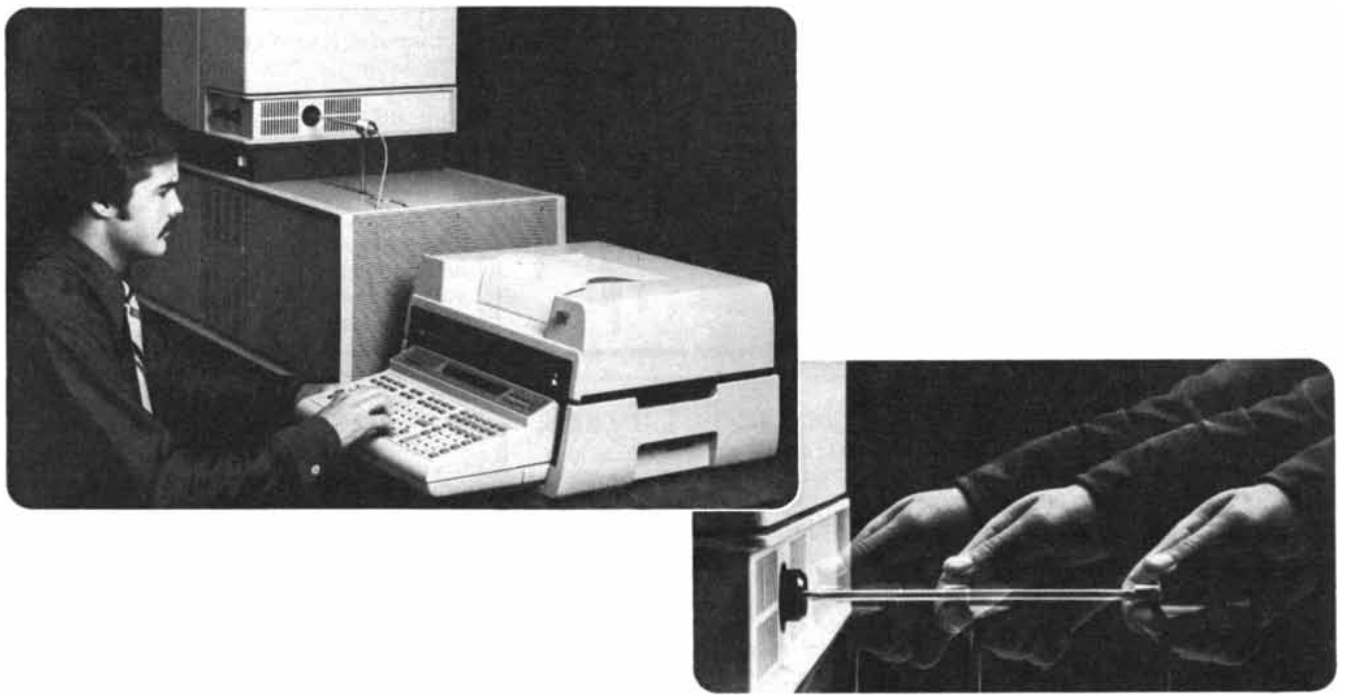


HP 3000 Series III

The top of the line Series III performs well in several roles such as supplying computer power to a large number of users, or centrally controlling a network of HP 3000's.



extend your possibilities.



HP's new benchtop gas chromatograph/mass spectrometer: a new standard in versatility, performance, and price.

The new HP 5995 is an economical GC/MS system with the versatility to meet a wide range of analytical needs. It is the first benchtop system with a direct insertion probe that gives the user the choice of bypassing the GC—a useful technique for rapid identification of pure substances, and an essential one for the analysis of compounds that cannot be chromatographed.

The HP 5995 incorporates recent advances in mass spectrometry such as hyperbolic rods in the quadrupole mass filter which improve spectral quality, and its HP computer technology eliminates the need for skill in tuning various mass spectrometer voltages.

An integral HP 9825 computer controls both the gas chromatograph and mass spectrometer, processes the data, and provides hard copy in a variety of formats on its thermal printer/plotter. Software supplied for the HP 5995 includes routines that tune and calibrate the system to Environmental Protection Agency standards, perform standard GC/MS analyses, do selected ion monitoring, plot and tabulate normalized spectra, and search spectral libraries for quick matching.

Like HP's other GC/MS systems, the HP 5995 is highly accurate, reliable, and easy to use.
Its price is $65,000*.

For more information, mail the coupon to Hewlett-Packard, 1505 Page Mill Road, Palo Alto, CA 94304. Or call the HP regional office nearest you: East (301) 258-2000, West (714) 870-1000, Midwest (312) 255-9800, South (404) 955-1500, Canada (416) 678-9430.

HEWLETT PACKARD

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

( ) HP 3000 Series 30 computer system
( ) HP 5995 gas chromatograph/mass spectrometer

Name _____
Company _____
Address _____
City _____ State _____ Zip _____

*Domestic U.S. prices only.

00950

the lore says. That supply matches the much newer flow to growing Los Angeles from the Colorado, but California mountain waters can generate hydropower on the way down, about enough to lift the Arizona waters over the intervening ridges. Both profiles are here.

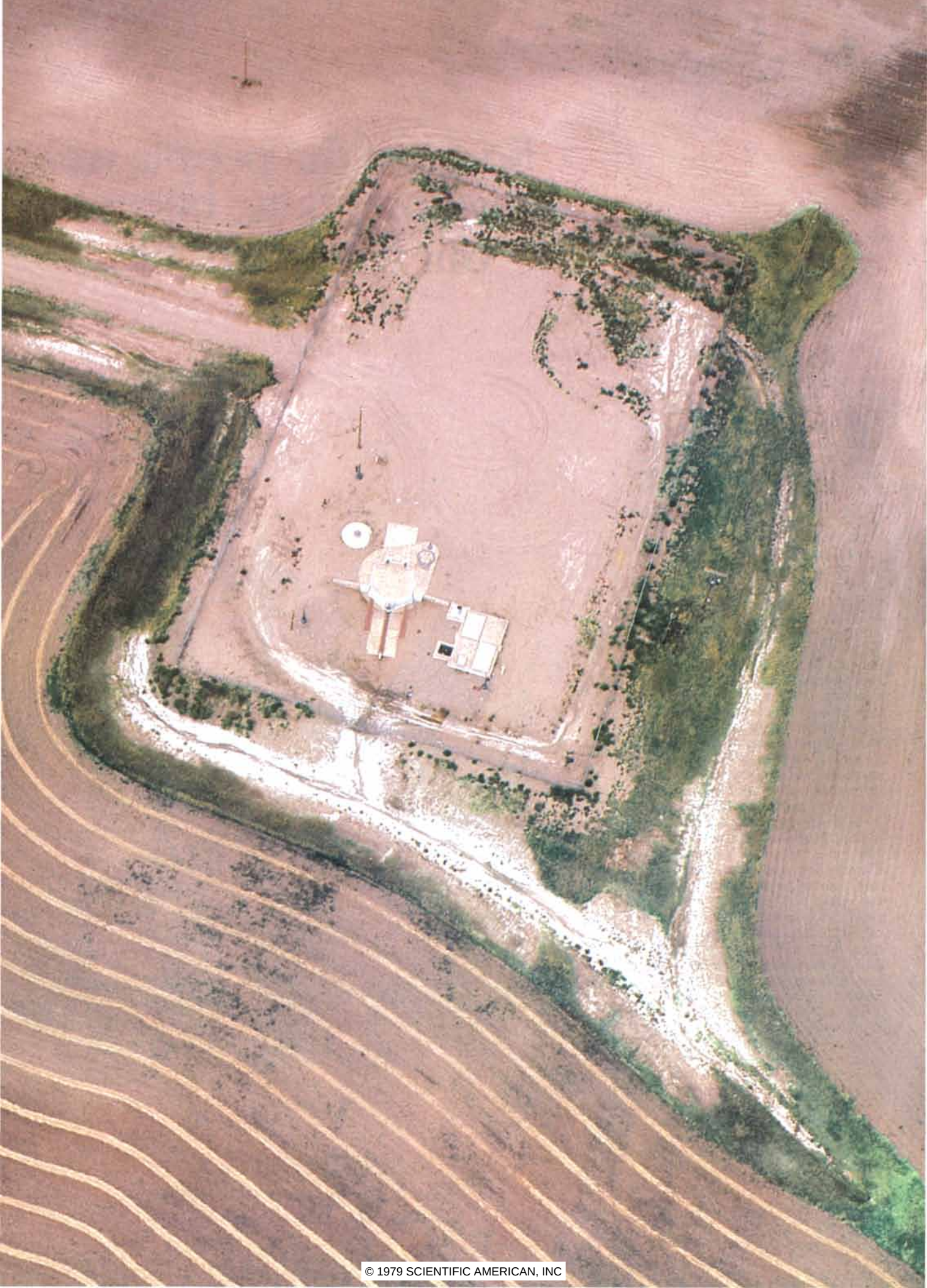
The questions are plain. Ground water is not as renewable as the snowpack. The salinity of the southern basins is steadily growing; a drainage canal to carry off the salts is stalled between engineering and economic difficulties. A similar tangle surrounds the plexus of the delta, where freshwater flow must be maintained to beat back the sea. The generous and proud works of the state and the Federal Government water the valley crops, but should the price of that water be kept at a fraction of its cost when almost half of the irrigation water feeds not people but their cattle? The text proposes no clear answers; its task is seen as being to provide a resource, not an adversary brief.

This striking book is the product of a wide collaboration. Its impetus came from the Governor's Office of Planning and Research, its data and expertise largely from the staff of the state Department of Water Resources and its cartography from an expert and modernly equipped team at California State University at Northridge. An entire parade of lawyers, artists, historians, journalists and other savvy people helped to make the volume a "source of pride for us all." It was done within a quite short time, mainly with state resources in the face of a fiscal drought, known nationwide as Proposition 13. The cost was a fourth of what some private firms estimated. A reader is pleased throughout, but another page or two of indexing apparatus, particularly a detailed list of graphs, tables and photographs, would have added a good deal.

MUYBRIDGE'S COMPLETE HUMAN AND ANIMAL LOCOMOTION: ALL 781 PLATES FROM THE 1887 "ANIMAL LOCOMOTION," by Eadweard Muybridge. Introduction by Anita Ventura Mozley. Dover Publications, Inc. ($100). Dark and light-starved, the relay-triggered sequence exposures on slow wet plates had made clear enough to Leland Stanford exactly how his swift horses ran. But Eadweard Muybridge, the romantic experimenter behind the work, left Palo Alto and his wealthy sponsor. The first Muybridge sequence photograph to appear in *Scientific American* was the cover of the issue for October 19, 1878. (The magazine was then a weekly.) By 1882, armed with much faster gelatin dry plates, Muybridge sought grandly to extend his pioneer studies of motion. He looked anew for well-heeled sponsors. He found them in prosperous Philadelphia, where the painter Thomas Eakins

and the wealthy sportsman-scientist Fairman Rogers were influential and enthusiastic. The provost of the University of Pennsylvania invited the project. There in the enclosure of the veterinary department, in the Gentleman's Driving Park and at the zoo the photographer set up his camera batteries, his scaled backgrounds and his wires. Between 1884 and 1887 the scheme was carried out on a satisfyingly heroic scale. Before his cameras moved athletes, mechanics, young boys, artist's models, racehorses, big cats and patients from the neurological clinic of Francis X. Dercum. The work was a clear success; its richness of form was high art; its wide concerns and its then unmatched division of time (measured by tuning fork) were of lasting interest to artist and analyst alike. Dercum published a study of the abnormal motions of his patients. The photographs did not, however, sell well to the public. The cost was very high. About three dozen complete sets of collotype plates—781 of them in all their variety, each a two-page plate with two or three dozen sequenced images—were bought by the very well-to-do and the finest libraries. Today the work has achieved a collector's notoriety for its rarity and its striking effects. More than 20 years ago the present publisher issued two selected volumes of plates, one on human subjects and one on animals and birds, which did a great deal to acquaint the world of the graphic arts with the work. The plates reproduced at that time were not quite half of the total, and the quality of the reproduction was not high. Another reprint publisher (Da Capo Press) then brought out a more expensive edition, with 100 excellent reproductions. Now we are offered in three volumes the entire corpus, quite strikingly reproduced, with the original invaluable (if hard to use) list of topics and technical data, and a helpful introduction by a Muybridge specialist, the curator of photography at Stanford University. The topics, as grouped by Muybridge himself (note that more than one version was originally offered for sale), are males and females (nude), males (with pelvis cloth), females (seminude or with transparent drapery) and children, males and females (draped) and miscellaneous subjects, abnormal movements, horses and other domestic animals, and wild animals and birds.

The books are a treasury of living form, whether one looks at the blacksmith at his anvil, the nude women serving and taking tea, the 340-pound model rising from the ground, the walk of locomotor ataxia or the "mastiff Smith aroused" by the report of a firecracker. The full price is now the same $100 Muybridge asked for a portfolio of 100 prints of the buyer's choice a century ago; today it is a luxurious bargain.



Established 1845 SCIENTIFIC AMERICAN November 1979 Volume 241 Number 5

Land-based Intercontinental Ballistic Missiles

Accurate multiple-warhead missiles are beginning to make fixed ICBM's appear vulnerable to a surprise attack. The proposed U.S. mobile missile, the MX, is viewed as an inappropriate response to the perceived problem

by Bernard T. Feld and Kosta Tsipis

From the early 1960's until quite recently the strategic nuclear forces of both the U.S. and the U.S.S.R. were generally regarded as being invulnerable to a preemptive "counterforce" attack by the other side. This effective military standoff was a central factor in the emergence of the concept of mutual deterrence (or, as it is sometimes called, mutual assured destruction). According to this view, if the strategic nuclear forces of both sides could be counted on to stay mutually invulnerable (that is, if each nation, even after absorbing a surprise first-strike attack by the other, could be sure of retaining a sufficient number of deliverable nuclear warheads to inflict unacceptable retaliatory damage on the attacking nation), then no rational leadership would be tempted to start a nuclear war.

Recent technological advances, for the most part introduced by the U.S. but closely emulated by the U.S.S.R., have contributed to a decline in the acceptance of deterrence as the guiding principle of the strategic confrontation between the two superpowers. These developments have also given rise to an erosion of confidence in the future invulnerability of one component of the U.S. strategic forces: land-based intercontinental ballistic missiles (ICBM's). In response to concerns about this anticipated problem it has been advocated that a new mobile missile, the MX, begin replacing the fixed Minuteman ICBM's sometime in the next decade.

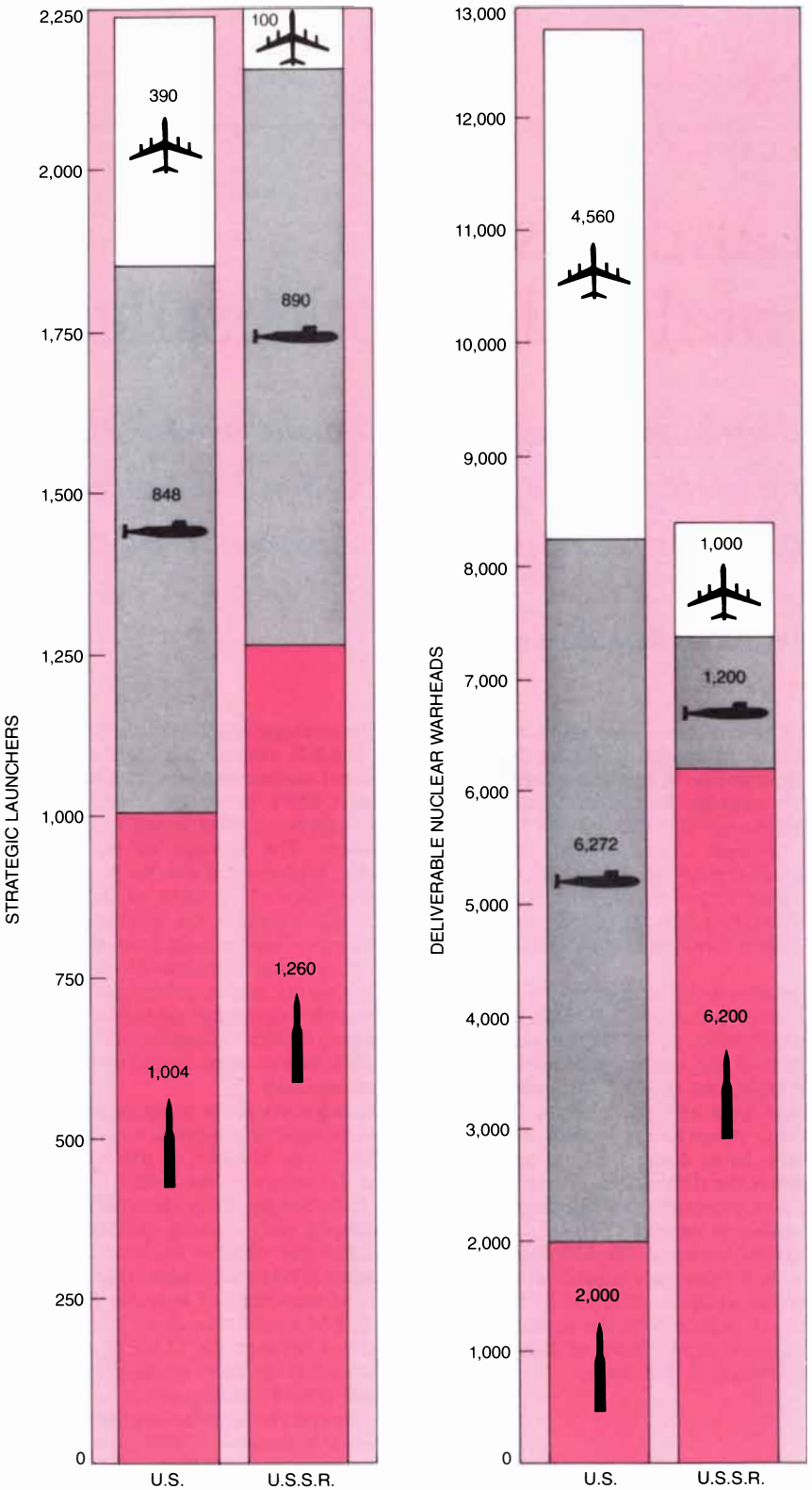
In this article we shall review the innovations that have led to the projected vulnerability of the Minuteman force. We shall also explore the various strategic situations in which this vulnerability could arise and try to predict what the actual vulnerability would be in those cases. In so doing we shall attempt to answer the difficult but crucial question of how vulnerable the Minuteman force is likely to appear to the military and political leaders of the U.S.S.R. Finally, we shall examine a number of technical options available to the U.S. that could help to reduce both the actual and the perceived vulnerability of this country's land-based ICBM force.

TARGET of a hypothetical "counterforce" attack by the U.S.S.R. against the land-based component of the U.S. strategic forces is seen in the aerial photograph on the opposite page as it might appear from the vantage of an incoming Russian nuclear warhead. The white-faced pointed shape at the head of the short tracked ways at the left center is the lid of an underground Minuteman missile silo at a location near Minot, N.D. The silo, the reinforced concrete and steel cylindrical structure that houses the missile, is designed to withstand a pressure of several thousand pounds per square inch over atmospheric pressure. In order to be sure of destroying the Minuteman in its silo a Russian nuclear warhead with an explosive yield of 750 kilotons would have to be detonated at ground level within about 550 feet of the silo. The adjacent structure at right center is the top of the missile's unmanned security and maintenance facility. An array of automatic sensors surrounds the silo within the rectangular perimeter fence. White streak at the bottom is a drainage ditch. Patterns on the fields are from contour plowing.

The strategic arsenals of the U.S. and the U.S.S.R. consist at present of three different nuclear-weapons delivery systems: ICBM's, submarine-launched ballistic missiles (SLBM's) and long-range bombers. The strategy of deterrence places a high premium on the postattack survivability of all three of these systems. For example, the fraction of the submarine-based missile force that is at sea at any time is essentially invulnerable to enemy action; submerged ballistic-missile submarines are hard to detect and even harder to destroy, particularly if all of them must be caught by surprise simultaneously.

Strategic bombers could in principle be vulnerable to a surprise attack by the ICBM's and SLBM's of the other nation. To minimize the vulnerability of the U.S. bomber force this country has developed early-warning systems: satellites that can observe the launching of Russian ICBM's and radars that can detect the launching of Russian SLBM's. An ICBM would take about 30 minutes to travel between the U.S.S.R. and the U.S., so that the early-warning satellites would provide ample notice to enable alert aircraft to fly away from their bases once a massive ICBM launch had been confirmed. SLBM's fired from Russian submarines near the U.S. coast, on the other hand, could take as little as five minutes to arrive on target, which might not be enough time for all the bombers to escape.

In order to minimize the vulnerability of the land-based ICBM's to a surprise attack both countries have emplaced these missiles in buried reinforced-concrete and steel structures called silos, where they are maintained in operation-

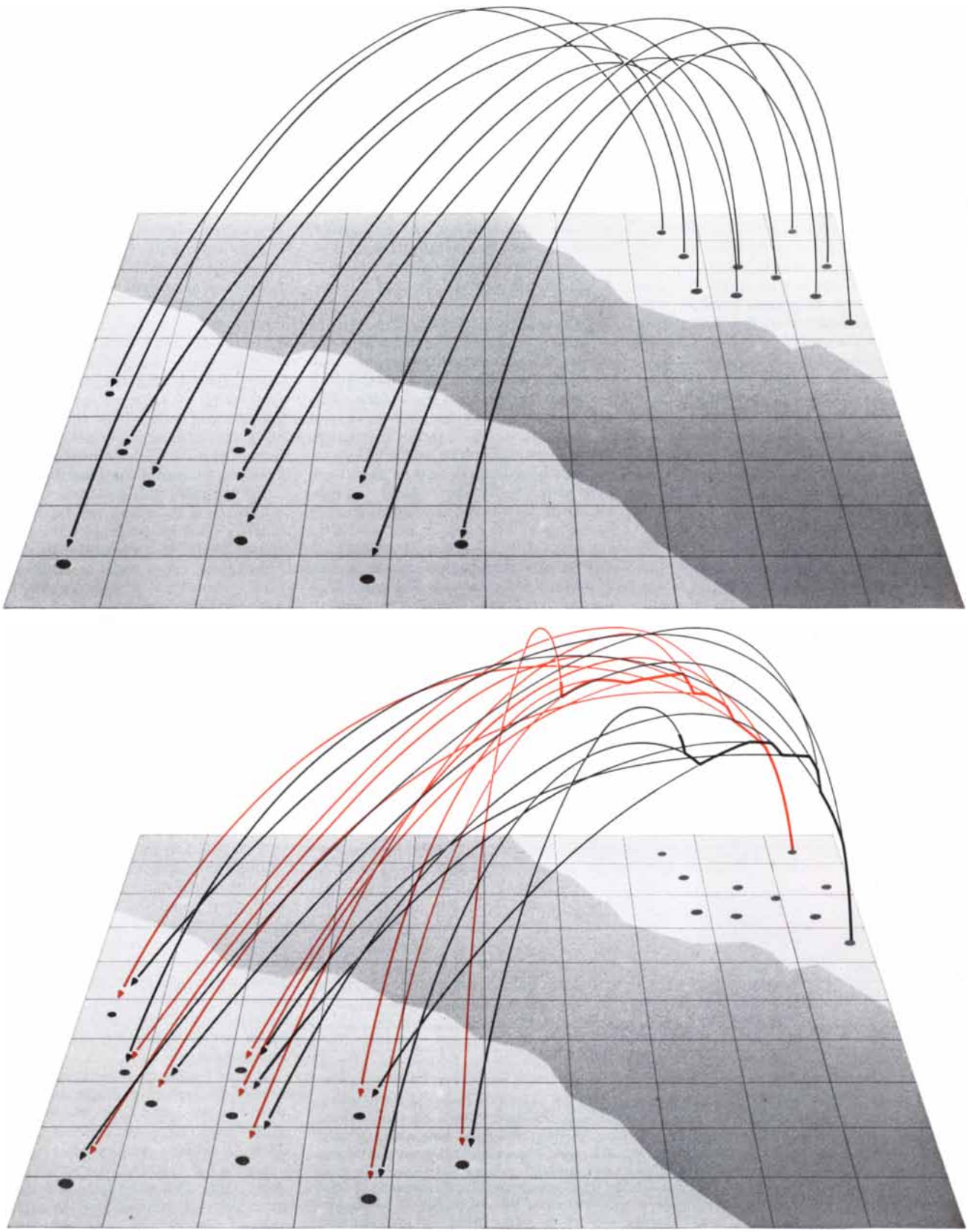


PROJECTED STRATEGIC FORCES of the U.S. and the U.S.S.R. in 1985, when the SALT II treaty is scheduled to expire, are represented here in two ways: in terms of strategic nuclear-warhead delivery systems (*left*) and in terms of deliverable independently targetable nuclear warheads (*right*). The SALT II treaty sets a ceiling of 2,250 on the total number of strategic launchers each side will be permitted through 1985. Within that total the U.S. is expected to maintain a rough balance among the three components of its strategic "triad": land-based intercontinental ballistic missiles (ICBM's), submarine-launched ballistic missiles (SLBM's) and long-range bombers. The U.S.S.R., in contrast, will presumably continue its traditional emphasis on land-based ICBM's. The contrast is particularly evident in the projected figures for numbers of warheads: whereas the U.S. will have only about 15 percent of its warheads on ICBM's in 1985, the U.S.S.R. will have close to 75 percent of its warheads deployed in this manner.

al readiness and protected from nuclear attack. A modern missile silo is designed to withstand a pressure of several thousand pounds per square inch over atmospheric pressure and to substantially attenuate the effects of thermal and nuclear radiation from a nuclear explosion. At present the probability that such a missile would survive a nuclear attack must be considered excellent. The survival probability depends in a direct (although not entirely predictable) manner on the characteristics of the attacking weapons, on the "hardness" of the silo and to a lesser degree on the characteristics of the targeted missile itself. (Another way to avoid the destruction of these missiles, of course, would be simply to launch them toward their predetermined targets on receipt of an early warning of a massive attack by the other country's ICBM force.)

Originally the concept of a "triad" of strategic weapons was adopted by the U.S. not only as a hedge against some unanticipated vulnerability of one of the three component systems but also as a way of giving both the Air Force and the Navy a "piece of the action" in strategic weaponry. Another purpose was to complicate an enemy's strategic-attack planning, since each weapons system has its own special combination of strengths and weaknesses. Gradually, however, two beliefs emerged that were eventually accepted as dogma by military and political leaders in the U.S. The first belief was that each component of the triad must by itself be capable of surviving a surprise attack in sufficient strength to enable it to inflict unacceptable retaliatory damage on the U.S.S.R. The second was that the land-based component of the triad was militarily and politically its most important part. The latter belief was supported by a number of advantageous features of land-based missiles, such as their exceptionally reliable command-and-control provisions, their rapid retargeting capability, the high reliability and superior accuracy of their warheads and their comparatively low operational cost. In spite of the many advantages of land-based missiles, the U.S. strategic forces are rather evenly distributed among the three components of the triad. In contrast the U.S.S.R., perhaps because of some of the features listed above, has traditionally emphasized the land-based component of its strategic triad.

Although both countries clearly attach major importance to their land-based ICBM's, their policies and actions in recent years have failed to give commensurate priority to maintaining the survivability of these forces. For example, the perceived vulnerability of the Minuteman force has resulted from a number of interrelated factors: the constant efforts of both countries to improve the accuracy of their ballistic mis-



EFFECT OF MIRVING (the deployment of missiles with multiple independently targetable reentry vehicles, or MIRV's) on the perceived vulnerability of fixed land-based ICBM's is indicated schematically in this pair of drawings. Before the advent of MIRVed missiles a preemptive attack by one country's ICBM force against the other country's ICBM force would have required the expenditure of at least one of the attacking country's missiles for each of the attacked country's missile silos (*top*). In this case the attacking country would have effectively disarmed itself without any certainty of having disarmed the attacked country; hence no advantage could be gained by attacking first. With a MIRVed missile force, however, one country could in principle devote only a fraction of its missiles to an attack against the other's missiles, expecting to destroy most if not all of them while retaining most of its own missiles safe in their silos (*bottom*). In this case it is conceivable some advantage could be gained by attacking first. Attacking MIRV's are shown cross-targeted, two to a silo.

siles, the development and deployment of multiple independently targetable reentry vehicles (MIRV's), first by the U.S. and soon afterward by the U.S.S.R., and the concurrent failure to ban MIRVed ICBM's in the bilateral strategic-arms-limitation talks (SALT).

In deciding to deploy MIRV's in the late 1960's the U.S. gave in to the temptation of what J. Robert Oppenheimer once called a "technically sweet" solution to a military problem. To the U.S. MIRV's offered an inexpensive but, as it is turning out, temporary military advantage over the U.S.S.R. Since a MIRVed ICBM carries more than one independently targetable nuclear warhead, it can attack more than one missile silo. Given a roughly equal number of missile silos in the U.S. and the U.S.S.R., it is evident that the side with a MIRVed ICBM force could devote only a fraction of its missiles to a surprise attack against the other side's missile silos and still expect (in principle at least) to destroy most if not all of them while retaining the bulk of its ICBM force safely in reserve.

One outcome of this turn of events is that a kind of nostalgia has developed among many military strategists in this country for the "good old days" of un-MIRVed ICBM's, when a surprise preemptive attack against the other side's missile silos would have required the expenditure of at least one of the attacker's missiles for each of the opponent's attacked silos, a ratio that made such a surprise countersilo attack unfavorable to the attacker and therefore improbable. Since it is difficult if not impossible to remedy the shortsightedness of the decision to deploy MIRVed missiles, however, the U.S. is left with the dilemma of having to choose between, on the one hand, the projected erosion of the survivability of the most highly valued component of its strategic triad and, on the other, the development and deployment of a new generation of expensive, somewhat arcane and ultimately counterproductive strategic weapons.

How is the vulnerability of an ICBM silo measured? If one makes the reasonable assumption that against such a structure the most destructive effect of a nuclear explosion is the sudden overpressure created by the blast wave, and if in addition one ignores numerous operational difficulties inherent in executing a coordinated, large-scale attack, then given the accuracy, yield and reliability of the attacking weapons and the hardness of the silos, it is possible to calculate a theoretical rate of destruction for silos subjected to the attack. The dominant variable that determines the destructiveness of the attack is the accuracy of the attacker's warheads. By convention the accuracy of a ballistic missile is expressed in terms of a quantity designated C.E.P. (for "circular error probable"), which is measured in nautical miles. The explosive yield of a nuclear warhead is expressed in equivalent tons of TNT, and the reliability is usually given as a percentage representing the probability that a given missile will function as intended.

C.E.P. is defined as the radius of the circle around a target within which half the warheads aimed at the target can be expected to land. The rated C.E.P. of any missile system is established in test flights, and it may or may not be duplicated in actual military operations. Furthermore, this measure of accuracy does not take into account the possibility of systematic aiming errors. The C.E.P. value for Russian missiles is presumably determined by U.S. intelligence experts from information gathered during Russian weapons tests by observing the launching of each missile and the return of its warhead (or warheads) to the surface. The U.S. Government, however, does not announce the results of its intelligence-gathering activities. As a result the C.E.P. values of the Russian ICBM's cited in unofficial public statements about the vulnerability of the Minuteman force may be neither the values measured directly by the Russians nor the values determined indirectly by the Americans. Since accuracy is the most important determinant of a missile's destructiveness against a silo, publicly quoted estimates of the C.E.P. values of Russian missiles are likely to be subject to powerful political pressures generated by highly motivated interests. The sensitivity of such estimates is easy to demonstrate: increasing the estimated C.E.P. of a Russian missile from .12 nautical mile to .25 would reduce its estimated lethality against a U.S. missile silo by more than half [*see illustration at left*].

Similar doubts may be held concerning the estimated yields of Russian warheads quoted in "off the record" public statements by U.S. officials. In general the recent public debate about the anticipated vulnerability of the Minuteman force simply overlooks the considerable uncertainty that surrounds such key variables in the calculation of missile-silo vulnerability.

The most authoritative estimate available of the actual Minuteman vulnerability is presented in a graph accompanying Secretary of Defense Harold Brown's defense-posture statement to Congress for 1980 [*see illustration on opposite page*]. The considerable range of uncertainty introduced into the calculation of the vulnerability of a silo by the uncertainties involved in the determination of the accuracy, yield and reliability of the Russian missiles, compounding the uncertainty inherent in predicting the actual response of a silo to a nuclear explosion, is correctly represented here by the broad band of probable error.

ACCURACY (C.E.P. IN NAUTICAL MILES)	YIELD (KILOTONS OF TNT)	RELIABILITY (PERCENT)	PROBABILITY OF DESTROYING A MINUTEMAN SILO BY ATTACKING WITH	
			ONE RUSSIAN WARHEAD	TWO RUSSIAN WARHEADS CROSS-TARGETED
.12	750	80	.64	.87
.12	750	66	.53	.78
.12	400	80	.51	.76
.12	400	66	.43	.68
.17	750	80	.44	.79
.17	750	66	.37	.61
.17	400	80	.33	.55
.17	400	66	.27	.47
.25	750	80	.24	.43
.25	750	66	.20	.36
.25	400	80	.17	.31
.25	400	66	.14	.27

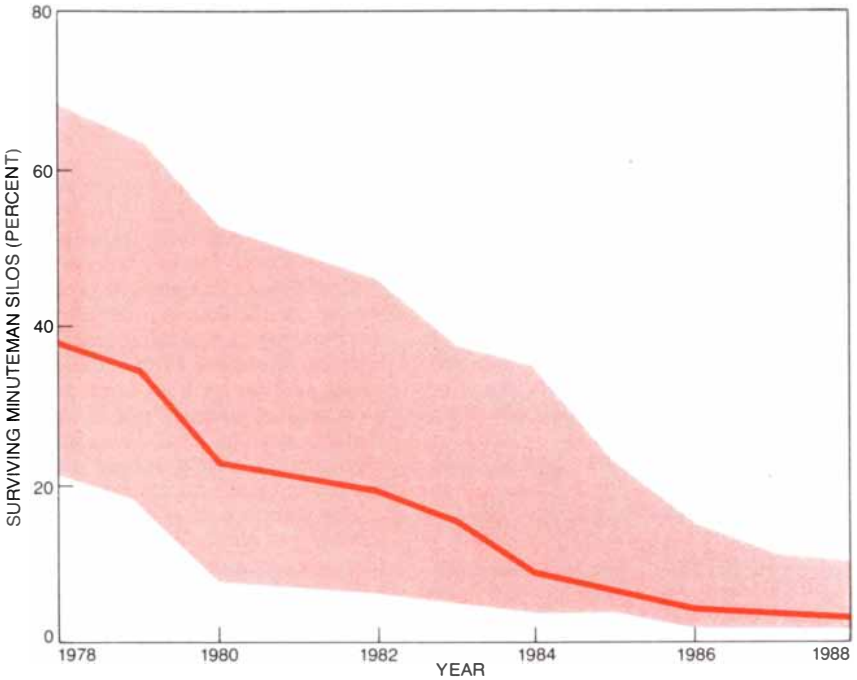
ESTIMATED DESTRUCTIVENESS of Russian MIRV's employed either individually or in cross-targeted pairs in a hypothetical surprise attack against a single hardened Minuteman silo depends on a number of factors, including the characteristics of the attacking weapons, listed in the first three columns at the left. The figures in the colored boxes are the "worst case" values for accuracy, explosive yield and reliability most often quoted in public discussions of the question of Minuteman vulnerability. The figures in the white boxes are different estimates of these values, based on somewhat less pessimistic assumptions about the capabilities of the Russian MIRV's. As the chart shows, the most important determinant of the lethality of such an attack is the accuracy of the attacking warhead (or warheads), which by convention is expressed in terms of the warhead's C.E.P., or "circular error probable," measured in nautical miles. (The C.E.P. of a reentry vehicle is defined as the radius of the circle around a target within which half of the warheads aimed at the target can be expected to fall.) The reentry vehicles of the two Russian MIRVed missiles, the SS-18 and the SS-19, are currently estimated by the U.S. to have a C.E.P. of .17 nautical mile; by 1985 they are expected to have a C.E.P. of .12 nautical mile. Increasing the estimated C.E.P. of reentry vehicles of either missile from .12 nautical mile to .25 would reduce expected lethality of attack by more than half (*last four rows of figures*).

Clearly arguments about the vulnerability of a given Minuteman silo are extremely sensitive to changes in the estimates of the characteristics of the Russian weapons.

What is more, attempts to calculate the vulnerability of the entire Minuteman force by merely extrapolating to all silos in that force the results of the calculation of the vulnerability of one silo can result in misleadingly inflated figures. This result is virtually inevitable not only because such simple calculations are flawed by the omission of numerous operational considerations and fail to take into account the many uncertainties inherent in the modeling of an attack by a single warhead against a single silo but also because they ignore the many additional uncertainties inherent in the complex operation of launching hundreds of missiles carrying thousands of warheads cross-targeted on more than 1,000 targets. Systematic departures from designed performance criteria are bound to happen. Although such an attack could be rehearsed, it could never be tested to determine such systematic errors. Hence any theoretical calculation of the level of damage that could be inflicted on an entire land-based missile force will surely warrant a lower degree of confidence than the calculation of single-silo vulnerability. Experienced military leaders are keenly aware of these uncertainties, which apply to the missile forces of both countries equally.

In short, we view assertions of the imminent vulnerability of the U.S. Minuteman force as premature, to say the least. A counterforce attack of the type visualized in current "worst case" scenarios is an immensely complicated and risky operation that cannot be described adequately by any mathematical model. Therefore it is wrong to treat predictions made on the basis of such models as pragmatic guides for action. Such predictions are based on calculations that are inherently approximate and tend to ignore many factors that, if they were taken fully into account, would in almost every instance tend to reduce the resulting vulnerability figures. Therefore the assertion that the U.S. Minuteman force can now or in the near future be destroyed with any degree of assurance by an all-out Russian counterforce attack seems to us to be a careless oversimplification, if not a deliberate exaggeration.

Nevertheless, there is little doubt that in the long run fixed land-based missiles will appear to become increasingly vulnerable to a MIRV attack. If this perception were to be transformed into a conviction in the minds of an opposing country's leaders, then under certain circumstances it is conceivable they might be persuaded to attempt such an attack. Conversely, if either side were to become convinced that its own ICBM force appeared vulnerable to the other side, then in a time of crisis the fear of a surprise attack by the opponent might provide a strong incentive to preempt the opponent by launching one's own missiles first. In both cases the appearance of vulnerability could be as dangerous as the vulnerability itself.



CONSIDERABLE UNCERTAINTY introduced into the calculation of the overall vulnerability of the Minuteman force by various uncertainties inherent in the determination of the accuracy, yield and reliability of the Russian MIRVed missiles is reflected in this graph, which is adapted from Secretary of Defense Harold Brown's defense-posture statement to Congress for 1980. The solid-color curve is Department of Defense's "best estimate" through 1988 of Minuteman "survivability" following a Russian attack. The light-color band represents range within which the estimate of Minuteman survivability has a "90 percent confidence level."

Under what circumstances would the appearance of Minuteman vulnerability be likely to tempt the Russian leaders to launch an attack against the U.S. ICBM force? Since most of the Russians' strategic nuclear warheads are carried by land-based ICBM's, and since in the absence of any new strategic system the Minuteman force is the only part of the U.S. strategic arsenal that is theoretically capable of attacking the Russian ICBM force with any chance of destroying it, one possibility is that the U.S.S.R. could decide to attack the Minuteman missiles preemptively in order to protect its own land-based missiles. It does not appear, however, that the Russians' ICBM force can realistically be considered to be threatened with extermination by an all-out Minuteman attack, because it can be shown that a substantial fraction of the Russian ICBM force would survive such an attack. Since there is no reason to believe the Russian leaders would feel that their land-based ICBM force was threatened by the Minuteman force, it follows that they would not be motivated by such a hypothetical threat to attack preemptively, even if they believed the Minuteman force was vulnerable.

There are several factors that would tend to minimize the likelihood of a surprise Russian attack intended to eliminate the Minuteman force. For one thing, the Russians could never be sure that the U.S. had not adopted a "launch on warning" policy for at least part of the Minuteman force. In addition there is little if any reason for them to believe the U.S. would not counterattack immediately against the U.S.S.R. with its surviving ICBM's, SLBM's and bombers. It would therefore be illogical for the Russians to attack only Minuteman silos; they would have to attack all three components of the U.S. strategic triad simultaneously, expecting to disable them all. This task, however, is impossible to carry out successfully because of the different time constraints of an attack against each of the three components, the varying degrees of vulnerability that each one has and the dependence of this vulnerability on the element of surprise.

Finally, in view of the severity of the outcome of any massive nuclear exchange, one would expect that prudent political leaders would join even a limited nuclear conflict reluctantly and would escalate it gradually, always with an eye to war-termination opportunities that would minimize their losses. It is difficult to escape the conclusion that given the current array of strategic sys-

tems on both sides, even if the military leaders of the U.S.S.R. were convinced that the Minuteman force was vulnerable, no responsible political leader in that country would be willing to risk the consequences of a preemptive, first-strike attack.

Nevertheless, the perception of Minuteman vulnerability, even if it is technically unfounded and irrational, may well generate political problems for the U.S. Government, both domestically and internationally. As a result, although there is no compelling military reason to move precipitately into measures designed to alleviate the perceived vulnerability of the Minuteman force, there are pressing political considerations pointing in that direction. The most pragmatic response, from both the economic and the strategic point of view, to the current problem would be to do nothing; it is not unreasonable, however, to explore alternatives that could offer some relief of the perceived Minuteman vulnerability.

In the remainder of this article we shall consider three such alternatives. The first would be to attempt to return to the "good old days" of unMIRVed ICBM's. The second would be to try to reduce the vulnerability of the Minuteman silos by applying existing defensive technology. The third would be to proceed with the construction of one or more entirely new missile systems designed to avoid the vulnerability problem.

The first option would require that both the U.S. and the U.S.S.R. abandon their large MIRVed ICBM's and construct a new land-based force consisting of comparatively small single-warhead missiles in superhardened silos. This approach would ensure that there would be no strategic advantage to attacking the other nation's land force preemptively, as long as the two forces remained approximately equal, since the attacker would have to expend at least one missile for each enemy missile attacked.

It is an attractive option, since it would make a surprise countersilo attack unprofitable and would reverse the strategic arms race. It suffers, however, from several shortcomings. First, eliminating MIRVed missiles could revive interest in anti-ballistic-missile (ABM) defenses, particularly the exotic space-based systems that have recently been speculated on. Second, it is only a matter of time before the new ICBM's in their silos would become vulnerable to very accurate sea-based warheads. In fact, with the aid of the "global positioning system" of satellites that the U.S. will probably have in full operation by the mid-1980's, warheads from SLBM's could be accurate enough to destroy fixed land-based missiles with a high degree of confidence. Although the operational difficulties of such a sea-based attack against an entire land-based ICBM force would be formidable, the land-based force of the U.S.S.R. could appear vulnerable to the MIRVed SLBM's of the U.S. as early as the late 1980's. The same would undoubtedly be true of U.S. land-based ICBM's some years later.

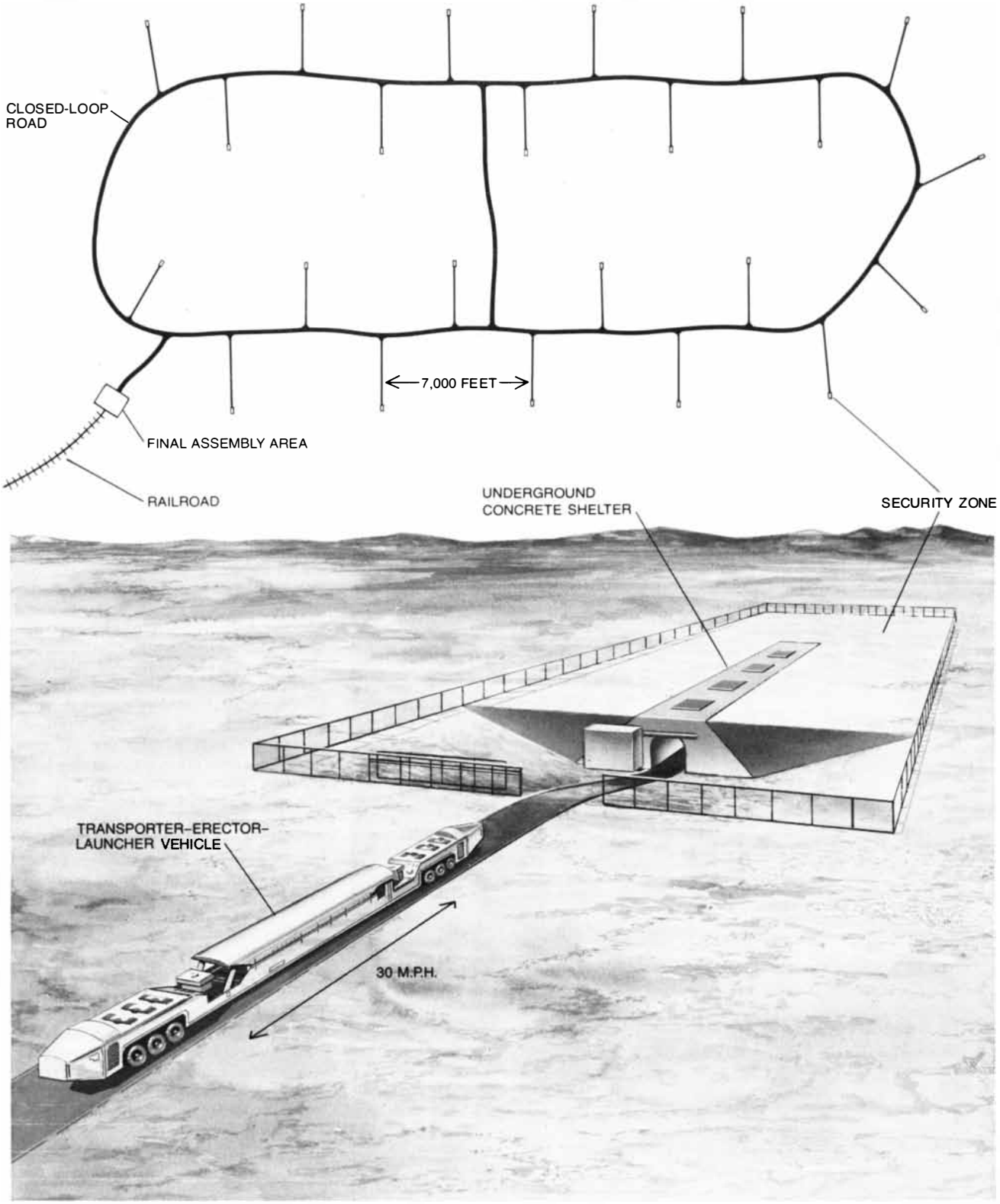
This "return to the good old days" option is one of several possible ways to seek strategic stability through arms-limitation negotiations. Indeed, it could be thought of as a stage in a process that would lead to the eventual elimination (perhaps by mutual agreement at some later stage of SALT) of all fixed land-based missiles on both sides. After all, nuclear disarmament, not just strategic stability, remains a valid long-term goal, even though it may appear utopian in the present political atmosphere.

The second option open to the U.S. would be to reduce considerably the actual and perceived vulnerability of the Minuteman force by installing local silo defenses, as was proposed a few years ago by Richard L. Garwin. Such defenses might consist of jamming devices to confuse the radar altimeters in the fusing mechanisms of the Russian reentry vehicle or radar-triggered buried nuclear explosives that would lift a curtain of debris into the path of the incoming warhead. Since dirt and dust take a long time to settle to the ground, such a curtain could protect the silo for up to an hour. One disadvantage of this scheme is that even if low-radiation nuclear explosives were used, it would still be politically unpalatable to detonate numerous nuclear devices on one's own soil and create a distinct, even if limited, radioactive-fallout hazard. To be sure, the existence of these defenses could make a Russian attack, which would create vastly larger quantities of fallout, much less probable. Nevertheless, local resistance to such a Minuteman defense scheme is likely to be considerable.

A variant of Garwin's silo-defense scheme would be to install two small upward-looking radars north of each silo in order to determine the approximate trajectory of an incoming warhead. Then, about a kilometer from the silo, the warhead would be met by tens of thousands of minirockets, each one weighing no more than about 200 grams, fired from hardened launchers. This protection system could ultimately be defeated by a combination of nuclear explosions near the radars and by the deployment of several warheads per silo. Since there are 1,000 Minuteman silos, however, the enemy would have to expend many thousands of warheads to penetrate the protective system and attack them. (The proposed system would not utilize the guided missiles and ABM radars prohibited by the ABM treaty; nevertheless, a separate agreement that would allow installation of the system might have to be negotiated between the U.S. and the U.S.S.R. so that it would not even appear to violate the treaty.)

The chief advantage of such silo-protection systems is that they could be deployed fairly quickly and therefore be in place by the mid-1980's, when the appearance of Minuteman vulnerability might otherwise become alarming. All the other policy options designed to counter this vulnerability cannot be expected to become effective before the early 1990's. The suggested protective devices could not protect the silos with absolute certainty, but they could decrease the expected number of destroyed silos to the point where the opponent would be forced to expend such a large portion of his counterforce warheads that the attack would be almost self-disarming. Furthermore, this approach would increase the uncertainty of the outcome of a countersilo attack and complicate it to such an extent that its planning would be made more difficult and the political decision to execute it would be made even more improbable. Such passive or quasi-passive defenses could be comparatively inexpensive. Since they would increase the safety of existing land-based ICBM's without threatening the assured-destruction potential of the opponent, they could reduce the fear of a surprise attack against Minuteman silos without interfering with the basic strategy of deterrence that has influenced the complexion of the strategic confrontation between the U.S. and the U.S.S.R. A weakness of this approach appears to be its lack of technological novelty, an aspect that seems to diminish its attractiveness to military planners.

The third option the U.S. could exercise to avoid the perceived vulnerability of the Minuteman force would be to construct an entirely new missile system and base it in such a way that it would be impervious to a surprise Russian attack. According to current plans, a new missile of this kind, known as MX ("missile experimental"), would be deployed; the missile would weigh about 95 tons, would carry 10 large and very accurate warheads and could be launched quickly and reliably. Its large throw weight, high accuracy and the contemplated number and size of its warheads are favored by those military and civilian defense analysts who believe the U.S. must have a countersilo capability comparable to that attributed to the Russians' large SS-18 missile. They believe a large new missile of this type would restore the strategic equality between the two countries that would be visibly altered after 1990, when the U.S.S.R. could have the perceived capability of destroying the Minuteman force. The main disadvantage of such a countersilo missile would be that no matter how it was based it would threaten the survivability of the land-based ICBM's of the U.S.S.R. It would there-



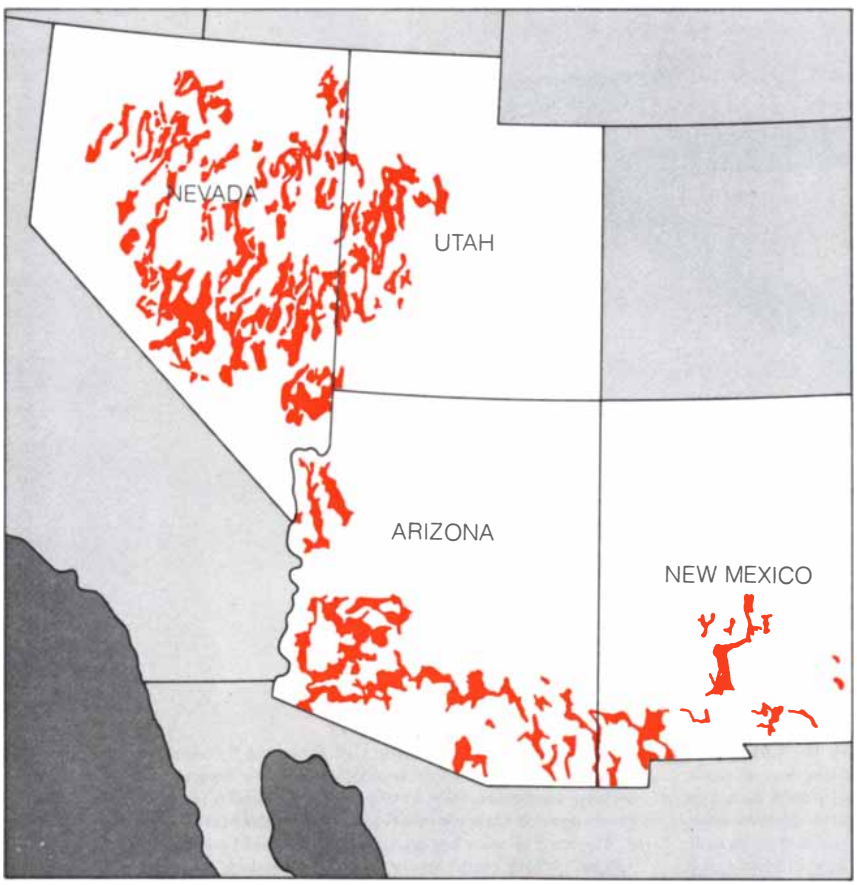
MOBILE BASING SCHEME currently favored by the Carter Administration for the MX missile calls for the construction of some 200 "closed loop" roads, along each of which there would be a system of spur roads leading to 23 horizontal protective shelters (*top*). Each missile would be assembled in a specially designated open area in full view of Russian reconnaissance satellites. The missile could then be carried around the road from shelter to shelter periodically by a 700,000-pound "transporter-erector-launcher vehicle" (*bottom*). In the event of an early warning of a Russian attack the vehicle would drive at a speed of about 30 miles per hour into one of the shelters, any of which could be reached during the 30 minutes it would take for a Russian ICBM to reach the U.S. Since the Russians could not know which shelter the missile would be in at the time of arrival of the attacking warheads, they would have to expend a large number of warheads against each mobile missile to be reasonably sure of destroying it. The roof of each horizontal shelter would be equipped with several "plugs," which could be removed at agreed-on intervals to enable the Russians to verify that there was only one MX missile per cluster of shelters. According to the authors' analysis a 200-missile MX force deployed in this way would be less vulnerable to a Russian counterforce attack than the 1,000-missile Minuteman force only if each shelter were superhardened and the shelters were built farther apart.

fore probably elicit from that country a response that could prove to be not in the ultimate interest of the U.S.

Several options have been proposed for the mode of deployment of this new missile that, in principle at least, would make it harder to target by Russian ICBM's. One option would be to conceal the missiles on vehicles circulating on tracks housed in underground tunnels. Because the hardness of the tunnel would be limited by economic factors, however, the missile would still be vulnerable; not only would the tunnel present a fairly soft, extended target to an attacker (in contrast to the "hard point" target presented by a Minuteman silo) but also by assigning widely separated aim points along the tunnel a massive Russian attack could avoid the "fratricide" effects that limit to two the number of warheads that can be effectively allocated against a hard-point target such as a silo.

Another possible land-basing mode for the MX would be to construct a system of approximately 25 concrete shelters connected by a road along which a missile could be transported rapidly from shelter to shelter in some random fashion, one such complex for each of 200 missiles. The idea here would be simply to make the number of targets a large multiple of the number of missiles. Of course, one could never be sure that an opponent would not be able at some time in the future to devise a technology, involving either reconnaissance satellites or clandestine ground observers, that would make it possible to detect which shelters housed a missile and which were empty at any one time, thereby vitiating the purpose of this basing mode. Accordingly either the rest of the shelters would have to be filled with dummy missiles indistinguishable in every detectable physical characteristic from the real missile or the transporter would have to be able to dash from one shelter to any other in the time available between the detection of a massive enemy attack and the arrival of the attacking warheads. (The latter scheme is reported to be the one currently favored by the Carter Administration.) Unless the U.S. were to agree with the U.S.S.R. beforehand on a mutually acceptable, common multiple-shelter basing mode, however, there is no guarantee that the U.S.S.R. would in turn deploy a multiple-shelter system that would meet U.S. verification requirements. If the U.S. were to proceed to deploy the MX missile in a basing system of its choice without prior agreement with the Russians, what would stop the Russians from deploying a similar missile in a basing system of their choice and then asserting that it is verifiable?

Given the obvious inadequacy of "national technical means" of inspection in such a situation, how long could any future arms-limitation agreement withstand the suspicions now prevailing in both the U.S. and the U.S.S.R. even toward agreements whose verifiability by national means can hardly be in any serious technical doubt? With or without dummy missiles, however, such a basing scheme would not offer long-term protection to the U.S. land-based force, since the U.S.S.R. could choose to attack all the shelters. Adding more shelters would probably be a losing game because the incremental cost of adding one warhead to the Russians' ICBM force would probably be less than that of adding a shelter to the proposed U.S. system.



POSSIBLE MX DEPLOYMENT SITES include large tracts of Federally owned desert-valley land in the southwestern U.S. (*colored areas*). According to current plans the loop roads would be open to the public. Each of 4,600 shelters would be fenced into a 2.5-acre security zone. Thousands of additional shelters would be built if perceived Russian threat were to increase.

The large area of such a horizontal shelter directly exposed to the overpressure from a nuclear explosion would result in hundreds of millions of pounds of load on the protective structure, even from the explosion of a 750-kiloton warhead as much as 500 meters away, calling for unusually strong and therefore extraordinarily expensive structures for reasonable protection. Unless the shelters were superhardened to withstand more than 1,000 pounds per square inch of pressure or were spread out over areas considerably larger than those contemplated in the schemes discussed publicly so far, it would be possible to target an attack so that one warhead could destroy more than one shelter. In the absence of any constraints on the number of warheads in the attacking nation's arsenal, any multiple-shelter system could ultimately be defeated. The U.S.S.R. would be constrained until 1985 by the SALT II agreement to about 7,000 land-based warheads. After that date, however, it would be free to deploy as many warheads as it wanted to unless a new strategic-arms-limitation agreement were to at least maintain this constraint. Hence the survivability of any multiple-shelter basing mode is completely dependent on the cooperation of the U.S.S.R. in future SALT negotiations.

In addition to the unfavorable protection-to-cost ratio the various multiple-shelter missile-basing modes would offer, they would suffer from several further disadvantages. First, their missiles would be more susceptible than the current Minuteman missiles to a "pin down" attack. (In other words, their launching could be impeded for a considerable time by the explosion of an enemy warhead in the atmosphere above the missile launcher, even though the explosion would not damage the

missile.) Second, these schemes would not start offering credible protection to the land-based strategic force of the U.S. until the number of existing missile shelters exceeded the very large number of warheads the U.S.S.R. would have available for a preemptive attack against them. That threshold, however, might not be reached before 1990, even if the ceilings established by SALT II were extended beyond 1985. Furthermore, since in case of war each cluster of shelters would require numerous ground-burst warheads for their destruction, they would introduce additional grave danger to the U.S. population by causing delayed radioactive fallout to the east of their sites. This effect might be mitigated by the decreased probability of a Russian countershelter attack, but only if these basing modes were to seem to the Russians to be virtually impossible to defeat. All the proposed multiple-shelter basing modes would have the additional environmental disadvantage of disrupting large areas in the proposed deployment regions in the southwestern U.S.

Consider also the opportunity such an approach would give the Russians for openly abrogating an agreement limiting the number of real missiles in their multiple-shelter system. Given a large number of available shelters, the Russians could manufacture and store clandestinely thousands of MIRVed missiles. Then they could suddenly abrogate the agreement and quickly deploy these missiles in the preexisting hardened shelters.

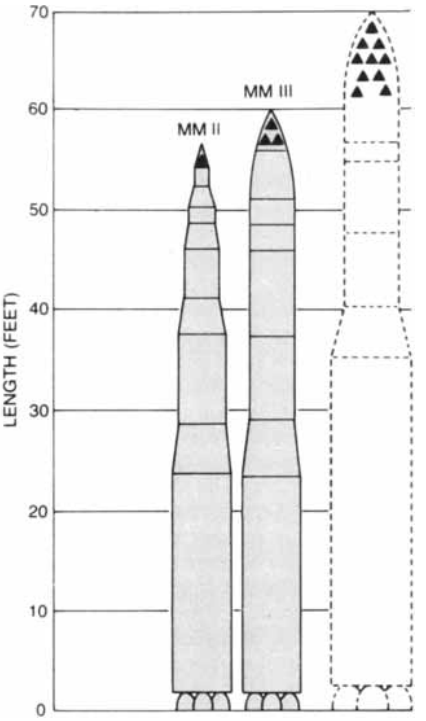
Such an augmentation of the Russians' land-based force would cost them only a small fraction of the expense of the entire force, since the protective shelters that would receive the missiles would already have been paid for. Hence a multiple-shelter basing scheme would militarily and economically tempt renunciation of numerical ceilings established in any future agreement between the U.S. and the U.S.S.R. Such a sudden abrogation of any future SALT agreement, accompanied by the overt deployment of thousands of additional warheads, would give the U.S.S.R. an enormous military advantage. Protected as these warheads would be in the preexisting shelters, they could be used to threaten violence to the U.S. without resort to a catastrophic nuclear war. Even if the U.S.S.R. did not resort to such chicanery, however, the security of the U.S. mobile force would still depend on the successful conclusion of future SALT agreements that constrained the number of Russian warheads.

In short, the deployment of very accurate MIRVed ballistic missiles appears to have permanently deprived land-based ICBM forces of the assurance of credible survivability they had in the past. The remaining choices for the U.S. are to learn how to live with land-based ICBM's that have some degree of vulnerability, to abandon land-based MIRVed missiles or to limit all new ICBM deployment to specially equipped aircraft and submarines.

Short-takeoff-and-landing (STOL) aircraft, for example, appear to offer better missile survivability than a multiple-shelter basing mode, given adequate warning of a Russian attack. This weapons system could consist of some 200 STOL aircraft, each carrying a single missile with between six and eight warheads. The system would presumably be designed to take advantage of the fact that the missile would be launched thousands of meters above the ground in order to make the ratio of the missile's payload to its gross weight greater than it is for ground-launched missiles. The planes could be deployed at about 100 bases, but they could be dispersed and moved at random among several thousand small airfields available around the country. On warning of a massive launch of Russian SLBM's the aircraft could escape to additional thousands of predetermined sites on the Federal highway network, making the preemptive targeting of all possible landing sites for the aircraft impossible.

Mobile ICBM's carried by transport aircraft offer several military advantages over a tunnel-based or multiple-shelter system. Their mobility, in conjunction with early-warning systems, quick alert and takeoff capabilities and several levels of basing, guarantees very high survivability, even in the case of an all-out barrage of SLBM's. The aircraft could be launched on warning of attack, but the missiles on board would not have to be fired toward their targets immediately. Although the missile that might have to be used in conjunction with an airborne basing mode would carry no more than eight MIRV's, its accuracy could be as good as that of a shelter-based missile. Finally, since an airborne system would be truly mobile, it would not tempt an enemy to acquire an unlimited number of warheads in the hope of making it vulnerable. Therefore the survivability of this system would depend neither on Russian adherence to agreed-on numerical ceilings on missiles and warheads nor on the outcome of future SALT negotiations.

In spite of the many advantages of an air-launched ICBM system, it would not be a short-term solution to the problem of the perceived vulnerability of land-based ICBM's. Certainly a decade and perhaps more would be necessary to develop, test and deploy such a system. The weakness of the system would be that in times of crisis hundreds of nuclear weapons would be circulating in the air over the country and would be moved from airfield to airfield with little physical security, exposing the weapons both to accidents and to possible sab-



	MINUTEMAN	MX
NUMBER OF MISSILES	450 (MM II) 550 (MM III)	200
BASING MODE	FIXED (SILO)	MOBILE (MULTIPLE SHELTERS)
NUMBER OF PROTECTIVE STRUCTURES PER MISSILE	1	23
TOTAL NUMBER OF PROTECTIVE STRUCTURES	1,000	4,600
NUMBER OF NUCLEAR WARHEADS PER MISSILE	1 (MM II) 3 (MM III)	10
TOTAL NUMBER OF WARHEADS	450 (MM II) 1,650 (MM III)	2,000
YIELD PER WARHEAD (EQUIVALENT KILOTONS OF TNT)	1,200 (MM II) 335 (MM III)	335
ACCURACY (C.E.P. IN NAUTICAL MILES)	.2 (MM II) .12 (MM III)	.05

SPECIFICATIONS for the proposed MX mobile ICBM call for a substantially larger missile than the two existing Minuteman ICBM's. The MX would be capable of carrying a payload more than three times greater than that of either of the other missiles, and hence it could deliver a much larger number of nuclear warheads. The accuracy of the MX, estimated here in terms of C.E.P. in nautical miles, would also be considerably improved.

otage. Moreover, the system would require an extensive and well-orchestrated (therefore often rehearsed) support network that could prove both expensive and awkward.

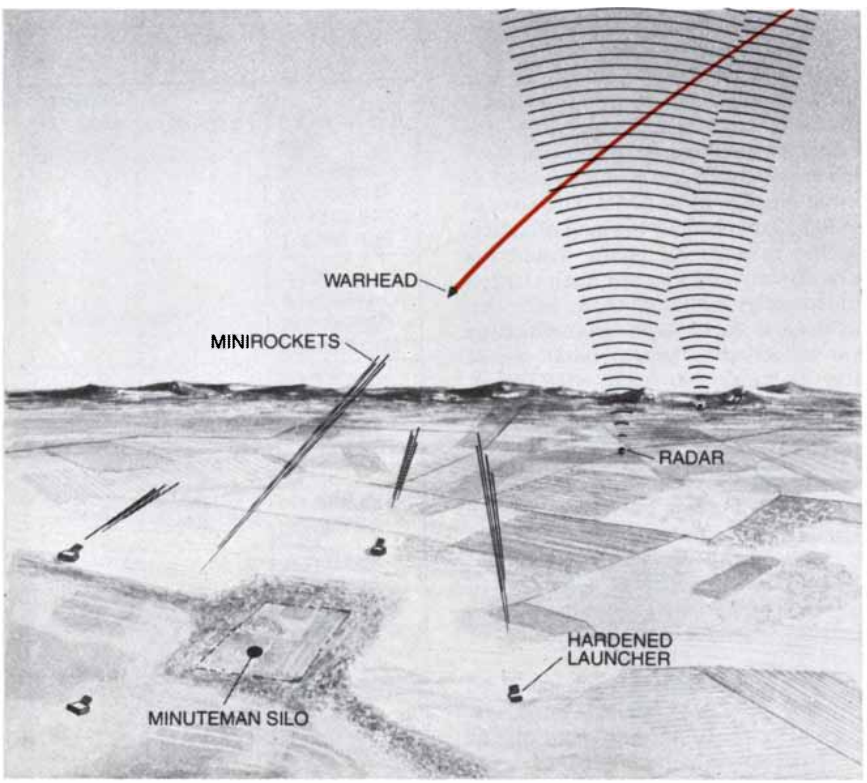
Another basing option, proposed in Congressional testimony by Garwin and Sidney D. Drell, would involve small submarines that could carry a few missiles each in canisters outside a pressurized hull; the submarines, which could be powered by inexpensive non-nuclear power plants capable of generating speeds of a few knots, would loiter over the continental shelf. Since the total area of the ocean over the U.S. continental shelf is 400,000 square miles, a saturation attack against such vehicles is practically impossible. The submersibles could be counted by national technical means of verification, and therefore they would not present any arms-limitation difficulties or open the way for unpleasant military surprises, such as a sudden abrogation of the SALT II treaty by the U.S.S.R. induced by the inability to verify their numbers confidently. The proposed submersibles could, for example, carry two missiles each, have an on-station endurance of a month and be operated by a crew of fewer than 20. Their survivability would not depend on any warning and would not be threatened by preemptive attack or by any Russian antisubmarine-warfare activities, since they would be deployed in areas of the seas firmly controlled by U.S. surface ships and submarines. Although at any one time some of these submersibles would be in port for crew changes and maintenance and therefore be vulnerable to a surprise Russian attack, the short distances the small submersibles would have to travel to reach their battle stations would make it possible to have about 80 percent of them on station (and therefore impervious to preemptive attack) at any given time. The accuracy of the missiles launched from such a system could be guaranteed to be as good as that of land-launched ICBM's.

Such a coastal submarine-based system could not be ready for deployment in the next decade or so. Therefore this system shares with the air-mobile system and the multiple-shelter basing mode the disadvantage that its unavailability for many years might mandate the need for intermediate remedial measures (such as local defenses for Minuteman silos), which would add to the overall cost of restoring invulnerability to the ICBM component of the U.S. strategic triad.

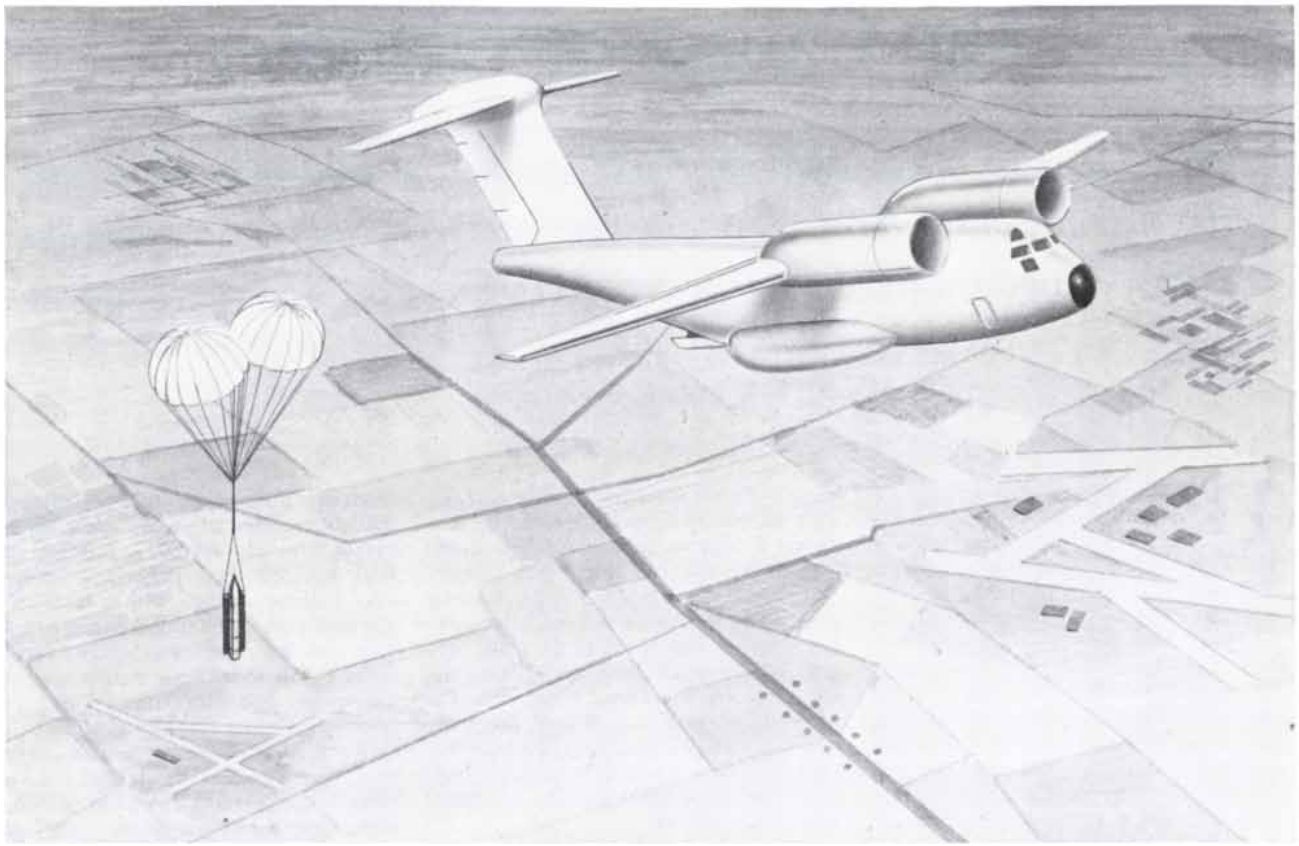
In sum, much as one might like to advocate a return to the "good old days" of unMIRVed land-based missiles, both as an immediate solution to the perceived problem of Minuteman vulnerability and as a step in the right direction toward limiting the strategic nuclear arsenals of the U.S. and the U.S.S.R., this option appears remote in time and distant as a political and military reality. On the other hand, the U.S. need not rush into a new land-based ICBM system out of fear of the immediate vulnerability of the Minuteman force. Alarmist public descriptions of this vulnerability are exaggerated and misleading because they ignore the inherent uncertainties of a forcewide attack and are based on the assumption of a strategic scenario that is implausible on both political and military grounds. The actual vulnerability of the Minuteman force is very much open to question. Our own view is that the most likely outcome of even a concerted Russian countersilo attack during the late 1980's would be the survival of a significant fraction of the Minuteman force.

More important, the land-based Minuteman force carries only about a third of the total equivalent megatonnage of the U.S. nuclear strategic arsenal; the other two-thirds is carried by the SLBM force and the bomber force. Since only a small fraction of the total U.S. equivalent megatonnage is assigned to industrial and other economic targets in the U.S.S.R. as a part of the strategy of deterrence, even if in the course of an unexpectedly successful Russian attack the U.S. were deprived of all its land-based ICBM's, it would still be left with about half of its strategic nuclear force available for retaliatory attacks against Russian military targets.

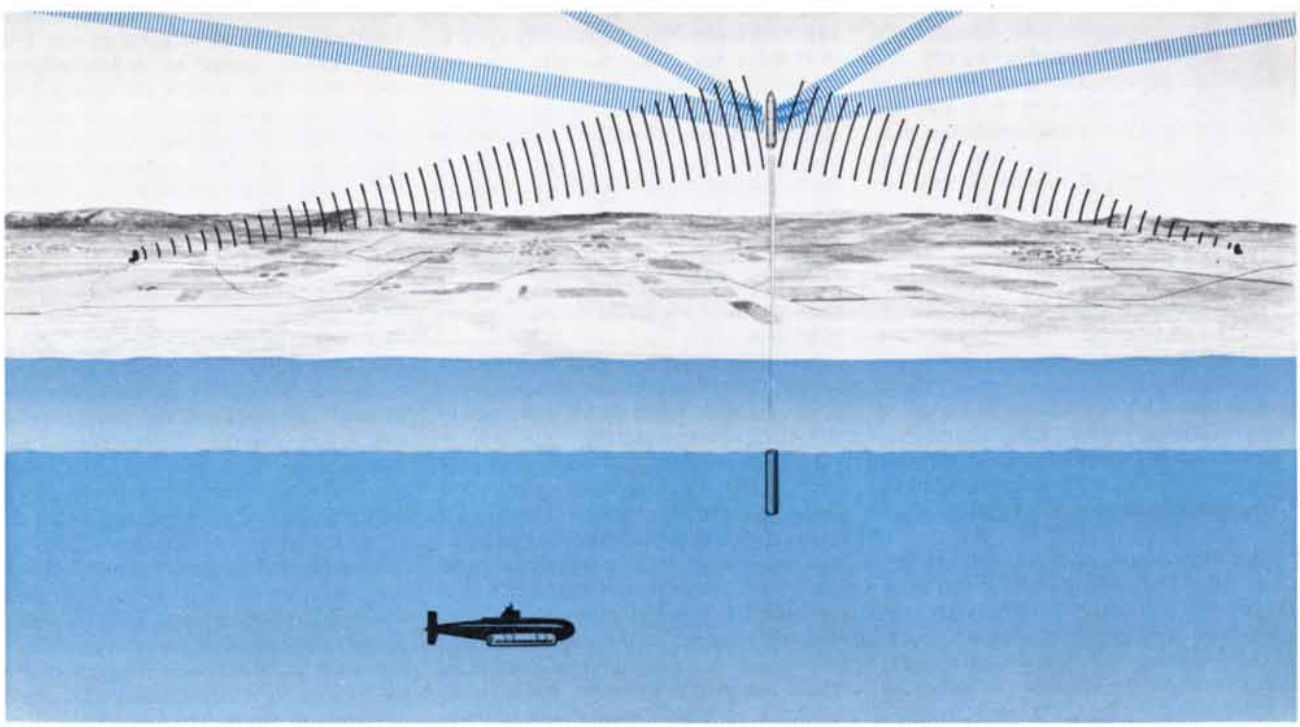
The installation of local silo defenses, after prior discussions with the U.S.S.R. aimed at maintaining the ABM treaty unchallenged, would increase the actual and perceived security of the Minuteman force, decrease crisis instability and minimize still further the probability of a large-scale counterforce nuclear exchange. In such a strategic environment the U.S. and the U.S.S.R. could then attempt in future SALT negotiations to return as far as possible to the "good old days" of unMIRVed ICBM's and perhaps even go beyond that. If, on the other hand, such negotiations were to appear unpromising, the U.S. would still have plenty of time to develop other strategic systems, such as an underwater ICBM force deployed along the continental shelf of the U.S., that would offer a better solution than the land-based MX to the perceived problem of Minuteman vulnerability.



LOCAL SILO DEFENSES, such as the system illustrated here, could be installed fairly quickly, applying existing technology, to reduce the actual and the perceived vulnerability of the Minuteman force to a Russian counterforce attack in the mid-1980's. This particular scheme, originally proposed by Richard L. Garwin, calls for the installation of two small upward-looking radars a few kilometers north of each silo to determine the approximate trajectory of an incoming warhead. Hardened rocket launchers situated about a kilometer from the silo could then fire tens of thousands of minirockets at the warhead in order to disable it. This comparatively inexpensive alternative to the mobile MX missile would have a number of military and political advantages, not least of which is that it would not upset the stability of the present strategic balance by threatening the deterrent capability of the Russian land-based ICBM force.



AIRBORNE ICBM's have been proposed as an alternative mobile basing mode to the multiple-shelter scheme. A system of this type could entail the deployment of some 200 short-takeoff-and-landing (STOL) aircraft, each carrying one missile, at some 100 major airfields. Given adequate warning of a Russian attack, the aircraft could escape to any of several thousand smaller airfields or even to predetermined sites on the Federal highway system. Such a system would appear to be much less vulnerable than the land-mobile MX system.



UNDERSEA ICBM'S carried by a fleet of small, non-nuclear submarines deployed over the U.S. continental shelf would provide another alternative to the land-based MX. The submarines could carry two missiles each in canisters outside their pressurized hull. Since the submarines would be deployed in areas of the seas firmly controlled by the U.S., they would be essentially invulnerable. With the aid of a ground-based or satellite-based "global positioning system" the accuracy of the missiles launched from such a system could be as good as that of land-launched ICBM's. Neither of the systems depicted on this page could be ready for deployment in less than a decade.

The Causes of Diabetes

There are two major types of diabetes: the maturity-onset form and the juvenile-onset form. Juvenile-onset diabetes appears to develop from a complex interaction between genetic makeup and environment

by Abner Louis Notkins

Diabetes mellitus and its complications are now thought to be the third leading cause of death in the U.S., trailing only cardiovascular disease and cancer. According to a report issued by the National Commission on Diabetes in 1976, as many as 10 million Americans, or close to 5 percent of the population, may have diabetes, and the incidence is increasing yearly. The direct and indirect effects of diabetes on the U.S. economy are enormous, exceeding $5 billion per year. If current trends continue, the average American born today will have better than one chance in five of ultimately developing the disease. The likelihood of becoming diabetic appears to double with each decade of life and with every 20 percent of excess body weight.

Moreover, although the acute and often lethal symptoms of diabetes can be controlled with insulin therapy, the long-term complications of the disease reduce life expectancy by as much as a third. Compared with nondiabetics, diabetics show a rate of blindness 25 times higher, of kidney disease 17 times higher, of gangrene five times higher and of heart disease twice as high.

Many aspects of diabetes remain mysterious, but recent work in three seemingly unrelated fields—genetics, immunology and virology—has supported the contention that diabetes is a heterogeneous group of diseases rather than a single one. This work has also indicated that diabetes arises from a complex interaction between the genetic constitution of the individual and specific environmental factors.

Diabetes mellitus is an ancient disease. The earliest description of its symptoms is found in the Ebers papyrus of Egypt, dating back to 1500 B.C. In the second century A.D. Aretaeus of Cappadocia named the disease diabetes, the Greek word meaning "to flow through a siphon." "Diabetes," he wrote, "is a strange disease that consists in the flesh and bones running together into the urine." This was an imaginative description of the striking symptoms of diabetes: a copious flow of urine accompanied by extreme thirst and hunger, but nonetheless resulting in the wasting away of both muscle and fat, often ending in coma and death. In the sixth century Indian physicians recognized that the urine from diabetic patients had a sweet taste. It was not until the 18th century, however, that the sweet-tasting substance was identified as the sugar glucose and the word mellitus, or "honeyed," was added.

One of the first clues to the pathology underlying diabetes came in 1889, when Oscar Minkowski and Baron Joseph von Mering, working in Strasbourg, sought to determine whether the pancreas was essential to life. By careful surgical procedures they removed the pancreas from dogs. A probably apocryphal story has it that the day after the operation the caretaker in the laboratory noticed that the dogs' urine attracted an unusual number of flies. In any event Minkowski and von Mering analyzed the urine and found in it high levels of glucose, indicating that removal of the pancreas gave rise to a syndrome resembling diabetes. This finding strongly implied that the pancreas was secreting a substance that regulated the metabolism of glucose.

In 1909 the hypothetical antidiabetic substance secreted by the pancreas was given a name: insulin. All attempts to alleviate experimental diabetes by feeding pancreatectomized dogs raw pancreas or by injecting them with crude pancreatic extracts, however, were unsuccessful. It is now known that such experiments were bound to fail because insulin is a protein, and that when it is given orally, it is destroyed by protein-cleaving enzymes in the gastrointestinal tract and by similar enzymes present in crude pancreatic extracts. As a result definitive proof of the existence of insulin was not forthcoming until two Canadian investigators, Frederick G. Banting and Charles H. Best, extracted insulin from dog pancreases that they had previously depleted of protein-cleaving enzymes. On July 30, 1921, Banting and Best injected their pancreatic extract into a diabetic dog. Within hours the glucose level in the blood began to fall. News of the experiment spread rapidly, and in a short time insulin was being widely and successfully employed to treat the acute symptoms of diabetes mellitus in human beings. The discovery of insulin was hailed as a cure for diabetes because it lowered blood-glucose levels, controlled the acute symptoms of the disease and prevented the death from coma that sometimes came within days after the onset of symptoms.

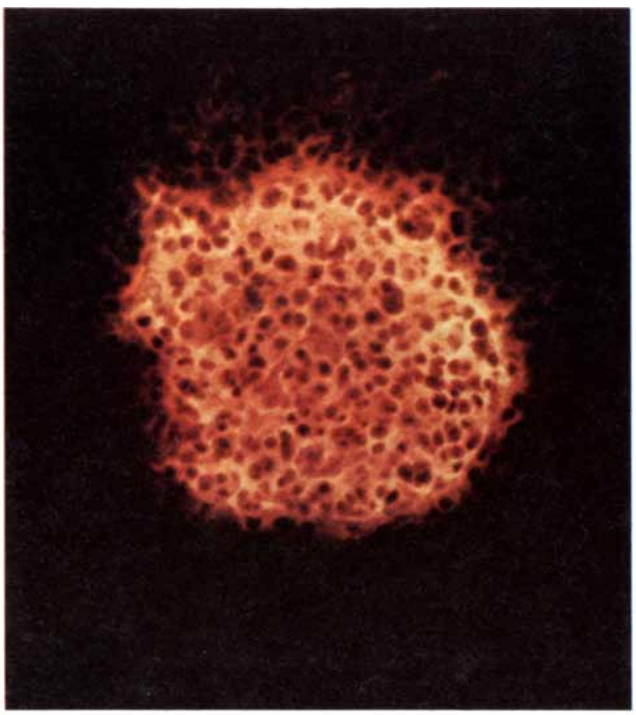
Evidence that all was not well only became apparent years later. Diabetics who had been on insulin for a long time were found to have an unusually high incidence of heart attacks, stroke, kidney failure, gangrene and blindness. Disorders of the nerves, the skin and the mouth were also common, and particularly serious complications arose during pregnancy. Insulin treatment thus controlled the early symptoms of diabetes but not the development of long-term complications.

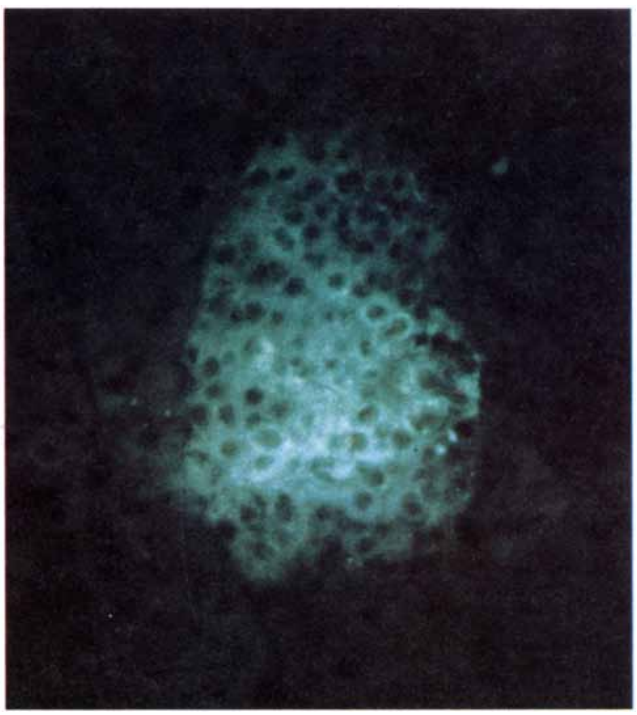
What causes diabetes and its complications? Since the time of Banting and Best much has been learned about insulin and how it controls glucose levels in the blood. The pancreas has two distinct components: the acinar cells, which manufacture digestive enzymes and secrete them into the duodenum (the first segment of the small intestine), and the islets of Langerhans, which secrete a variety of hormones into the bloodstream. The pancreas has between one and two million islets, each islet about 200 microns in diameter and together accounting for about 2 percent of the mass of the organ. The islets are highly vascularized, and each islet cell is in close proximity to a capillary.

There are at least four different types of cells in each islet. The alpha cells, which make up about 20 percent of a

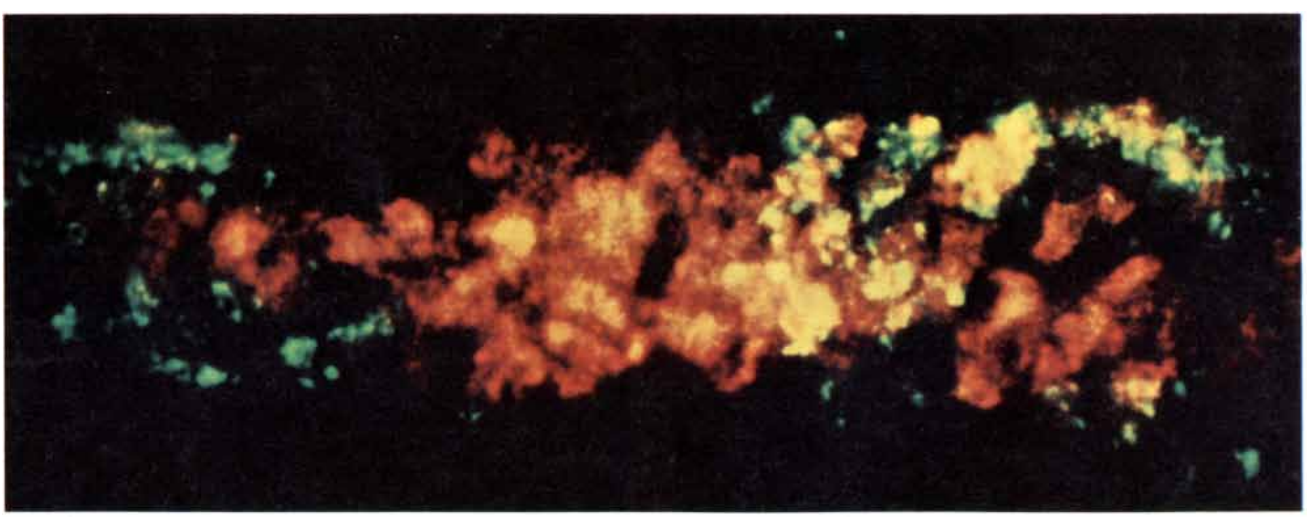
typical islet, secrete the hormone glucagon. The beta cells, which make up about 75 percent of the islet, secrete insulin. Insulin and glucagon act in different ways; whereas the secretion of insulin lowers the level of glucose in the blood, the secretion of glucagon raises it. The delta cells secrete the hormone somatostatin, which inhibits the secretion of both insulin and glucagon, and the PP cells secrete pancreatic polypeptide hormone, the function of which is not yet clear. Although the concentration of glucose in the blood is maintained at a more or less constant level by the actions of insulin, both glucagon and somatostatin play an important modulating role.

In normal individuals the concentration of glucose is usually less than 115 milligrams per 100 milliliters of plasma,

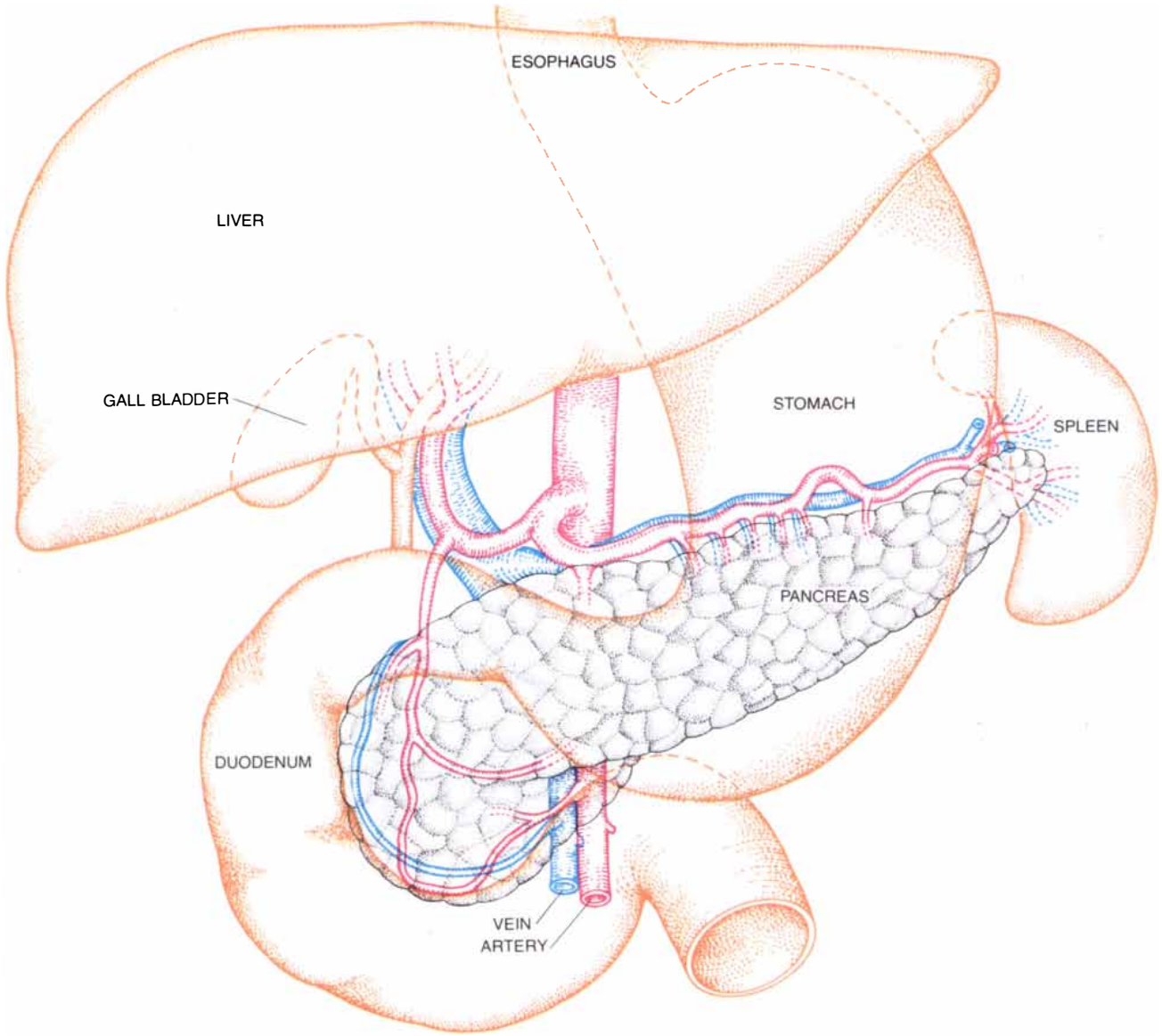


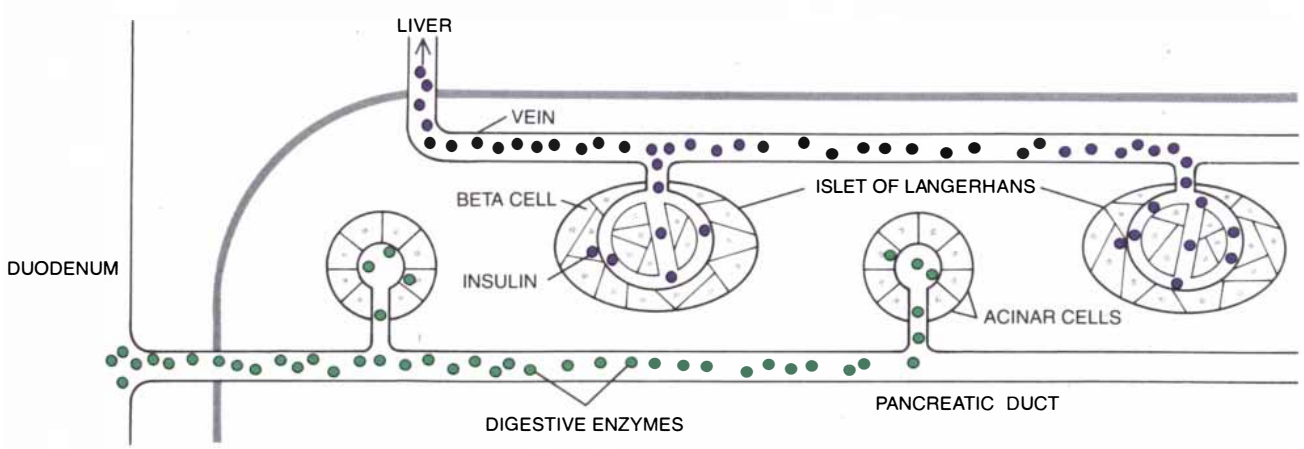


INSULIN-CONTAINING BETA CELLS within an islet of Langerhans in the pancreas glow orange in the fluorescence micrograph at the left. They were made visible by treating a section of human pancreas with an antibody specific to insulin and labeled with rhodamine dye, which glows orange when it is illuminated with ultraviolet. The dark spot within each fluorescent cell is the cell nucleus; the surrounding acinar cells of the pancreas do not contain insulin and hence appear black. The micrograph at the right is of a section of pancreas from a mouse infected three days earlier with encephalomyocarditis (EMC) virus. It was stained with antibody to the virus that was labeled with fluorescein dye, which fluoresces green when it is illuminated with ultraviolet. Only the cells within the islet of Langerhans are infected; the surrounding acinar cells are not infected and appear dark. Such experiments have shown that viruses can destroy beta cells and induce diabetes in experimental animals. The micrographs were made by A. Bennett Jenson and Kozaburo Hayashi in author's laboratory.



VIRUS-INFECTED BETA CELLS are revealed in this immunofluorescence micrograph. A section of pancreas from a mouse infected with reovirus was stained with two antibodies: a rhodamine-labeled antibody to insulin and a fluorescein-labeled antibody to the virus. Depending on the filter employed in the fluorescence microscope, the beta cells could be made to glow orange and the virus-infected cells to glow green. In the double exposure shown here the fluorescent colors mix so that some of the virus-infected beta cells appear yellow.





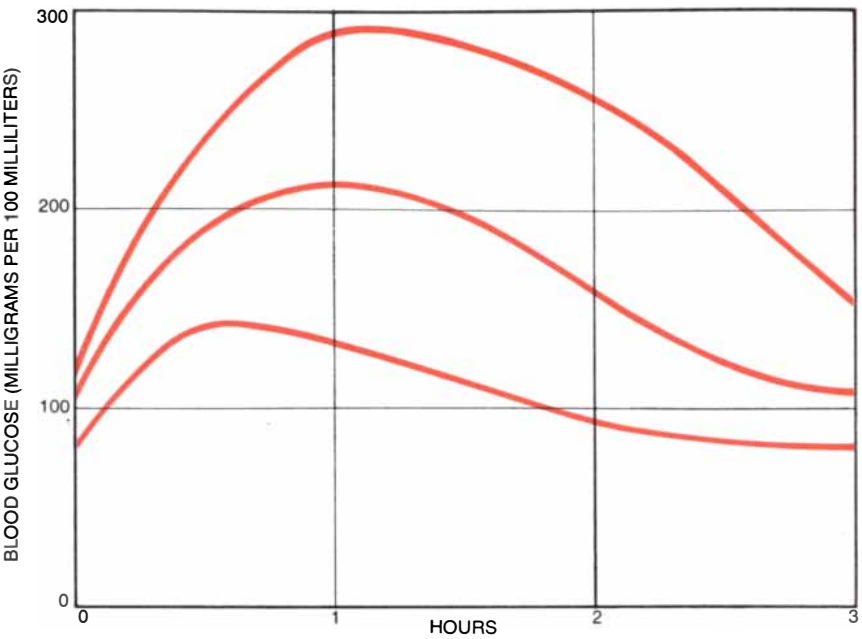
PANCREAS is situated in the abdominal cavity just below the liver and under the stomach. It is bordered on one side by the duodenum (the first segment of the small intestine) and on the other side by the spleen. The pancreas is made up of two functionally distinct components: the acinar cells and the islets of Langerhans, as is shown in the schematic diagram. The acinar cells, which make up the bulk of the pancreas, manufacture digestive enzymes that enter the duodenum by way of the pancreatic duct. The islets of Langerhans represent only 1 to 2 percent of the total mass of the pancreas and secrete several hormones (such as somatostatin and glucagon) in addition to insulin. These hormones reach the bloodstream by way of numerous small veins in the islets of Langerhans that drain into the liver.

but in diabetics it is much higher and in severe cases may reach 1,000 milligrams. Because diabetics have a particularly difficult time removing excess glucose from the blood following the ingestion of carbohydrates they can be easily diagnosed by means of a glucose-tolerance test. A standard load of glucose is given by mouth, and the blood-glucose concentration is measured periodically over the next few hours. In the normal individual the blood glucose returns to base-line levels within two hours. In the diabetic, however, the blood-glucose values rise much higher and take longer to return to the base-line levels.

After the ingestion of a meal the beta cells of a normal individual respond to the resulting rise in blood-glucose levels by secreting more insulin into the bloodstream. The insulin travels to the various organs of the body, interacting with specific receptors on the surface of the target cells. The binding of the hormone to its receptors initiates a series of events within the cells that results in the increased uptake of glucose into the cells, where it is converted into metabolic energy or stored as glycogen (animal starch) and fat.

It is obvious that pathological processes intervening anywhere along this pathway (such as in the pancreas, the blood or the peripheral tissues) could result in abnormal glucose metabolism. For example, the glucose in the blood would be elevated if the beta cells of the pancreas did not manufacture enough insulin, or if there were antagonists to insulin in the bloodstream, or if the peripheral tissues of the body did not respond properly to the action of insulin.

The causes of the long-term complications of diabetes are even more perplexing. One of the many complications is the thickening of the basement membrane that surrounds the wall of capillaries. This thickening is believed to contribute to poor peripheral circulation and to be at least partially responsible for the fact that many diabetics suffer from disorders of more than one organ. A number of hypotheses have been put forward to explain the pathological processes that give rise to complications. One hypothesis holds that in the diabetic there is a premature aging of cells, so that the diminution of function that might be expected late in life comes much earlier. Indeed, changes characteristic of aging have been observed when cells from diabetics are cultured in laboratory glassware, and it has been suggested that such changes might be genetically controlled. Another hypothesis suggests that specific metabolic intermediates of glucose, such as sorbitol, accumulate to high concentrations in tissues such as nerves and the lens of the eye. This accumulation could lead to changes in osmotic pressure, causing the cells to swell and thereby damaging the tissue.



GLUCOSE-TOLERANCE TEST is a common method for diagnosing diabetes in individuals whose fasting glucose is not unequivocally elevated. Seventy-five grams of glucose are administered orally and changes in blood-glucose levels are monitored over several hours. Bottom curve represents the response of a normal individual. Middle curve shows the response of an individual who has an impaired glucose tolerance but is not considered truly diabetic. Top curve represents the response of a diabetic and shows that the blood-glucose level remains elevated.

A hypothesis that is currently receiving considerable attention contends that high concentrations of glucose in the blood cause glucose molecules to form chemical bonds with the amino groups of cellular proteins, the reaction known as glycosylation. This process was first recognized when it was found that in the blood of diabetics glycosylated forms of hemoglobin are unusually common. Many investigators are now beginning to believe proteins in other tissues, such as the eye, the nerves and the blood vessels, may become glycosylated to a much greater degree in diabetics than in nondiabetics. The critical question is whether these changes and others caused by high concentrations of glucose are genuinely responsible for the thickening of the basement membrane of the blood vessels and the other long-term complications of diabetes. The question is not merely academic; if a link between high levels of blood glucose and complications is firmly established, then the most prudent course would be strict therapeutic measures to keep glucose levels close to the normal base line.

The different and often conflicting findings about the nature of diabetes and its long-term complications have led many to conclude that it is not a single disease but rather a heterogeneous group of diseases, all of which ultimately lead to an elevation of glucose in the blood. Indeed, two major types of diabetes have been distinguished on the basis of clinical evidence: the maturity-onset type and the juvenile-onset type. Long-term complications can develop in both types, although there is considerable variation among individuals.

Maturity-onset diabetes is the much more prevalent type, representing more than 90 percent of all the cases. It most often occurs in people who are over 40 and overweight. The onset is slow, and pathological changes in the pancreas are not always apparent. Moreover, the clinical manifestations of maturity-onset diabetes are often mild, and the high glucose levels in the blood can usually be controlled by diet alone. Although diabetes is traditionally viewed as a condition caused by a deficiency of insulin, many maturity-onset diabetics have a sufficiency or even a surplus of the hormone in the blood. For these individuals diabetes arises not from a shortage of insulin but probably from defects in the molecular machinery that mediates the action of insulin on its target cells. Maturity-onset diabetes is accordingly referred to as non-insulin-dependent diabetes.

Again a number of hypotheses have been put forward to explain maturity-onset diabetes, and there is little unanimity in the field. One of the most intriguing hypotheses comes from the work of Jesse Roth, C. Ronald Kahn

and their colleagues at the National Institute of Arthritis, Metabolism, and Digestive Diseases. By measuring the binding to target cells of radioactively labeled insulin molecules they showed that the number of insulin receptors was decreased in obese patients with maturity-onset diabetes. When the patients were put on a weight-reducing diet, however, the number of insulin receptors returned to normal. Roth and Kahn contend that the increased food intake associated with obesity leads initially to the secretion of an excessive amount of insulin into the circulation. The secreted insulin then acts through a negative-feedback process of some kind to reduce the number of insulin receptors on the target cells. This decrease in receptors is thought to make the cells less responsive to insulin and hence less capable of utilizing glucose. Other investigators propose, however, that the primary defect actually arises within the target cells after insulin has bound to the receptors. Regardless of the mechanism, most would agree that the proper control of diet and body weight is of major importance in treating this form of diabetes.

Juvenile-onset diabetes is much less common than the maturity-onset type, representing well under 10 percent of all the cases. It usually develops in people younger than 20, and its onset is more abrupt. The disease is characterized by a marked decline in the number of beta cells in the pancreas (often to less than 10 percent of normal), leading to a deficiency of insulin and an elevation of glucose in the blood. The deficiency of insulin accelerates the breakdown of the body's reserve of fat, resulting in the production of ketones and organic acids. These metabolites lower the *p*H of the blood, producing a condition known as diabetic ketoacidosis that can result in death. Because injections of insulin are required to regulate the level of blood glucose, this form of diabetes, which is generally more severe, is referred to as insulin-dependent diabetes.

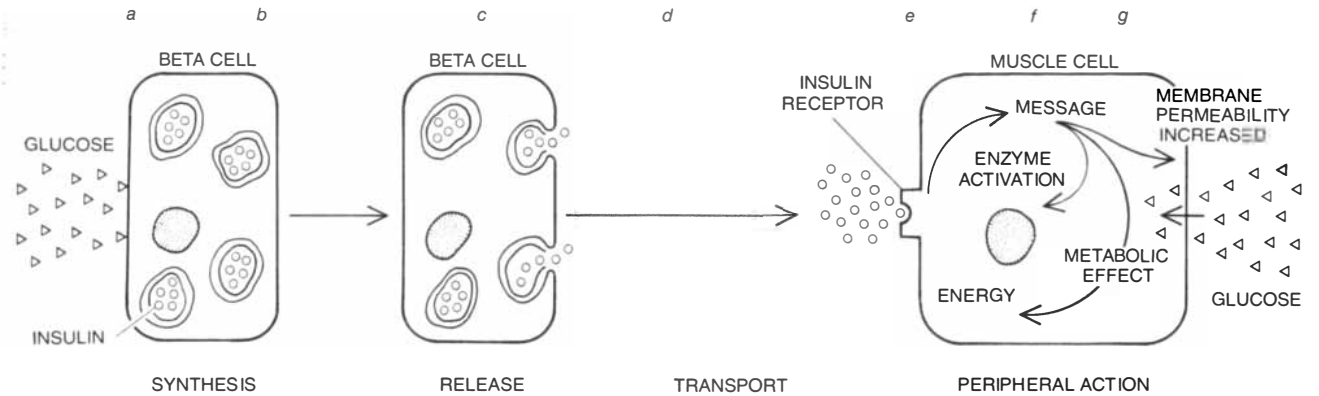
Although there is no general agreement on the cause of juvenile-onset diabetes, the reduction in the number of beta cells and the decline in insulin levels long ago suggested that the primary deficit is at the level of the beta cell. Over the past few years new information has emerged on some of the factors that might cause beta-cell damage. The new leads concerning the relation between genetic and environmental factors are what I shall now discuss in depth.

It has been known for some time that maturity-onset diabetes tends to run in families and that the likelihood of an individual's developing the disease is increased if one parent is diabetic and further increased if both parents are. In contrast, the likelihood of developing juvenile-onset diabetes is not substantially increased in the offspring of diabetic parents. The precise risk of developing the different forms of diabetes has been difficult to establish, however, because of the variety of the diagnostic criteria and because such variables as diet, obesity, age, sex and ethnic background are not always taken into account. Moreover, since family members share the same diet and environment, a high incidence of diabetes in a given family does not necessarily prove that genetic factors are involved.

In the hope of distinguishing between genetic and environmental factors in diabetes geneticists began some 40 years ago to study identical twins: offspring derived from the same egg who share the same genes. If diabetes were caused solely by inheritance, then if one identical twin developed the disease, the other twin would be expected to develop it. The degree of genetic involvement in diabetes can therefore be estimated from the degree of concordance (both twins developing diabetes) as opposed to discordance (only one twin developing diabetes).

In the early 1970's David A. Pyke and his colleagues at King's College Hospital in London reported their findings on more than 100 pairs of identical twins, the largest series of twins ever studied. They found that when one twin of a pair developed diabetes after age 50, the other twin developed the disease within several years in almost every case. If one twin developed the disease before age 40, however, the other twin developed it in only half of the cases. Strikingly, the majority of the twins over 50 had non-insulin-dependent diabetes, whereas most of those under 40 had insulin-dependent diabetes. Pyke's findings generated considerable excitement because they demonstrated that genetic factors are predominant in maturity-onset diabetes but that additional factors, presumably environmental, are needed to trigger juvenile-onset diabetes.

A different approach to the genetics of diabetes has been to investigate the relation between diabetes and the histocompatibility antigens. These antigens—proteins on the surface of all body cells with nuclei—are responsible for the fact that tissue transplanted from one individual to an unrelated individual will be recognized as foreign and will be rejected by the recipient's immune system. In man the histocompatibility antigens are referred to as the HLA system. The genes coding for the HLA antigens are on chromosome No. 6 and occupy four loci along the chromosome designated *A, B, C* and *D*. The genes at a given locus are not always the same, and the different forms of each gene are referred to as alleles. An individual can have two different alleles at a given locus, one allele



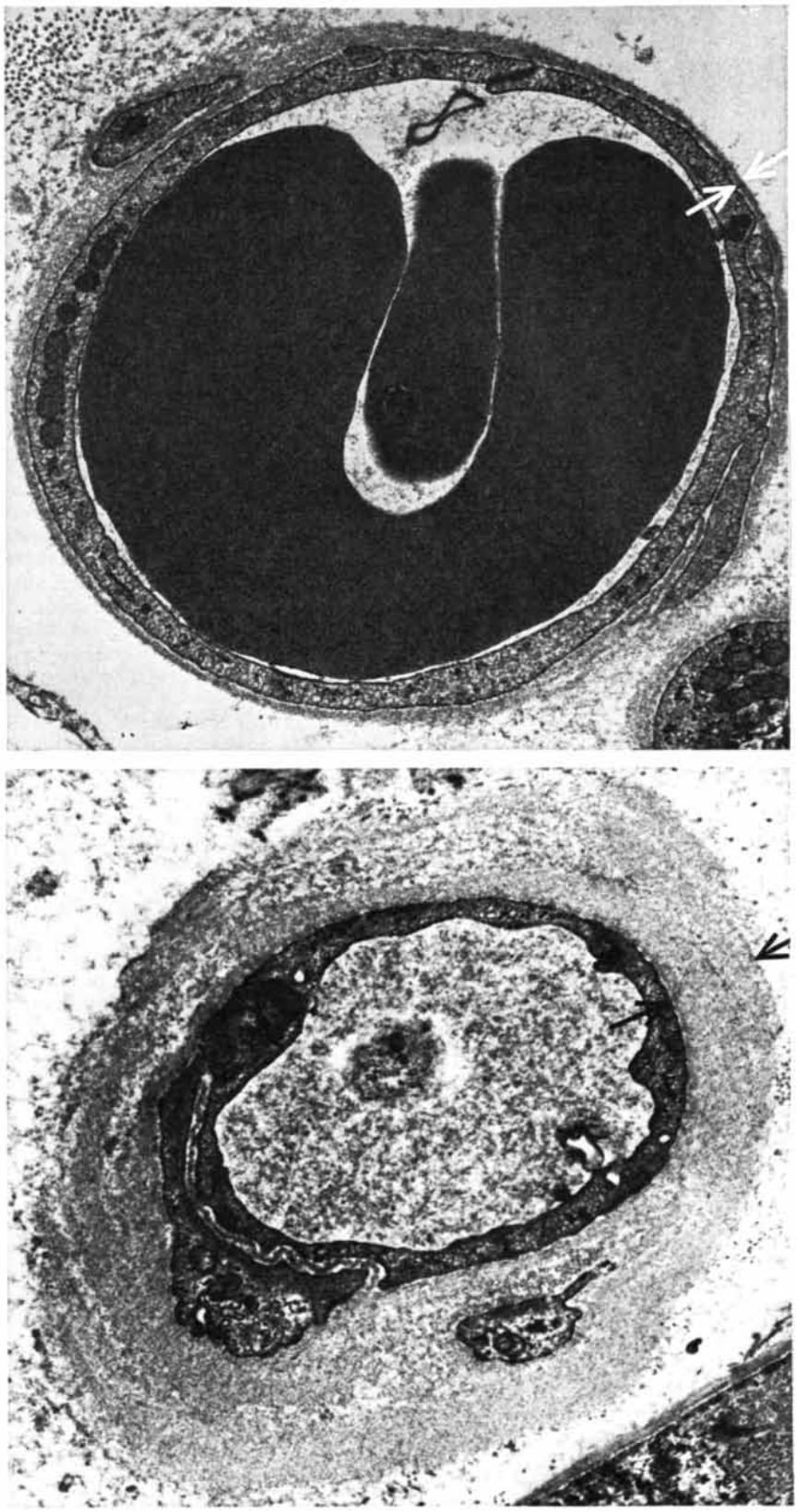
ACTIONS OF INSULIN are summarized. The rise in blood glucose associated with a carbohydrate meal induces the beta cells in the islets of Langerhans to secrete insulin into the circulation. The insulin is then carried in the bloodstream to target cells throughout the body, where it binds to receptor molecules on the cell surface. This interaction triggers a series of events inside the cells that enhances the uptake of glucose from the blood and its subsequent breakdown for metabolic energy or storage as glycogen (animal starch) and fat. A defect anywhere along this pathway could result in diabetes. Possible causes include destruction of beta cells (*a*), abnormal synthesis of insulin (*b*), retarded release of insulin (*c*), inactivation of insulin in the bloodstream by antibodies or other blocking agents (*d*), altered insulin receptors or a decreased number of receptors on peripheral cells (*e*), defective processing of the insulin message within the target cells (*f*) and abnormal metabolism of glucose (*g*). Current evidence points to the beta cell as site of the primary defect in juvenile-onset diabetes.

contributed by each parent. In the HLA system both alleles are expressed as cell-surface proteins, which can be identified by laboratory tests and can serve as the basis for the typing of tissues for transplantation.

A few years after tissue typing became a common procedure a completely unexpected discovery was made. Certain of the HLA antigens were found at unusually high frequency in patients with specific diseases. For example, the risk of developing the deformity of the spine known as ankylosing spondylitis was found to be 100 times greater in individuals carrying the HLA antigen *B*27. This observation spurred investigators in many parts of the world to look for an association between HLA antigens and other diseases, including diabetes. Jørn Nerup and his associates at the Steno Memorial Hospital in Copenhagen found that the HLA antigens *B*8 and *B*15 were two to three times commoner in diabetics than they were in nondiabetics. When the Danish workers further analyzed their data, they made an interesting discovery. The increased frequency of these HLA antigens was associated only with juvenile-onset diabetes; there was no change in the frequency of the antigens associated with the maturity-onset disease. It soon became apparent that there was an even stronger association between juvenile-onset diabetes and HLA antigens at the *D* locus. Moreover, when more than one high-risk allele was present in the same individual, the likelihood of developing juvenile-onset diabetes was increased by as much as tenfold.

Further support for an association between the HLA system and diabetes has come from the work of A. G. Cudworth of St. Bartholomew's Hospital in London, Jose Barbosa of the University of Minnesota Medical School and Pablo Rubinstein of the New York Blood Center, all of whom looked at families in which two or more siblings had juvenile-onset diabetes. When they compared diabetic and nondiabetic siblings in the same family, they found that the siblings with diabetes had inherited identical groups of HLA alleles in a significantly greater percentage of cases than the siblings without diabetes. Taken together with the fact that several different high-risk alleles have been identified, it now appears that one or more genes in close proximity to the HLA complex along chromosome No. 6 actually may be the important determinants of juvenile-onset diabetes. Recently, however, the situation has been complicated by the finding that certain HLA alleles are associated with a significant decline in the incidence of juvenile-onset diabetes, implying protective genes exist as well.

How genes might act to cause diabetes or prevent it is far from clear. One clue comes from the work of Hugh O. McDevitt of the Stanford University



THICKENING OF BASEMENT MEMBRANE that surrounds the blood vessels is one of the long-term complications of diabetes. The electron micrograph at the top shows a capillary from a normal male; the basement membrane (*between arrows*) is a thin sheath surrounding the epithelial cells of the vessel. (The dark structures inside the capillary are red blood cells.) The electron micrograph at bottom shows a capillary from a man with diabetes of 26 years' duration. Note that the basement membrane is markedly thickened. This change is thought to contribute to poor peripheral circulation, damage to the retina of the eye and an acceleration of the atherosclerotic process. The micrographs, the magnification of which is some 23,000 diameters, were supplied by Joseph R. Williamson of Washington University School of Medicine.

School of Medicine and Baruj Benacerraf of the Harvard Medical School. They found that specific genes in the mouse control the immune-response (Ir) genes and that these genes are in the same region of the chromosome as the genes for the major histocompatibility antigens of the mouse. On the basis of this evidence it is conceivable that genetically controlled differences in the immune response influence the development of diabetes. For example, the high-risk alleles associated with the HLA complex might code for a deficient immune response to agents that preferentially attack beta cells, thereby allowing beta-cell damage and diabetes to result. Conversely, the protective alleles might enhance the host's immune response to such invaders.

There are other ways the immune response might influence the development of diabetes. Under certain circumstances the host can turn against itself and react immunologically to its own proteins and so damage its own tissues; the phenomenon is known as autoimmunity. Some five years ago G. Franco Bottazzo of the Middlesex Hospital Medical School in London, Richard Lendrum of St. Mary's Hospital in London and James C. Irvine of the Royal Infirmary in Edinburgh discovered that serum from newly diagnosed juvenile-onset diabetics contains an antibody that reacts with the alpha, beta and delta cells in the islets of Langerhans from normal, nondiabetic people. The fraction of juvenile-onset diabetics who possess this autoantibody is as high as 85 percent at the time of diagnosis but decreases to less than 25 percent after two years. Patients with maturity-onset diabetes, however, rarely possess the islet-cell antibody.

It is tempting to speculate that the high-risk HLA alleles associated with juvenile-onset diabetes influence the development of the islet-cell antibody. What, however, triggers the production of the antibody and what role does it play in causing diabetes? Some investigators believe an imbalance among the cells of the immune system is responsible for the development of islet-cell antibody. Others contend that the antibody represents an immune response to components of the islet cells that have been altered by viruses or toxic chemicals. Still others argue that the islet-cell antibody has little to do with the basic disease process, since it is present in the serum of some individuals who do not show any sign of diabetes.

The controversy may be resolved by some forthcoming experiments. Åke Lernmark and his colleagues at the University of Chicago recently demonstrated that the islet-cell antibody can react with antigens on the surface of beta cells grown in culture. These workers plan to exploit the cell-culture system to determine whether the islet-cell antibody can actually destroy beta cells. Other investigators are pursuing the possibility that the critical role in beta-cell damage is played not by antibody but by white blood cells such as lymphocytes and macrophages. Regardless of whether autoimmunity is an important cause of beta-cell damage, the islet-cell antibody appears to be a valuable marker for identifying juvenile-onset diabetes and differentiating it from maturity-onset diabetes.

Both the genetic and the immunological studies I have mentioned raised the possibility that an environmental agent such as a virus might trigger juvenile-onset diabetes, perhaps by inducing an autoimmune reaction. In 1965 Willy Gepts of the University of Brussels examined the pancreas of a number of juvenile-onset diabetics who had died shortly after the appearance of their disease. Gepts found in the islets of Langerhans of many of the patients white blood cells typical of what one might expect in response to infection or an autoimmune reaction. Additional support for an infectious process in diabetes comes from the observation that juvenile-onset diabetes often begins abruptly and that the incidence of the disease is higher in fall and winter, when infectious diseases are more prevalent.

FEATURES	JUVENILE-ONSET (INSULIN-DEPENDENT)	MATURITY-ONSET (NON-INSULIN-DEPENDENT)
AGE AT ONSET	USUALLY UNDER 20	USUALLY OVER 40
PROPORTION OF ALL DIABETICS	LESS THAN 10 PERCENT	GREATER THAN 90 PERCENT
SEASONAL TREND	FALL AND WINTER	NONE
APPEARANCE OF SYMPTOMS	ACUTE OR SUBACUTE	SLOW
METABOLIC KETOACIDOSIS	FREQUENT	RARE
OBESITY AT ONSET	UNCOMMON	COMMON
BETA CELLS	DECREASED	VARIABLE
INSULIN	DECREASED	VARIABLE
INFLAMMATORY CELLS IN ISLETS	PRESENT INITIALLY	ABSENT
FAMILY HISTORY OF DIABETES	UNCOMMON	COMMON
HLA ASSOCIATION	YES	NO
ANTIBODY TO ISLET CELLS	YES	NO

DIFFERENCES between juvenile-onset and maturity-onset diabetes are summarized. The maturity-onset form is commoner and not as severe. It does not require injections of insulin.

The possibility that the infectious agent might be a virus goes back to the turn of the century, when H. F. Harris, a Philadelphia physician, observed that one of his patients developed diabetes shortly after having had mumps. Since that time there have been scattered reports documenting a temporal relation between the onset of certain viral infections and the subsequent development of diabetes. The virus most often implicated has been that of mumps, but there is still no firm evidence that the relation between mumps and diabetes is anything more than a chance association. Indeed, if the mumps virus does infect beta cells and cause diabetes in human beings, it must do so under very special circumstances. In such cases a rare strain of mumps virus may be involved, or the individual who develops diabetes may have an unusual and possibly genetically determined susceptibility to the virus.

Evidence that another common virus, that of rubella ("German measles"), might be a rare cause of diabetes comes from the more recent work of Margaret Menser, Jill Forrest and their colleagues in Australia. They found that in children and adults who had contracted rubella while still in the uterus the incidence of diabetes was considerably higher than normal. Because congenital rubella is notorious for giving rise to a variety of malformations, however, the cause of diabetes in these individuals is far from clear.

Probably the best evidence that viruses may be involved in the genesis of diabetes has come from studies with experimental animals. In 1968 John E. Craighead, who is now at the University of Vermont College of Medicine, was studying a variant of encephalomyocarditis (EMC) virus. This virus causes encephalitis and myocarditis in mice and occasional febrile illness in human beings. Craighead found that many of the mice infected with the EMC variant developed diabetes. When he examined the pancreas of the infected animals, he noticed inflammatory white cells in the islets of Langerhans and found that many of the beta cells were damaged.

Proof that the virus was actually responsible for these changes in the islets came from experiments conducted by Kozaburo Hayashi, a visiting fellow in my laboratory at the National Institutes of Health. Hayashi prepared an antibody to EMC virus and labeled it with fluorescein dye, which fluoresces a brilliant green when it is irradiated with ultraviolet. He removed the pancreas from the infected animals and incubated sections of the organ with the labeled antibody. The antibody bound only to cells containing EMC virus, so that he was able to identify the infected cells by their fluorescence. His experiments showed that EMC virus had indeed infected beta cells and that the replication of the virus in these cells was primarily responsible for their destruction.

Subsequent work in Craighead's laboratory and mine revealed that the initial effect of the virus-induced destruction of beta cells is the release of large amounts of insulin into the circulation, which lowers the level of glucose in the blood. Within a few days, however, the animal's insulin reserve is depleted, in some cases to less than 10 percent of normal, and the glucose level rises. Many of the animals now begin to excrete glucose in their urine and to consume increasing amounts of water and food, the symptoms of juvenile-onset diabetes.

The investigation of the involvement of viruses in the genesis of diabetes took a new turn when it became apparent that only certain inbred strains of mice develop the disease. Three postdoctoral fellows in my laboratory, Wark Boucher, Michael Ross and Takashi Onodera, investigated this phenomenon in detail. They infected 24 different inbred strains of mice with EMC virus and divided the strains into three groups on the basis of their glucose response. The first group, designated "susceptible," responded to the infection by manifesting elevated levels of glucose in the blood. The second group, designated "glucose-intolerant," showed abnormally high glucose levels after infection only when a large glucose load was administered. The third group, designated "resistant," showed no sign of diabetes after infection.

Breeding experiments revealed that the differences among strains of mice in susceptibility to EMC-induced diabetes are genetically controlled. When mice from two susceptible strains were mated, their offspring were also susceptible. Similarly, the offspring of two resistant strains were resistant. A cross between a susceptible mouse and a resistant one yielded a first generation that was resistant to the development of virus-induced diabetes. When the first-generation mice were mated with one another, however, many of their offspring were susceptible, indicating that susceptibility to virus-induced diabetes had been inherited as a recessive trait. A backcross experiment was also performed: when the resistant first-generation mice were crossed with the resistant parental strain, none of the offspring developed diabetes. When the resistant first-generation mice were crossed with the susceptible parent strain, however, approximately half of the offspring developed diabetes, suggesting that a single gene locus plays a major role in controlling susceptibility to EMC-induced diabetes.

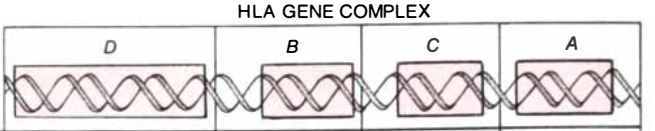
As often happens in science, these studies answered some questions and raised many others. Ji-Won Yoon, a staff scientist in my laboratory, and Onodera wanted to know how inheritance actually influences which mice develop diabetes. One possibility was that certain genes control the susceptibility of beta cells to EMC infection. The more susceptible the cells are, the more of them would be infected and the greater would be the likelihood that the mouse would develop diabetes. To see if this was the case, sections of pancreas from infected mice were stained with fluorescent antibody to EMC virus. Counts of infected cells showed that islets from susceptible strains of mice (such as *SJR*) contained approximately 10 times more infected beta cells than islets from diabetes-resistant strains (such as *C57BL*).

These experiments did not prove, however, that differences among strains of mice in susceptibility to the virus resided at the level of the beta cell; it was possible that the differences were at the level of the immune response. In order to differentiate between these alternatives we grew beta cells from different strains of mice in tissue culture and then examined their susceptibility to infection. In this way we were able to circumvent the immune response of the mice. When cultures highly enriched in beta cells were infected with EMC virus, we found that the beta cells from susceptible strains were destroyed more readily than those from resistant strains.

A clue to what makes the beta cells from different strains of mice differentially susceptible to EMC infection has been provided by recent work in my laboratory. In 1959 John J. Holland, who is now at the University of California at San Diego, showed that viruses would not attach to and infect certain cells unless those cells possessed viral "receptors": cell-surface proteins of some kind that the virus recognizes and binds to before infecting the cell. Recently Ruben Chairez, working as a visiting scientist in my laboratory, found that two to three times more EMC virus bound to beta cells from diabetes-susceptible mice than bound to beta cells from diabetes-resistant mice. This finding suggests that the degree of susceptibility to the virus might be a function of the number or type of viral receptors on the surface of the beta cells. These experiments must be interpreted with caution: it is still not possible to obtain pure cultures of beta cells, and the number of viral receptors may change when the cells are grown in culture. Nevertheless, the viral-receptor experiments are a beginning and, if they are confirmed, they will provide a plausible explanation of why only animals with the "right" genetic makeup develop virus-induced diabetes. It is certainly intriguing to speculate that similar factors might operate in human beings, and that the high-risk alleles associated with the HLA complex might control the susceptibility of human beta cells to viral infection.

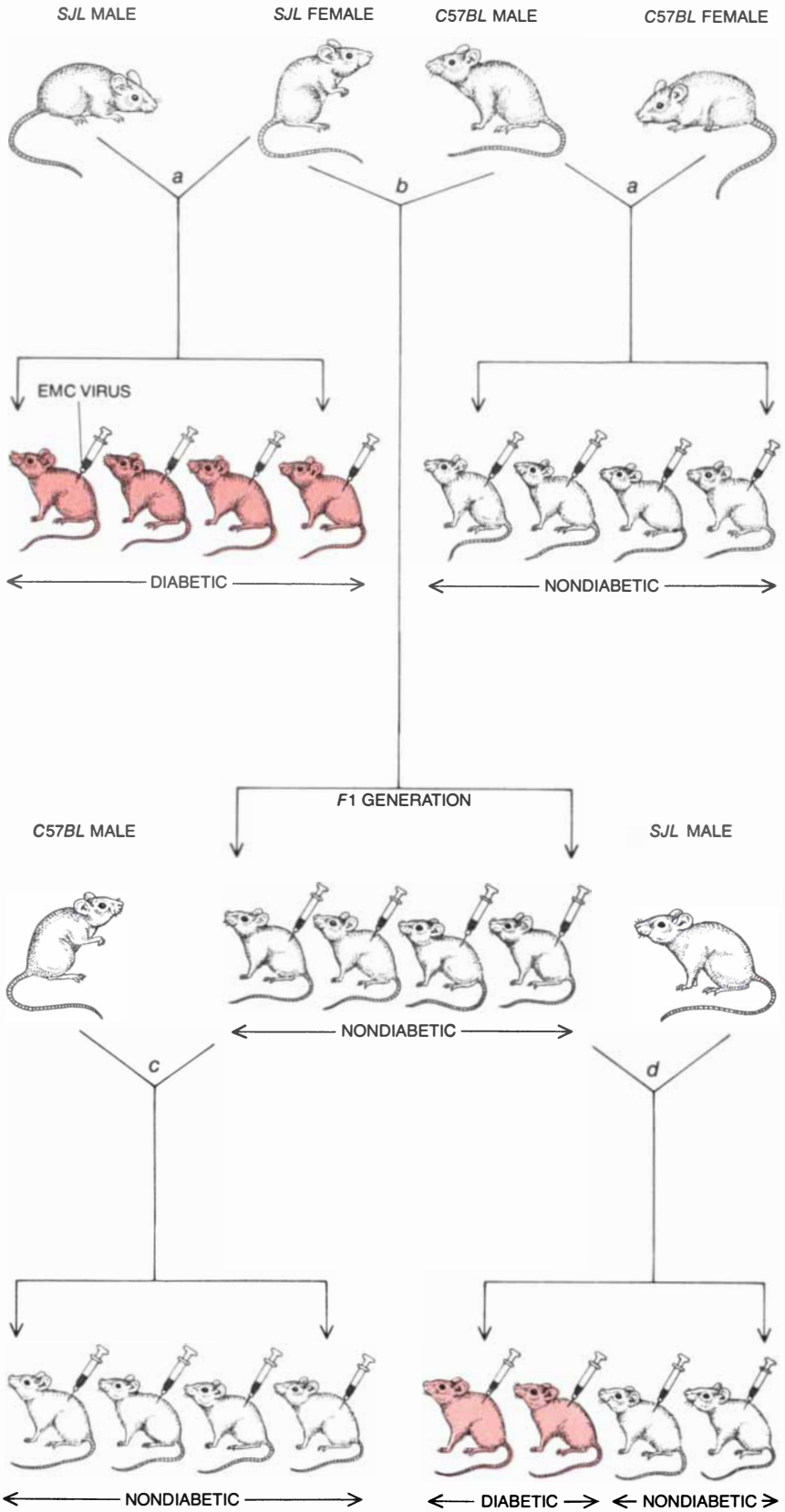
With the information gained from the

HLA GENE COMPLEX



ALLELES ASSOCIATED WITH INCREASED SUSCEPTIBILITY	*DW* 3 *DW* 4 *DRW* 3 *DRW* 4	*B* 8 *B* 15 *B* 18 *B* 40 *BW* 22	*CW* 3	*A* 1 *A* 2
ALLELES ASSOCIATED WITH DECREASED SUSCEPTIBILITY	*DW* 2 *DRW* 2	*B* 5 *B* 7		*A* 11

HISTOCOMPATIBILITY ANTIGENS are cell-surface proteins that provide each individual's tissues with a unique biological label. The major category of these antigens in man, known as the HLA complex, are coded for by a series of genes on chromosome No. 6. The genes at each position or locus on the two homologous chromosomes are not always identical, and a large number of alternate genes (alleles) may be present in a population. Certain HLA alleles have been shown to be associated with an increased or decreased risk of developing juvenile-onset diabetes. Such correlations imply that genes closely linked to the high-risk HLA alleles (perhaps those genes controlling the immune system) may play a role in the genesis of diabetes.



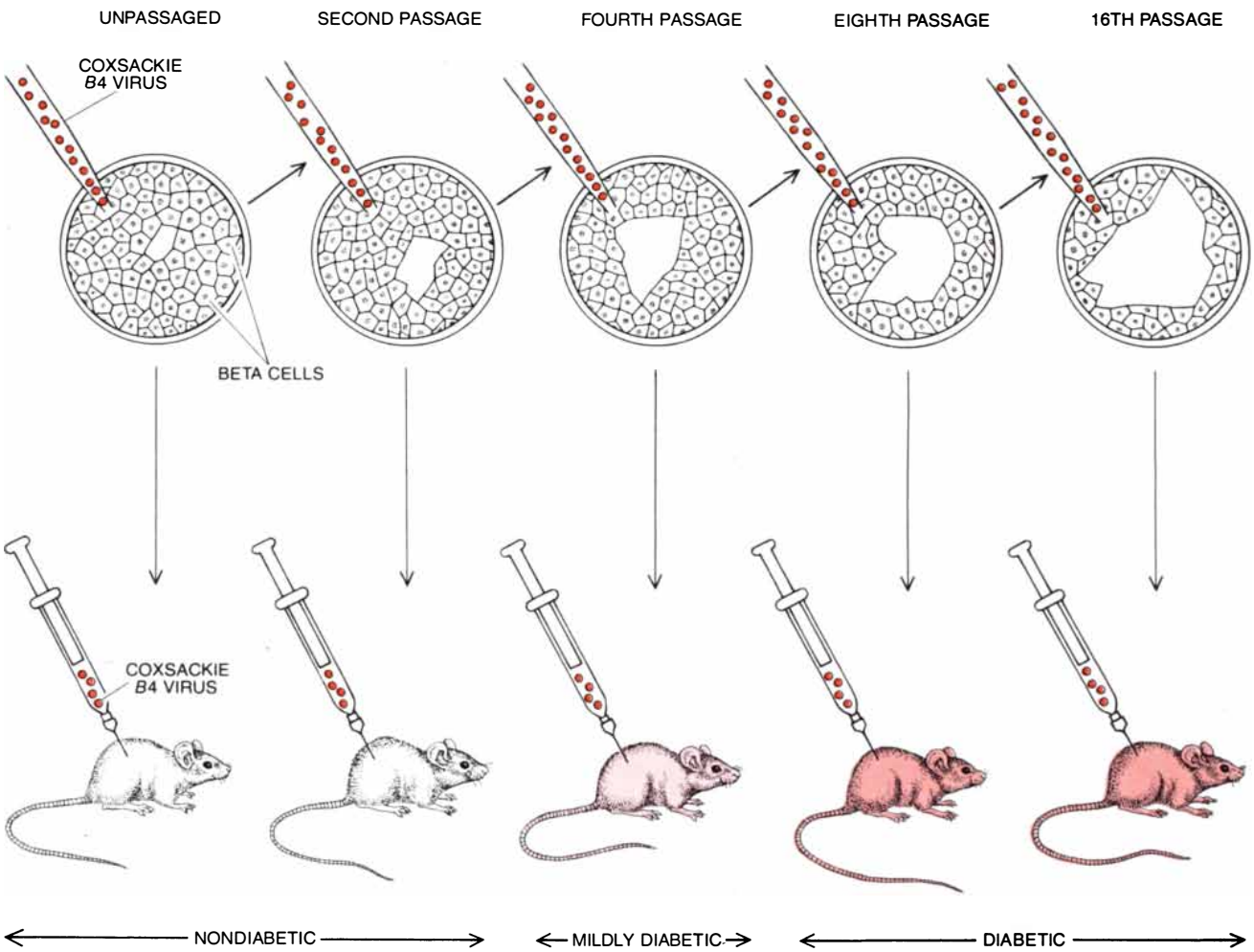
GENETIC SUSCEPTIBILITY to virus-induced diabetes can be demonstrated by breeding experiments in mice. Mice of the strain designated *SJL* develop diabetes when they are infected with the EMC virus, whereas those of the strain designated *C57BL* do not (*a*). When the susceptible and resistant strains are crossed, their offspring (the *F*1 generation) are resistant to EMC-induced diabetes (*b*). Similarly, a backcross between the *F*1 mice and the resistant parental strain gives rise to resistant offspring (*c*). When the *F*1 mice are backcrossed with the susceptible parental strain, however, a high percentage of the offspring are susceptible (*d*). These results suggest that susceptibility to EMC-induced diabetes is inherited as a recessive trait.

EMC model we began to look for other viruses that might induce diabetes in mice, particularly those viruses that also infect human beings. One likely candidate was the Coxsackie family of viruses. First isolated in the 1940's from a group of patients living in Coxsackie, N.Y., these viruses have RNA as their genetic material. They can cause upper-respiratory distress and muscle pain and can infect the heart and the brain. Soon after the discovery of the Coxsackie viruses Gilbert Dalldorf of the New York State Department of Health and A. M. Pappenheimer of the Harvard Medical School found that infection of mice with members of the virus group designated Coxsackie *B* caused severe destruction of the acinar cells in the pancreas but spared the adjoining islets of Langerhans. Twenty years later D. Robert Gamble of the West Park Hospital at Epsom in England, who had thought for some time that there might be a connection between Coxsackie virus infection in children and diabetes, and George E. Burch of the Tulane University School of Medicine undertook to repeat those experiments. They too found destruction of acinar cells, and they also noted that some cells in the islets of Langerhans were damaged. Gamble observed that there was a transient rise in the levels of glucose in the blood of some of the infected mice, but he had difficulty reproducing his experimental results.

Attempts in my laboratory to induce diabetes in mice with Coxsackie *B* virus also failed. Then Yoon, Onodera and I decided to take a different approach: to make the virus more virulent by adapting it to the beta cells. Exploiting the fact that beta cells can be grown in culture, we infected the cultures with a strain of Coxsackie virus known as *B*4. After several days we harvested the virus and infected fresh beta-cell cultures with it. We repeated the process several times. Initially the virus replicated poorly, but with repeated passaging through the cultures it became more virulent, presumably because the passaging selected for variants of the virus that replicated better in beta cells.

When the virus had been passaged 14 times, it was injected into mice. Within a week more than half of the animals had developed diabetes: they had low insulin levels and high glucose levels, many of the islets showed damage and viral antigens were detected in some of the beta cells. It soon became apparent that our success in inducing diabetes was due not only to the adaptation of Coxsackie *B*4 virus but also to the selection of the appropriate strain of mouse. As in the case of EMC virus, only certain strains of mice developed diabetes when they were infected with Coxsackie *B*4. Once again genetic factors were clearly influencing susceptibility to diabetes.

Another virus that we tested for its



PASSAGING A VIRUS repeatedly through cultures of beta cells increases its ability to induce diabetes. It is thought that passaging selects for those variants of the virus that reproduce most successfully in beta cells. For example, Coxsackie *B4* virus does not normally cause diabetes in mice. If the virus is passaged several times, however, its capacity to kill beta cells and cause diabetes is increased.

ability to induce diabetes was reovirus. The prefix reo- stands for respiratory-entero-orphan virus. The virus was so named in 1959 by Albert B. Sabin of the University of Cincinnati College of Medicine because it was found in the respiratory and gastrointestinal tract of many children and yet appeared to be an "orphan": it did not cause serious illness. In mice, however, reovirus attacks the acinar cells of the pancreas, although it spares the islets of Langerhans. Encouraged by our success in making Coxsackie *B4* virus more virulent for beta cells by passaging it several times through beta-cell cultures, we did the same thing with reovirus before injecting it into mice. When we examined the pancreas of the reovirus-infected mice several days later, we found virus particles not only in beta cells but also in alpha and delta cells. The total number of islet cells damaged by reovirus was considerably less than that damaged by Coxsackie *B4* infection, and the blood-glucose levels of most of the mice were relatively normal. When the reovirus-infected mice were stressed by injection with a large dose of glucose, however, it became quite clear that their ability to metabolize glucose had become impaired. The glucose levels remained elevated for a considerably longer period of time than normal, a condition resembling the abnormal glucose tolerance of human diabetics.

Elliot J. Rayfield of the Mount Sinai School of Medicine in New York observed similar subtle shifts in the metabolism of glucose and the release of insulin in hamsters infected with Venezuelan equine encephalitis virus, which occasionally causes fever, headache and damage to the nervous system in human beings. Taken together these studies show that variants of several common viruses known to infect human beings can attack islet cells in experimental animals and induce symptoms resembling those of diabetes.

The first direct evidence that viruses are capable of causing diabetes in human beings came in the spring of 1978. A 10-year-old boy was brought to a hospital in the Washington, D.C., area with an illness lasting for three days that resembled influenza. He was admitted because he suffered from severe lethargy bordering on coma. Laboratory tests revealed that the child had very high levels of blood glucose (600 milligrams per 100 milliliters of plasma) and was suffering from diabetic ketoacidosis. In spite of intensive therapy his condition rapidly deteriorated, and within a week he died. At autopsy Robert Marshall Austin of the National Naval Medical Center found inflammatory white cells in the islets of Langerhans and observed that many of the beta cells had been destroyed.

Since these signs were reminiscent of those observed in mice infected with diabetes-causing viruses, Austin froze a small piece of the child's pancreas for further examination. A few weeks later he told us about the case, and we eagerly agreed to look for a virus. We homogenized the sample of pancreas and added the homogenate to tissue-culture cells known to be sensitive to a variety of viruses. Within a few days the cells showed the signs characteristic of infec-

tion. With standard techniques of virology we then succeeded in isolating from the cells a virus that had properties similar to but not identical with the virus Coxsackie *B*4.

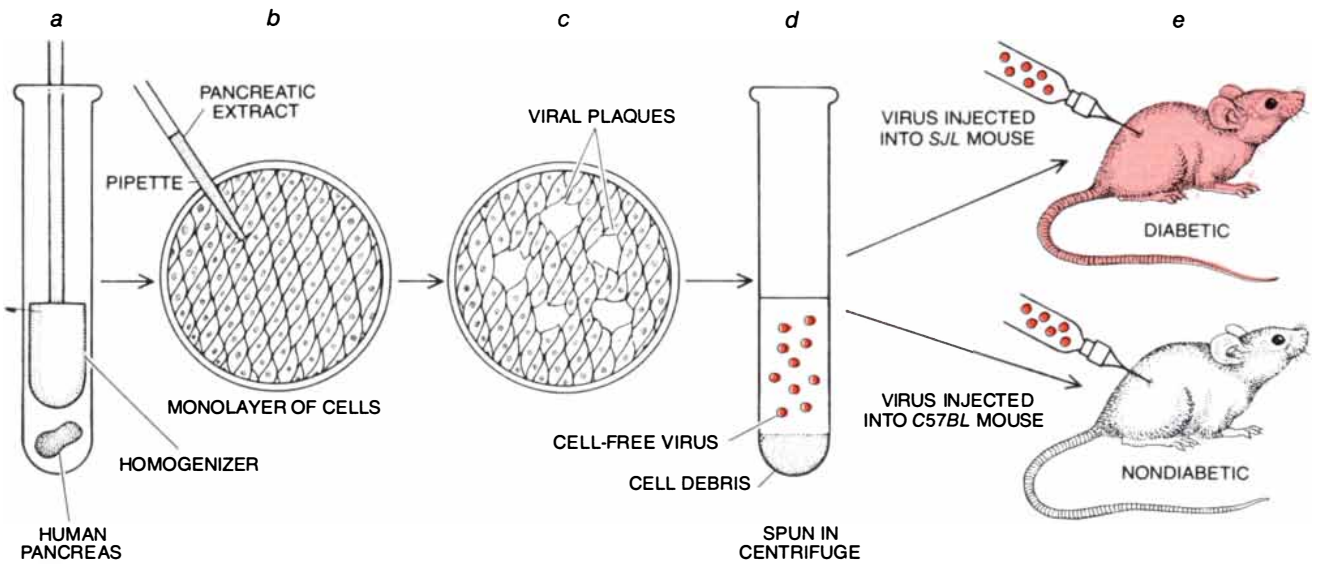
In order to prove that the virus actually came from the patient and was not an inadvertent laboratory contaminant we looked for signs of infection in the samples of the patient's blood that had been taken during the course of the disease. It is well known that early in the course of a viral infection there is little if any antibody to the virus in the patient's serum because the immune system has not had enough time to respond to the invader. Later on specific antibody can easily be detected, and a rise in the levels of antibody over a period of several weeks usually implicates the virus as the cause of the illness. Serum obtained from the diabetic child on his admission to the hospital contained no antibody to Coxsackie virus, but there were significant amounts of antibody in serum obtained about a week later. This finding strongly implied that the virus was not a laboratory contaminant, but it left open the possibility that the child's infection might have been fortuitous and unrelated to the cause of his diabetes.

Here our animal model proved to be invaluable. Strains of mice known to be either susceptible or resistant to diabetes induced by Coxsackie *B*4 were inoculated with the virus isolated from the child's pancreas. Within a week a high percentage of the susceptible mice had developed diabetes but the resistant mice had not. In the susceptible mice inflammatory white cells were seen in the islets of Langerhans and viral antigens were detected in the beta cells. These experiments gave strong evidence that the child's diabetes had indeed been induced by the virus.

Nevertheless, diabetes does not seem to be a common consequence of Coxsackie *B*4 infection. Almost 50 percent of the adult U.S. population has at one time or another been exposed to the virus, yet most people come through such infections with no apparent damage to the pancreas and no diabetes. Moreover, many juvenile-onset diabetics do not have antibody to Coxsackie *B*4. The case I have described might therefore have been a rare exception in which there was the right combination of a virus toxic to beta cells and a genetically susceptible individual. On the other hand, viruses capable of killing beta cells might be far commoner than has been suspected, giving rise to a spectrum of disease ranging from frequent subclinical infections with minimal beta-cell damage to relatively rare instances of moderate to severe beta-cell damage and overt diabetes. The situation might be analogous to the effects of the virus of poliomyelitis before the development of vaccines against it: many people had subclinical infections but only a few developed paralytic disease. There is, however, no evidence that a healthy individual can "catch" diabetes from an already diabetic person.

Since many juvenile-onset diabetics do not have antibody to Coxsackie *B*4, the search is on for other viruses that cause diabetes in man. It may be that a series of viral infections in childhood, each infection causing some beta-cell damage, finally results in overt diabetes when the reserve of beta cells has been sufficiently depleted. Along these lines it is interesting to speculate whether some children are born with a relative deficiency of beta cells or an impaired capacity to repair or regenerate beta cells once they are damaged. If that is the case, viruses might more readily produce diabetes in these already deficient individuals.

It is also possible that viruses may be only one of many causes of diabetes (and perhaps a minor one) and that other insults from the environment such as drugs and toxic chemicals might similarly damage beta cells and give rise to diabetes. In this connection it has been known for 35 years that the drug alloxan can destroy beta cells and induce diabetes in experimental animals. The drug is highly selective in its effects: damage to beta cells can be observed within minutes after injection. In the early 1960's another drug, streptozotocin, was also shown to be toxic to beta cells and capable of inducing diabetes in experimental animals. Recently Arthur A. Like and Aldo A. Rossini of the University of Massachusetts Medical School showed that whereas a single large dose of streptozotocin is directly toxic to the beta cells of experimental animals, multiple small doses appear to act indirectly, presumably by altering the beta cells so that they become vulnerable to attack by the animal's own immune system. Under the latter conditions only certain inbred strains of mice show signs of inflammation in their islets and develop diabetes, again suggesting the importance of genetic factors.



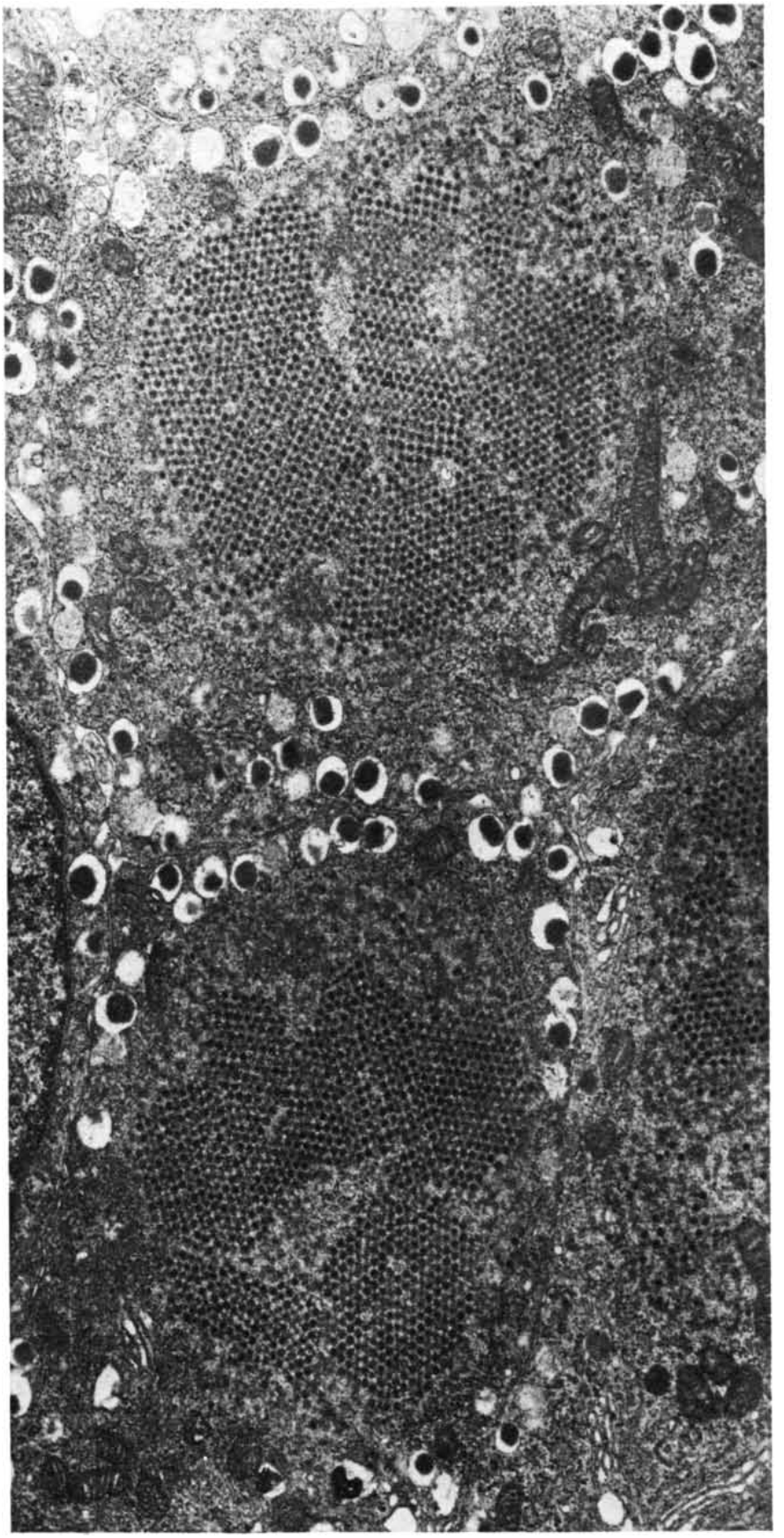
HUMAN DIABETES VIRUS was isolated in the author's laboratory from the pancreas of a boy who had died of juvenile-onset diabetes. A small piece of the child's pancreas obtained at autopsy was homogenized (*a*), and the extract was placed on a layer of tissue-culture cells known to be sensitive to a variety of viruses (*b*). Within several days the dissolution of virus-infected cells gave rise to plaques, or holes, in the sheet of cells, indicating that a virus was present (*c*). The cells were frozen and thawed several times to liberate the virus, which was separated from other cell components by spinning in a centrifuge (*d*). The fluid containing the virus was then injected into strains of mice known to be susceptible or resistant to Coxsackie *B*4-induced diabetes. After several days the susceptible mice developed diabetes but the resistant mice did not (*e*), implying that the child's diabetes had been induced by the virus isolated from his pancreas.

Paradoxically, because of its toxicity and its high specificity for beta cells, streptozotocin has been used with some success to treat a rare tumor of human beta cells known as an insulinoma. In general, however, drugs and chemicals that kill beta cells were employed merely as tools for investigating the pathogenesis of diabetes in experimental animals. Then in 1975 a rodent poison known as Vacor, which has a molecular structure resembling that of streptozotocin, was introduced into the U.S. Since that time it has been accidentally or deliberately ingested by a number of people with serious consequences. Some died, and as many as 20 of the survivors developed acute diabetes calling for treatment with insulin. In two of the fatal cases examined at autopsy there was clear evidence of beta-cell destruction. A small number of other drugs and chemicals have been shown to be toxic to beta cells, although in general the damage is mild and transient.

What of the thousands of other chemicals, natural and man-made, to which human beings are exposed daily and whose effects on the beta cells have not been examined? Do some of them damage the beta cells and cause juvenile-onset diabetes? If they do, the causes of diabetes could turn out to be as numerous as those of the common cold. In any case it seems unlikely that anything as simple as a vaccine to prevent diabetes will be available in the near future.

To sum up, work in laboratories around the world has now firmly established that diabetes is not a simple disease with a single etiology. Even the juvenile-onset type may have multiple causes. Of major importance are the recent findings that genes closely linked to the HLA complex influence the risk of developing juvenile-onset diabetes and that most newly diagnosed patients have in their serum an autoantibody to islet cells. The current hope is that it will eventually be possible to identify those individuals who are most susceptible to beta-cell damage and find some way to protect them.

A second important implication is that although some cases of juvenile-onset diabetes may be explained primarily on a genetic basis and others on an environmental basis, still other cases appear to arise from a complex interaction between the genetic background of the individual and his environment. The relative importance of different environmental insults, such as viruses and toxic chemicals, and of genetic factors and autoimmunity is still not clear, but research in these areas is receiving increased attention. Although it may be some time before it is possible to prevent or cure juvenile-onset diabetes and its complications, the mysteries that have long surrounded this ancient disease are gradually being dispelled.



REOVIRUS PARTICLES form large crystalline arrays inside a beta cell from a virus-infected mouse, as is shown in this electron micrograph. Also visible in the cytoplasm of the cell are numerous secretory vacuoles containing dark granules of insulin. Although reovirus infection does not induce severe diabetes in mice, it does impair the animals' ability to remove glucose from the blood as measured by glucose-tolerance tests. The magnification is 17,300 diameters.

Shape-Memory Alloys

If one of these new alloys has been formed into a shape at a certain temperature and is then deformed at another temperature, it can "remember" the original shape. The effect has many applications

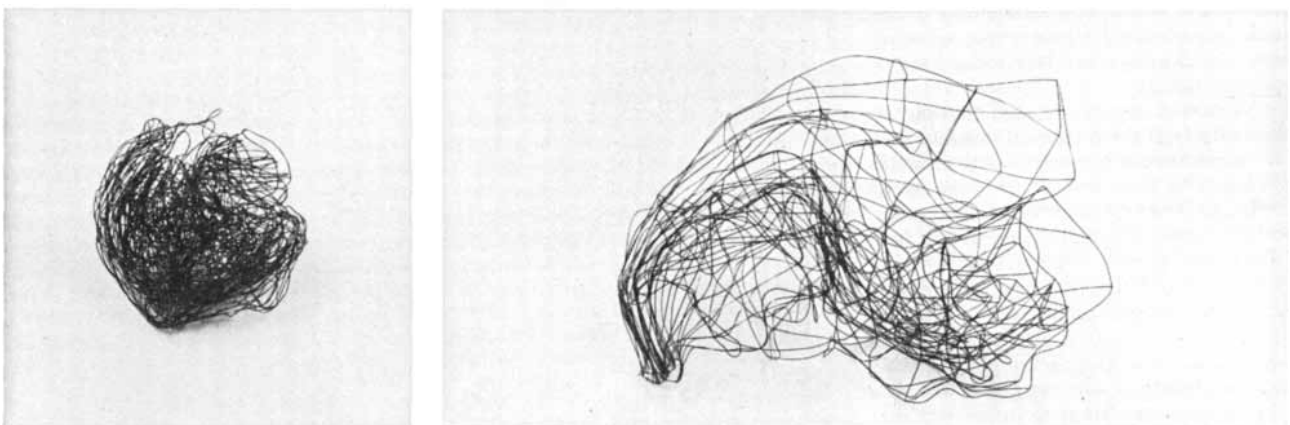
by L. McDonald Schetky

Metals are characterized by such physical qualities as tensile strength, ductility, malleability and conductivity. To these, in the case of a new family of alloys, one can now add the anthropomorphic qualities of memory and trainability. The new alloys exhibit what is called the shape-memory effect. If such alloys are plastically deformed at one temperature, they will completely recover their original shape on being raised to a higher temperature. In recovering their shape the alloys can produce a displacement or a force, or a combination of the two, as a function of temperature. Because of these novel and remarkable properties shape-memory alloys are helping to solve a wide variety of problems. In one well-developed application shape-memory alloys provide simple and virtually leakproof couplings for pneumatic and hydraulic lines. They are also serving as tight seals and couplings in electronic assemblies. Most recently the alloys have been exploited in mechanical and electromechanical control systems to provide, for example, a precise mechanical response to small and repeated temperature changes. Several promising applications are being tested in medicine. More remote, but nonetheless tantalizing, is the possibility of harnessing shape-memory alloys to convert low-grade heat (such as the heat of water at the surface of the ocean) into mechanical energy.

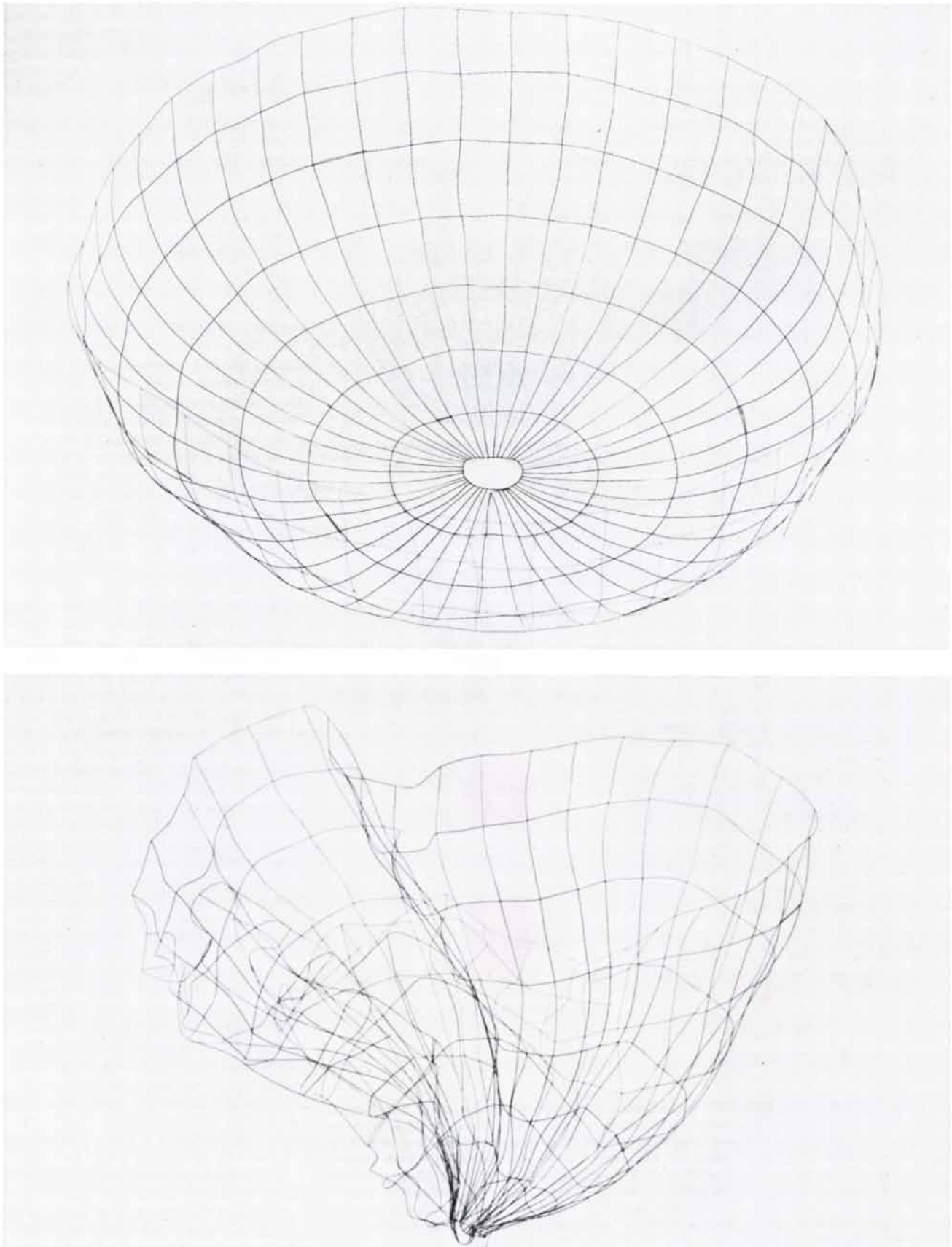
In order for an alloy to exhibit the shape-memory effect it must have a crystal structure that can shift into the configuration known as martensite when it is subjected to a certain temperature or stress and then shift out of it. In a simple example a wire of shape-memory alloy might be bent at room temperature into the form of a four-leaf clover. Then the wire is heated until its crystal structure assumes a high-temperature configuration called the beta or parent phase. Next the wire is rapidly cooled so that the atoms in the metal rearrange themselves into the crystal form of martensite. One can now bend or twist the wire into any other form. If the wire is later heated to a temperature above that at which the martensite reverts to the parent phase, there is an orderly shift of large groups of atoms that restores the original cloverleaf form. Since the martensite transformation is essential for the shape-memory effect, alloys that exhibit memory are also known as marmem alloys.

The shape-memory effect had been noted as early as 1938, when Alden B. Greninger of Harvard University and V. G. Mooradian of the Massachusetts Institute of Technology showed that the martensite phase in brass (an alloy of copper and zinc) could be made to form and disappear with a change in temperature. At about the same time G. V. Kurdyumov, a Russian metallurgist noted for his early work on the crystallography of martensites, particularly in steel, studied the phase relations in brass between the high-temperature beta phase and the martensite formed by rapid cooling. Later Thomas A. Read and his associates at the University of Illinois investigated the shape-memory effect in gold-cadmium alloys and demonstrated the forces that could be developed by the phase transitions. Alloys as diverse as iron-platinum, indium-cadmium, iron-nickel, nickel-aluminum and stainless steel have been observed to exhibit shape memory in varying degrees.

It was not until 1962, however, that the phenomenon came to worldwide attention with the announcement of shape



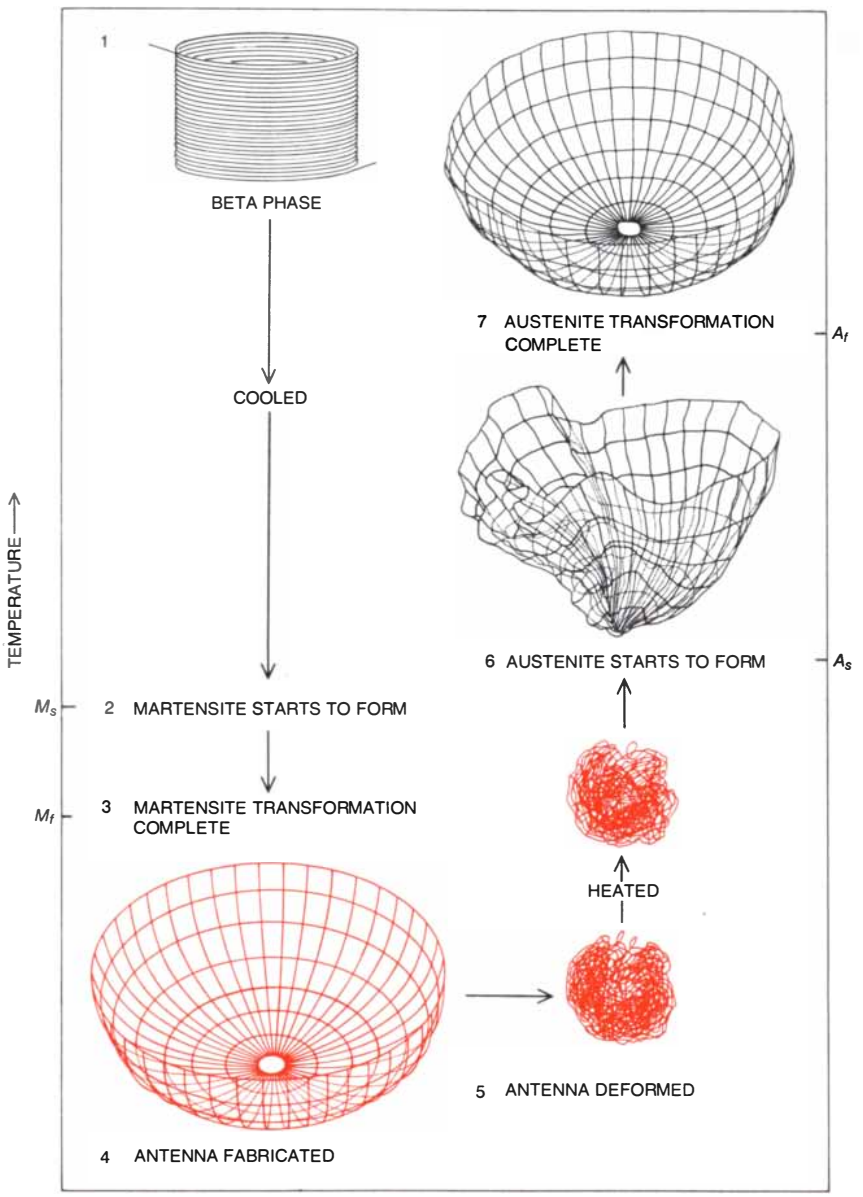
SPONTANEOUS UNFOLDING of a wire hemisphere that could serve as a small antenna for a spacecraft is shown in this sequence of four photographs. The antenna is made of an alloy of nickel and titanium (Nitinol) that exhibits the shape-memory effect. In the photograph at the left the fully formed antenna has been crushed at room temperature into a tight ball less than five centimeters across. As the temperature of the mass of wire is increased the antenna gradually unfolds. When the temperature reaches 77 degrees Celsius, the struc-



ture assumes its original configuration (*top right*). The diameter of the unfolded antenna is about 25 centimeters. The restoration of the antenna's shape coincides with the disappearance of crystals of martensite in the material and their replacement by crystals of austenite, which has a simpler crystal structure. The crystallographic history of the sequence is illustrated on the next page. The antenna was designed by the Goodyear Aerospace Corporation. Although it has never seen actual service, it vividly demonstrates the properties of such alloys.

memory in an alloy of nickel and titanium. The alloy was discovered by William J. Buehler of the U.S. Naval Ordnance Laboratory (since renamed the Naval Surface Weapons Center) in Silver Spring, Md. Buehler's alloy was dubbed Nitinol (for nickel-titanium Naval Ordnance Laboratory).

Before going into the various ways the shape-memory effect can be exploited I shall describe in somewhat more detail the phenomenon itself, the conditions of temperature and stress needed to give rise to it and the crystallography that underlies it. The martensite crystal structure that is essential for the effect can be produced in two general ways: by subjecting an alloy to a stress whose magnitude is related to temperature or by rapidly quenching a suitable alloy to some critical temperature. In the second method, which is the one most commonly applied, the martensite structure appears spontaneously through a shearing motion of the atoms in the alloy or through a process of nucleation and crystal growth. In each case the shift is classed as a "diffusionless transformation" resulting from the coordinated movement of large blocks of atoms.

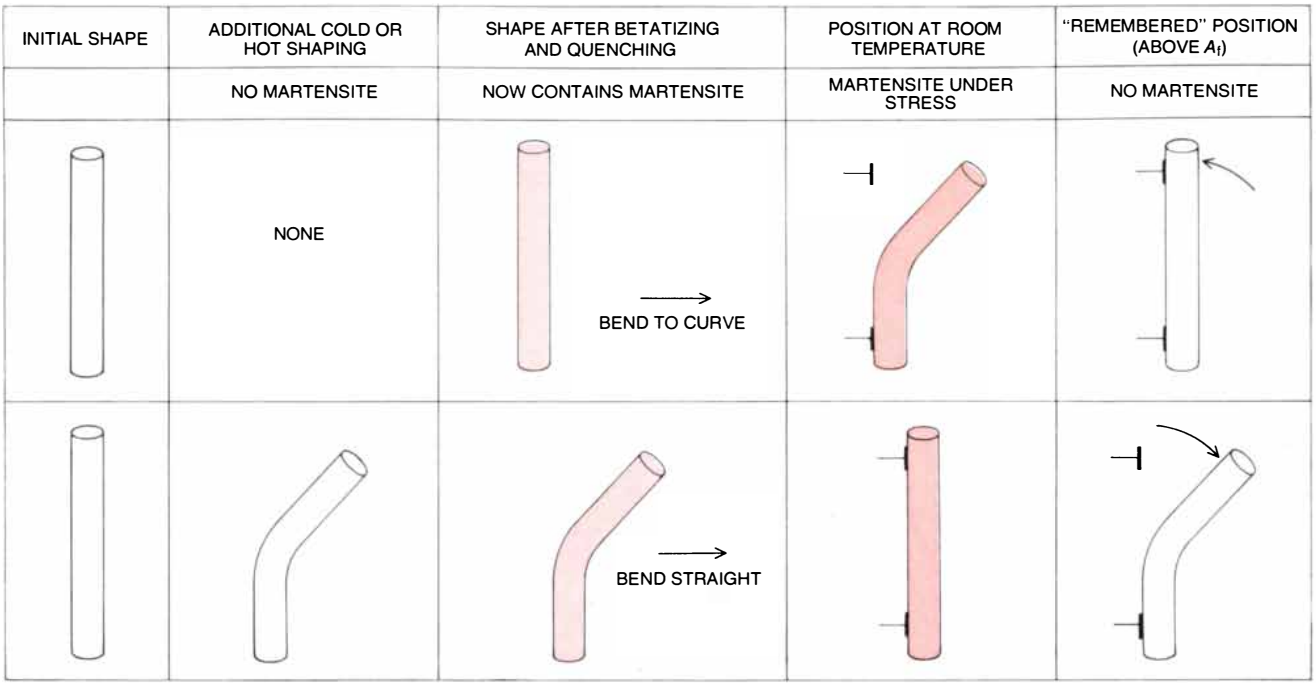


CRYSTALLOGRAPHIC HISTORY of the spacecraft antenna shows the role of temperature in the shape-memory effect. The construction material, Nitinol wire in a large coil, is raised to a temperature of 650 degrees C. and is stabilized so that the crystal structure of the material is entirely in the beta, or "parent," phase (*1*). This phase is often the crystal structure austenite. The wire is now cooled. At the temperature M_s (60 degrees C.) the new crystal phase martensite starts to form, replacing the beta phase (*2*). At M_f (52 degrees) the transformation to martensite is finished (*3*). While the Nitinol wire is held at a temperature below M_f it is cut into short lengths, which are gently bent to form the segments of the intended hemisphere (*4*). Where the segments cross one another they are fastened by tack welding. One can now crush the antenna into a small volume (*5*). In order to restore the original shape the crushed structure is heated. At temperature A_s (71 degrees) austenite begins to replace martensite (*6*). On reaching A_f (77 degrees) the antenna has unfolded completely (*7*). In this particular case what the shape-memory alloy "remembers" is not the actual configuration of the antenna but the gentle curves of the coiled wire from which the antenna was constructed. As the wire tries to straighten itself it is constrained to the shape of a bowl by the multiple welds at the crossover points.

The establishment of a thermodynamic equilibrium between the parent phase and the martensite phase depends on the composition of the alloy, the temperature and the internal stress. The nucleation and growth of martensite are controlled by shear strains that develop between adjacent martensite regions as the alloy is cooled or stressed. The balancing of such internal strains and any external stress causes the martensite regions to grow as an array of self-accommodating plates. The orientation of a plate with respect to the orientation of its neighbor is the one that is energetically stablest in that particular strain field. Whether the field results from an applied stress or from a change in temperature, in the material as a whole the plates assume a variety of orientations and a range of sizes.

Although the structure of the parent phase and of the martensite varies in detail with each alloy, the typical elevated-temperature phase is a disordered body-centered-cubic one (where the atoms form a cubic lattice with one atom in the center of each cube of eight other atoms). As the temperature is lowered that phase shifts into a structure that is either an ordered body-centered-cubic one or a superlattice. In the ordered structure atoms of one species in the alloy take up preferred sites in the cubic lattice with respect to the atoms of the other species. The superlattice can be visualized as interpenetrating cells in which the unit cell of the crystal, the smallest unit that when repeated and extended in all directions defines the crystal structure, consists of several dozen atoms or more. This complexity of crystal structure makes it difficult to describe the relative movements of atoms that result in the martensite transformation. Often invoked to describe the movements are terms such as shuffling and shear. In the marmem alloy of copper, zinc and aluminum four variants of martensite form out of the parent beta phase, each with a crystal orientation displaced by 60 degrees from the others.

The strain associated with one variant compensates the strain in the other variants, so that the growth of multiple self-accommodating fields of martensite plates is energetically favored over the growth of a single plate. The boundaries between the adjacent plates are, however, quite mobile and move readily under the influence of an applied stress. As a result a specimen can be deformed not



TEMPERATURE-ACTUATED SWITCH can be designed so that it opens or closes above a particular temperature. The temperature depends on the alloy selected and coincides with the temperature (A_f) at which martensite is replaced by austenite. If the switch is designed to close above the temperature A_f (*top*), a straight rod of alloy is heated to the "betatizing" temperature and then quenched. The rod now contains martensite (*color*). For the rod to serve as a switching element it is bent, which puts the martensite under stress. When the rod is heated above A_f, the martensite then disappears and the rod straightens, closing the switch. If the switch is designed to open above A_f, the rod must be bent before betatizing and quenching (*bottom*). The rod is then straightened out before it is placed in the switch.

through the slippage of adjacent plates, the usual mechanism of plastic deformation, but through the growth and compensating contraction of adjacent plates.

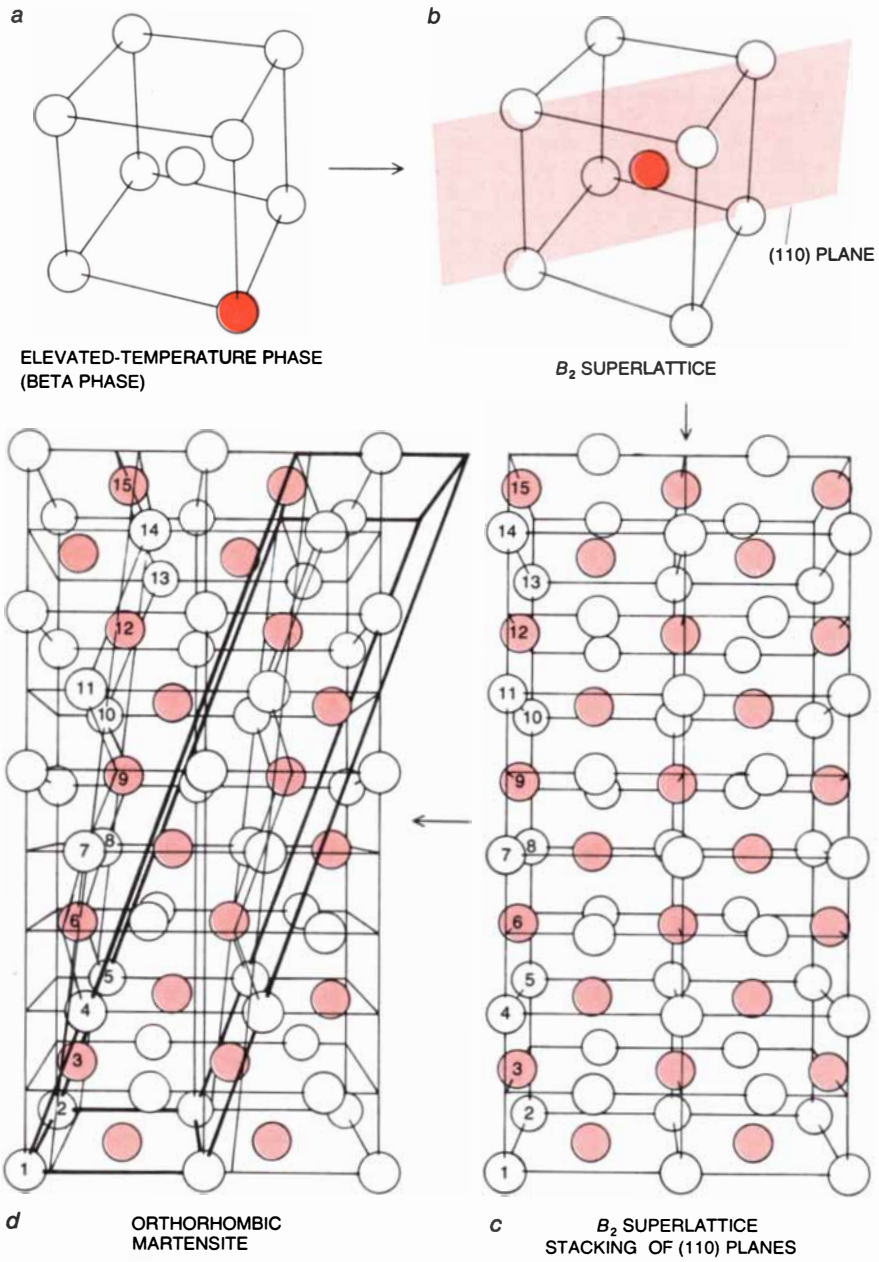
In an alloy that has a beta phase capable of producing martensite under stress one can observe an unusual elastic property called pseudoelasticity or superelasticity. In a typical alloy with this property the metal exhibits normal elastic behavior under stress (that is, it gets longer in some dimension) until the critical stress is reached at which martensite plates begin to form. With further stress the specimen continues to elongate, as if it were being plastically deformed, but when the stress is removed, the martensite plates that have formed revert to the parent phase and the specimen contracts to its original dimension, showing no permanent deformation.

When the temperature of a marmem alloy is reduced to the critical temperature at which martensite starts to form, the transformation proceeds spontaneously. As the temperature is increased above the critical one martensite can still form, but only if an increasing amount of stress is applied. In the temperature range where martensite can form under the influence of stress the elasticity of the material can exceed the elasticity of ordinary alloys by a factor of 10 or more: the material is superelastic. Although superelasticity is not the principal shape-memory effect in marmem alloys, it can provide an additional amount of configurational memory (or equivalent force) for special purposes. Hence an alloy that has already been transformed into martensite will exhibit superelastic strain resulting from the reversible motion of the boundaries between martensite plates. The reorientation of the plates will cause one variant of martensite to grow at the expense of another, depending on the direction of stress with respect to the internal strain energy associated with each boundary.

Trainability, a feature of marmem alloys mentioned at the outset, yields a specimen with memory for two different configurations, a "two-way" memory effect. Training is accomplished by limiting the number of variants of martensite formed when an alloy is repeatedly heated and cooled below the critical temperature. Such limitation inhibits the self-accommodation of the martensite plates and increases the amount of internal strain. The number of variants can be limited by stressing the specimen while it is cooling from the elevated temperature of the beta phase down to the critical temperature. The stress favors the initial formation of particular variants of martensite in the same way that stress at constant temperature favors the growth of one variant at the expense of another. Hence a structure can be trained by repeating many times the sequence of "betatizing" (heating the alloy until it goes into the beta phase), quenching, deforming and betatizing again. Ultimately a change of shape occurs spontaneously; the change is in one direction at a temperature corresponding to the beta phase and in another direction at a temperature below the martensite-transformation one.

Let me now summarize the conditions involved in the attainment of shape memory and therefore the variables that must be controlled if a useful device is to be fabricated. In order to exhibit memory an alloy must undergo a martensite phase transformation. The temperature at which the memory is accomplished is a function of the temperature at which martensite formation begins or the higher temperature at which the alloy reverts to the parent phase (usually the beta phase). The martensite-transformation temperature depends on the composition of the alloy but can be shifted by an applied stress. If the recovery of shape is restrained, a proportional force will be available for doing work or gripping another object. In order to achieve 100 percent recovery of shape the deformation of a part must be limited to an internal strain of between 3 and 9 percent, depending on the particular alloy. Finally, memory can be one-way or, with training, two-way.

The number of uses for marmem alloys has been growing steadily. The original Nitinol alloys were in spite of their high cost soon exploited for a variety of tasks in spacecraft. One of the first applications was a latching device for a British satellite in which a torsion tube



of Nitinol triggered the release of three instrument booms. The release had to be rapid and reliable in order to avoid any dynamic imbalance that might affect the stability of the spinning satellite itself.

In another early application Nitinol alloys solved the problem of coupling hydraulic-fluid lines in the F-14 jet fighter built by the Grumman Aerospace Corporation. Grumman engineers were seeking an alternative to the difficult task of brazing lines that lie close to the aircraft's aluminum skin. The Raychem Corporation, which had had wide experience in heat-shrinkable plastics, proposed a coupling in which a Nitinol alloy was shrunk to a tight fit by being raised above its martensite-transformation temperature. The coupling would have to maintain its gripping integrity down to −120 degrees Celsius, so that the alloy chosen would have to have a transformation temperature in the cryogenic region.

The transformation temperature of the Nitinol family of alloys can be manipulated over a remarkably wide range, from −273 to 100 degrees C., by altering the nickel-titanium ratio and by adding small amounts of other elements. For the hydraulic coupling an alloy with a transition temperature below −120 degrees C. was found. The coupling, in the form of a sleeve, is machined at ordinary shop temperature to have an inner diameter about 4 percent smaller than the outer diameter of the tubes it is designed to join. The coupling is then cooled to below the temperature at which martensite forms. While it is at this low temperature it is forcibly expanded to a diameter 4 percent greater than the tube diameter, for an overall internal strain of about 8 percent. Still at a cryogenic temperature to maintain the martensite phase, the coupling is fitted over the ends of the tubes that are to be joined. When it is warmed to room temperature, the martensite is replaced by

ORTHORHOMBIC CRYSTALS of martensite form spontaneously in a great variety of alloy systems when they are cooled rapidly from an elevated temperature. In the beta, or high-temperature, phase shape-memory alloys usually take the form of disordered body-centered-cubic crystals (*a*). As the temperature is lowered one of the atomic species (*color*) in the alloy takes up a preferred site in the crystal cell. One such structure is a B_2 superlattice (*b*). The martensite transition is best seen by considering an array of B_2 superlattice crystals in which the planes designated (110) are stacked horizontally. Such a stack, made up of 10 (110) planes that consist alternately of eight atoms and seven atoms, is shown in *c*. On cooling, the cubic arrangement of the B_2 superlattice is replaced by the orthorhombic structure of martensite (*d*). The numbered atoms help to clarify the nature of the rearrangement. In a ternary alloy consisting of copper, zinc and aluminum the parent beta phase gives rise to variants of martensite offset from one another by 60 degrees (*e*).

the parent phase, causing the coupling to shrink tightly around the ends of the tubes. The seal can be made still tighter by machining circular ribs into the inner surface of the coupling. Similar but larger couplings are now being tested experimentally for joining hydraulic and air lines on merchant ships.

Engineers at the Raychem Corporation have also developed an electrical connector that forms a high-compression fit yet can be quickly released and recoupled. Fabricated from a Nitinol alloy, the device is in the shape of a ring that closes the fingers on a pin connector. The connector is opened by chilling it with a blast of a cold fluorocarbon gas from an aerosol can; it closes when it warms up. The forces exerted by such shape-memory coupling devices are about 200 times greater than the force that would be exerted by the expansion and contraction of a bimetallic element of the same weight. Moreover, shape memory is effected at a predetermined temperature rather than over a broad range of temperatures, as would be the case for a device dependent on thermal expansion.

Several promising uses for Nitinol alloys are being examined in medicine. Artificial limb joints are fast becoming commonplace, but they are not without problems. In a large joint such as the hip the artificial ball and socket are now usually joined to the bone by a cement, with attendant problems of misalignment and even bone fracture. An alternative fastening method relies on Nitinol "butterflies" attached to the part of the artificial joint that is inserted into the central hollow of the bone. The butterflies are inserted cold and expand to form a tight lock on reaching body temperature. At the Mississippi Methodist Rehabilitation Center in Jackson, James Hughes and his associates have established that Nitinol does not react adversely with living tissue and have undertaken long-term studies with experimental animals.

Another orthopedic problem is the pulling together of the parts of a fractured bone to ensure their alignment and to encourage rapid calcification. The standard method of compressive fixation calls for pins, screws and plates, and it is also slow. With plates of Nitinol, which are installed chilled, body heat pulls the fractured parts of the bone into alignment. The technique is being studied by Alan A. Johnson of the University of Louisville and Frank P. Alicandri of the Polytechnic Institute of New York.

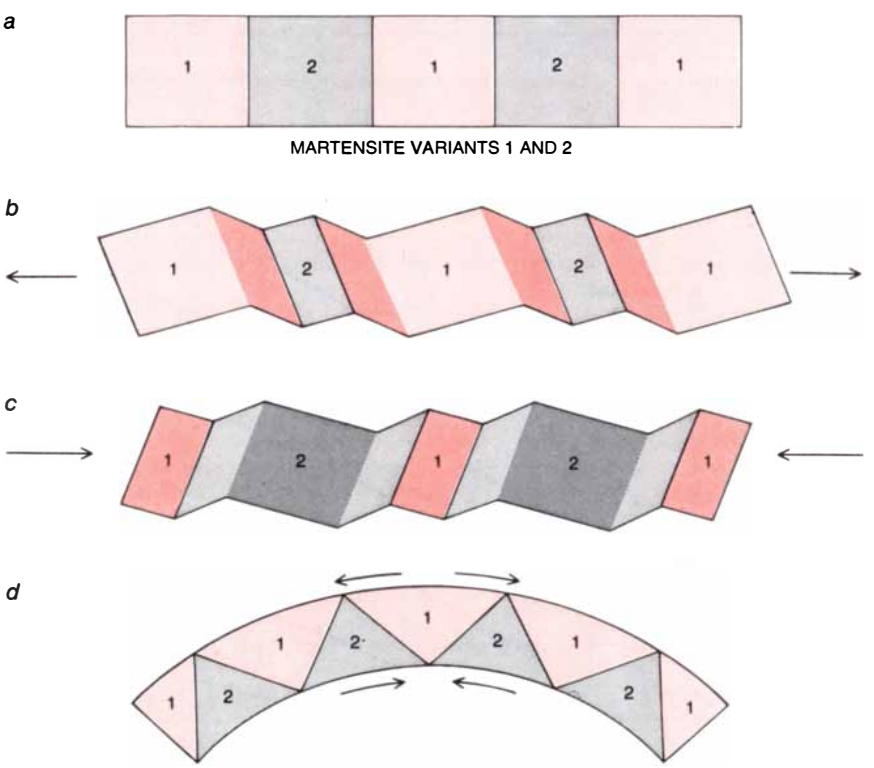
A rather different medical application of the shape-memory effect has the goal of filtering blood clots out of the circulatory system before they can do serious harm. Most clots of dangerous size form in the legs and lower trunk and travel through the veins to the heart and lungs, where they can block a vital blood vessel. Anticoagulant substances, although they are extremely useful, carry risks of their own. Morris Simon of Beth Israel Hospital in Boston and the Harvard Medical School conceived the idea of fabricating a screenlike filter with a mesh size of about two millimeters from a continuous length of Nitinol wire. The wire can be straightened out when it is cooled below the martensite-transformation temperature, chosen to be well below body temperature. As the wire is chilled to maintain its straightened condition, it can be inserted through a catheter in an arm vein into the vena cava, the large vein that feeds into the heart. As the wire warms up it assumes the screenlike form. Experiments on dogs have been encouraging.

Engineers are always interested in eliminating the weak link in mechanisms that must perform reliably with little or no attention. A case in point is the pen-drive mechanism found in tens of millions of recording and industrial-control instruments. In standard instruments the pen arm is actuated by a galvanometer, a venerable electromechanical device that responds to the presence of an electric current.

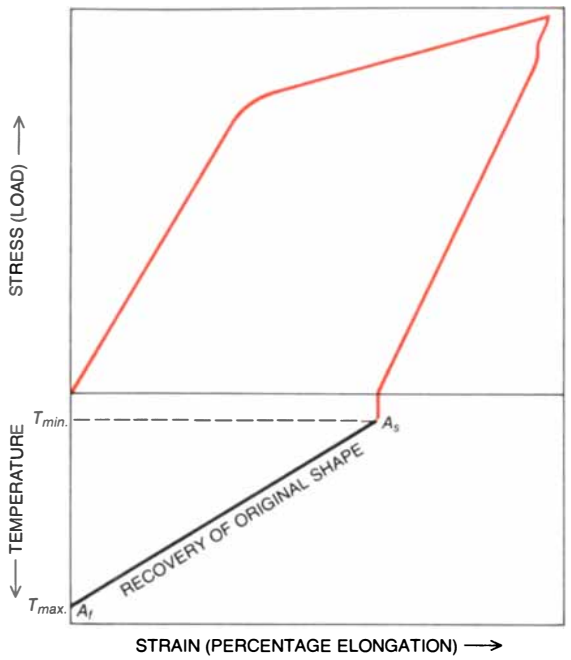
In a search for a still simpler and more rugged mechanism engineers of the Foxboro Company have devised a pen drive that exploits the shape-memory response of Nitinol wire. The wire, under tension restraint, lengthens and shortens in response to the amount of heat supplied by a small induction coil. The coil in turn is energized in response to the voltage input to the recorder from a transducer connected to whatever the recorder is measuring. The Nitinol wire exerts much larger forces than a galvanometer drive does, so that the mechanism requires fewer bearings and pivots. More than 600,000 of the new drives are now in service, representing by far the most intensive application of shape-memory alloys to date.

So far I have dwelt almost entirely on Nitinol and its applications. Recently a family of shape-memory alloys has been developed based not on nickel and titanium but on copper, zinc and aluminum. Because these ternary (three-element) alloys are much cheaper than Nitinol and much easier to machine and fabricate, they seem destined for wide exploitation. They were developed through work at the University of Louvain in Belgium, the Delta Metal Co., Ltd., and the Fulmer Research Institute, Ltd., in Britain collaborating with the Raychem Corporation.

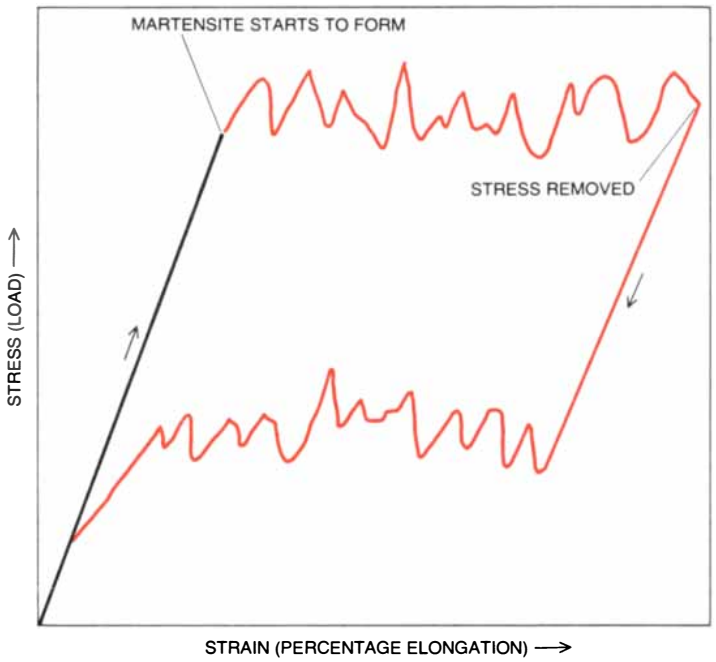
The new ternary alloys consist of between 68 and 80 percent copper. The remaining 20 to 32 percent consists of zinc and aluminum in various proportions. The alloys are remarkable in that only slight shifts in composition can



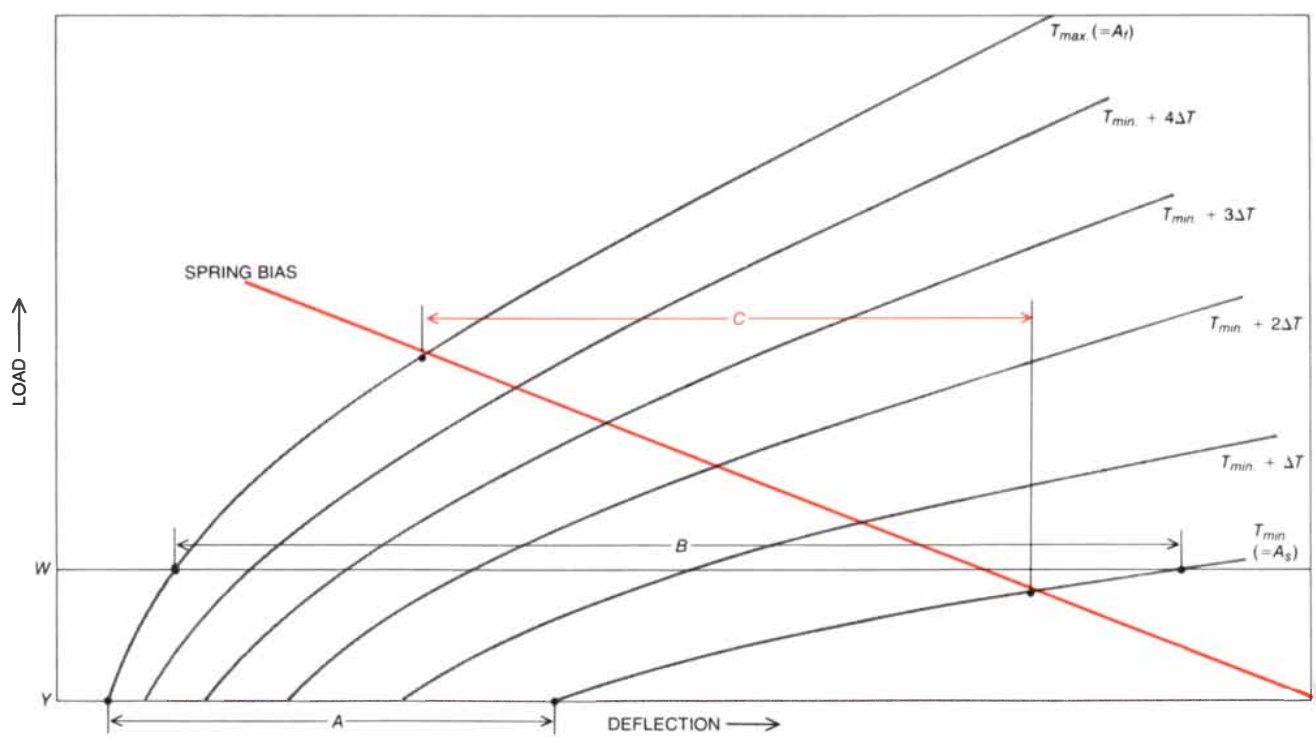
MARTENSITE VARIANTS grow at different rates when they are subjected to stress. Here for simplicity only two variants are depicted. In the absence of stress (*a*) plates of the two variants develop with equal probability. If the specimen is subjected to tensile stress (*b*), plates of variant 1 grow at the expense of those of variant 2. If the specimen is subjected to compression (*c*), the preference is reversed: variant 2 grows at the expense of 1. If the specimen is bent (*d*), the variants grow according to whether the crystals are locally under tension or compression.



SHAPE-MEMORY EFFECT can be represented in a stress-strain curve that includes the effect of temperature. The curve above the horizontal line shows the response of a shape-memory alloy to deformation at a temperature that preserves the martensite phase. Under stress the specimen elongates. With the removal of stress considerable strain remains. If the specimen is now heated to the temperature A_s (or $T_{min.}$), austenite starts to replace the martensite. With increasing temperature the strain drops rapidly. At A_f ($T_{max.}$) conversion into austenite is complete, and specimen regains original shape.



SUPERELASTICITY, or pseudoelasticity, is exhibited by certain alloys in which the transformation to martensite can be induced by stress. Such alloys follow a typical stress-strain curve up to the point where further stress causes martensite plates to form. The specimen then appears to elongate plastically, as if it were being permanently deformed. With the removal of stress the martensite plates revert to the parent phase and the specimen regains its original shape. The elastic elongation resulting from the martensite transformation can exceed the elasticity of ordinary alloys by a factor of 10 or more.



LOAD-DEFLECTION CURVES show the responses of a shape-memory element at various temperatures with and without a bias spring that opposes the deflection. With no load (Y) and no bias spring a change in temperature from $T_{min.}$ to $T_{max.}$ produces a deflection, A. At $T_{min.}$ austenite starts to replace martensite; at $T_{max.}$ the replacement is complete. Under a small load, W, the same temperature change causes a much larger deflection, B. By opposing the deflection with a bias spring (*color*) the deflection can be made to vary in response both to load and to temperature for a total deflection of C. The purpose of the bias spring is to "trim" the element to perform in that part of the load-deflection-temperature spectrum where significant amounts of energy can be harnessed to carry out the desired function.

give rise to large changes in the martensite transformation, from values as low as −105 degrees C. to as high as 299 degrees. For alloys that can be readily fabricated, however, the range is narrower: roughly from −100 degrees to 100 degrees. For a given alloy the shape-memory effect extends over a range of about 80 degrees. At higher temperatures there is a tendency for the martensite to become unstable, so that the usual upper temperature for operation is 150 degrees. Some of the applications for Nitinol are also attractive for the copper-based ternary alloys: couplings, springs and actuators.

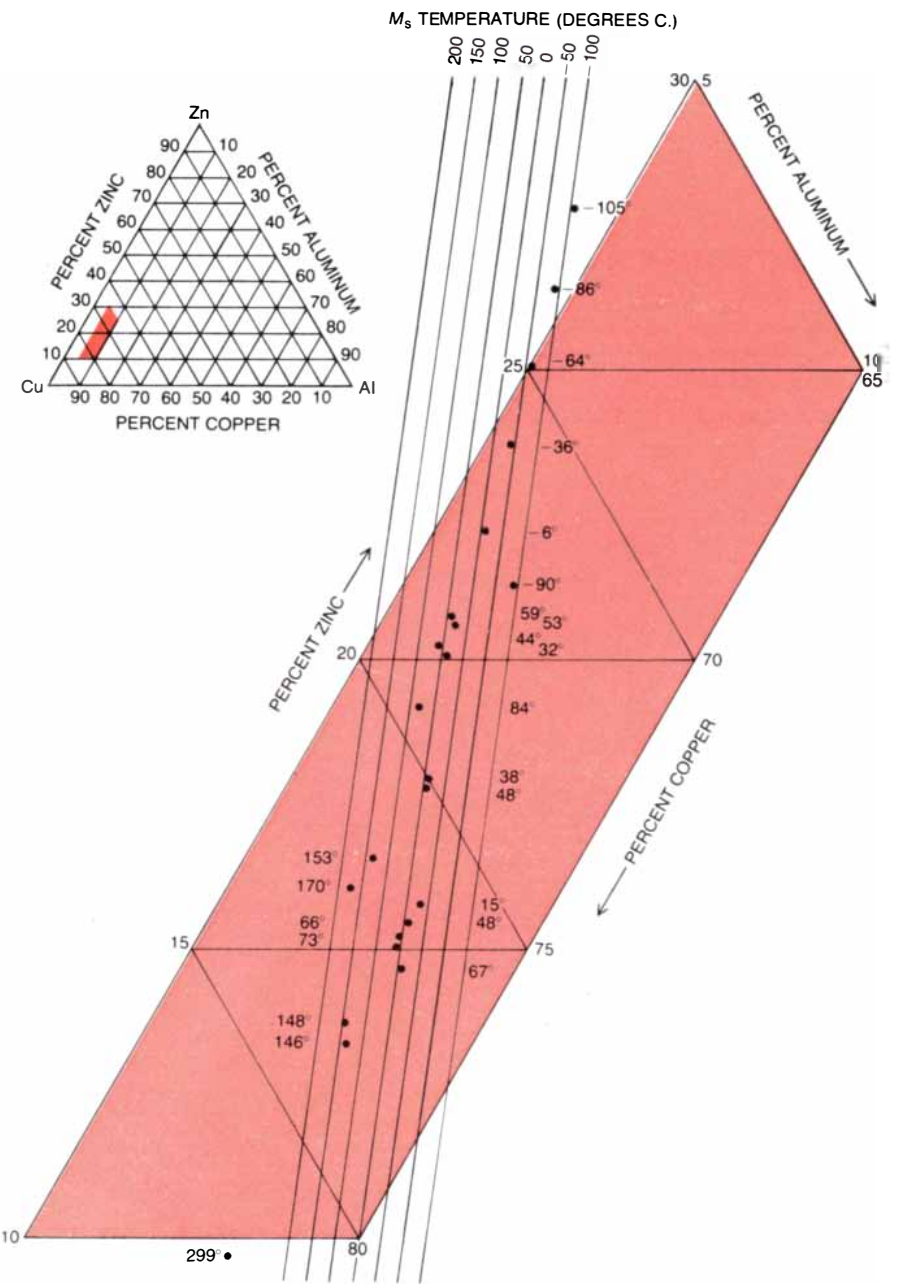
In exploring new applications the Delta company has developed a series of devices in which the shape-memory alloy serves as both a temperature sensor and an actuator. One such device is a thermostatic valve for individual radiators in a home or office heating system. The temperature at which the valve opens can be set by turning a knob that alters the compression on a spring acting against a spring made of the memory alloy. The characteristic of the memory spring is such that the force it exerts increases as its temperature rises. In crystallographic terms, as the temperature increases an increasing fraction of the martensite is transformed into the parent phase, which is associated with a memory of the spring in its expanded position.

A similar memory spring can regulate the clutch system that couples and uncouples the radiator fan in an automobile cooling system. Present devices operate through a viscous-fluid clutch controlled by a bimetallic spring. The memory-alloy clutch eliminates the need for fluid seals and provides a more flexible speed control. Automotive engineers are exploring the possibility of resorting to shape-memory systems to replace small electric motors, solenoid-actuated valves and various engine-control devices. One promising application is a variable-throat or variable-jet carburetor to provide optimum fuel economy over a range of air and fuel temperatures.

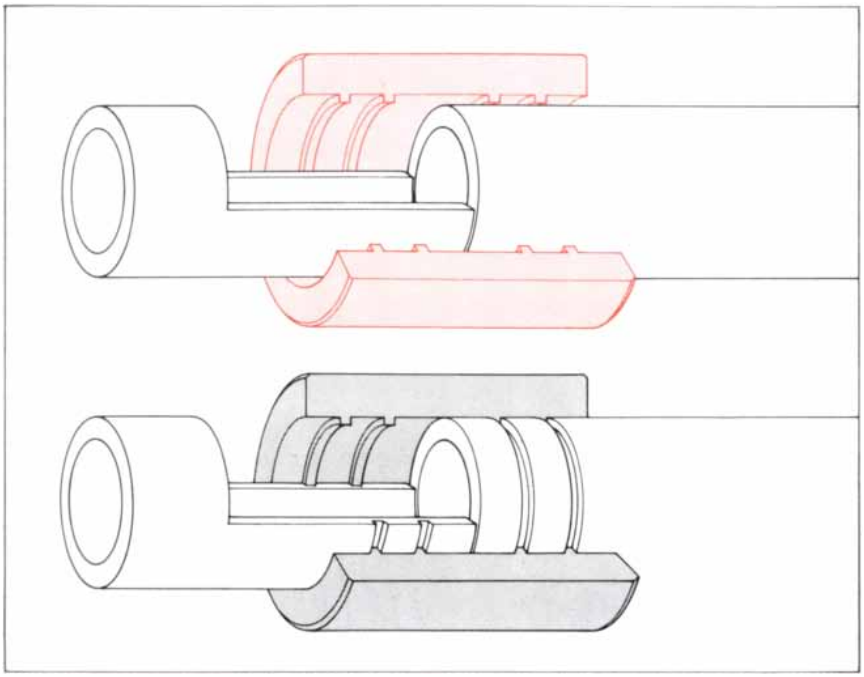
Each application involving shape-memory actuators must take into account operating temperature along with the force and deflection wanted and whether the motion is to be linear or torsional. By working with a series of charts and graphs developed by the Delta company a designer can determine for each particular alloy the amount of bias, or compensating, force necessary to have the shape-memory device operate at a specific temperature and deliver a given force. The number of devices proposed in patents is now approaching 100 and is accelerating.

One final application of the shape-memory effect deserves mention: the recovery of energy from waste heat or low-temperature water by the direct conversion of the heat into mechanical energy. Such a solid-state engine was considered soon after the discovery of the shape-memory effect in Nitinol. The first patent issued to Buehler and David M. Goldstein for a shape-memory engine in 1966 has been followed by what now seems an endless procession of energy devices. The first engine to really show the feasibility of generating useful amounts of energy was one invented by Ridgway Banks of the Lawrence Radiation Laboratory of the University of California. Possibly because Banks has had professional training in music, his devices give something of the aesthetic pleasure of stringed instruments.

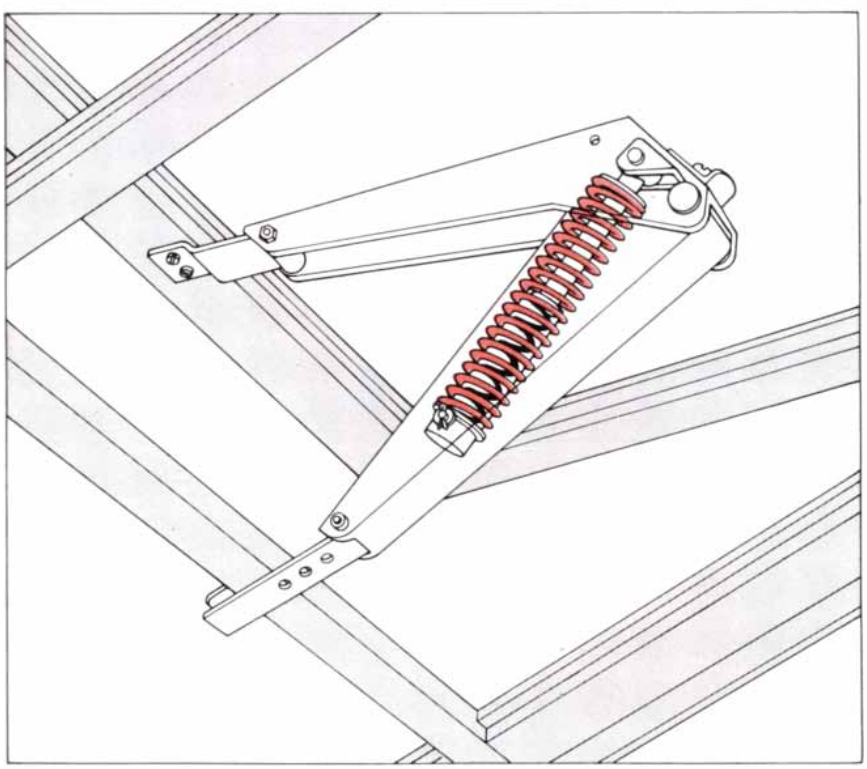
The theoretical efficiency of a solid-state engine can be analyzed by taking into account the latent heat of transformation, the additional stress delivered for an additional change in temperature and the difference in free energy between the martensite phase and the elevated-temperature phase. By means of such an analysis L. Delaey of the University of Louvain has arrived at a theoretical efficiency of 4 or 5 percent, which



MARTENSITE-FORMATION TEMPERATURES, M_s, for ternary alloys of copper, zinc and aluminum can be made to vary over more than 400 degrees C. as a result of only small changes in composition. The shape-memory alloys all lie in the copper-rich corner (*color*) of the triangle representing ternary mixtures. The amount of aluminum varies from about 4 to 10 percent, the amount of zinc from just under 10 to 28 percent. Balance in each case is copper.



HYDRAULIC-TUBE COUPLINGS for the Grumman F-14 jet fighter are fabricated from a Nitinol alloy that has a martensite-formation temperature, M_s, in the cryogenic region below −120 degrees C. The sleevelike coupling is machined at normal temperature to have an inner diameter 4 percent less than the outer diameter of the tubes to be joined. The sleeve is then cooled below the M_s temperature and is mechanically expanded to have an inner diameter 4 percent greater than the tubes' outer diameter (*top*). Still held at a temperature below M_s, the sleeve is placed around the ends of the tubes. When the coupling is warmed, it shrinks to form a tight seal (*bottom*). Ribs machined into the sleeve enhance seal by "biting" into tubes.



SIMPLE WINDOW OPENER that would be suitable for a greenhouse is actuated by a spring fabricated out of a copper-rich shape-memory alloy. At temperatures below about 18 degrees C. (65 degrees Fahrenheit) the spring is fully contracted and the window remains closed. When the temperature rises above 18 degrees, shape-memory spring overcomes the force of a bias spring and begins to open the window. At 25 degrees the actuating spring is fully extended.

for an engine operating on a modest temperature differential of 20 degrees C. corresponds to a fifth of the standard Carnot thermal-efficiency figure.

The early Banks engine operated with wire loops of Nitinol that moved up and down on spokes attached by a bell crank to a wheel. As the wheel turned, the loops dipped first into cold water and then into hot water, causing the loops to open and close. The device resembles a multipiston radial engine. In another arrangement strips of a ternary memory alloy open and close like an accordion as they are heated and cooled, giving rise to an axial motion that can be converted into a rotary motion by a crank as it is in the Banks engine.

One of the latest engines demonstrated by Banks makes use of the walking-beam principle of the 18th-century English engineer Thomas Newcomen. The "piston" consists of 88 lengths of Nitinol wire about 15 inches long. As the wire drive is alternately submerged in hot and in cold water the contraction and expansion of the wire is transmitted through the walking beam and transverse links to an output wheel. A secondary crank arm (sometimes called the Charleston because it resembles the dance popular in the 1920's) controls the dwell time in the two water baths, much as a variable linkage controls the valve action in a steam engine.

Given the low efficiency of the system, one may well ask what kinds of source could supply low-grade heat in quantities sufficient to generate useful power. One source is of course the ocean, particularly in the Tropics, where the year-round temperature difference between the sun-heated surface and the cold deep water is at least 25 degrees C. Another large potential source of low-grade thermal energy has recently been pointed out by J. S. Cory. He has observed that behind every tall hydroelectric dam there is a substantial temperature gradient from the top of the reservoir to the bottom, varying from a few degrees in the winter months to 18 degrees in the summer. Surprisingly, the thermal energy in such a large body of water is far greater than the energy potential of the gravitational head. The ability to extract this energy, as with the ocean temperature difference, calls for some form of heat engine. The shape-memory engine is a possible candidate.

Like many phenomena discovered earlier, shape memory is a solution looking for problems. I have mentioned a few applications that have already been commercially realized. Between them and the more speculative applications I have also mentioned are new devices now in the experimental stage, ranging from shallow-well oil pumps to a wide variety of automatic-control devices. One may be sure there will be more of them.

A tale (with numerous threads) from the manufacturing chemist business

1 We are pleased to announce that diethyl 4,4′-azoxydicinnamate is *again* available from lab suppliers as KODAK Laboratory Chemical No. 10086. Below 136°C it is an ordinary crystalline solid. Above 260°C it is an ordinary liquid. Between those temperatures its long, flat, rather rigid molecules arrange themselves:

not like this

NEMATIC MESOPHASE

and not like this

CHOLESTERIC MESOPHASE

but like this.

SMECTIC MESOPHASE

That makes it a smectic liquid crystal.

2 In May 1978 we incinerated our entire previous stock of No. 10086. On purpose. Wasn't selling well enough to keep in inventory.

3 A bit earlier the same year, *Journal of Chromatography* published a paper by Ralph Lester of the Tropical Products Institute in London about lepidopterous sex pheromones (*156*, 55-62). Ignored it. Not our sort of thing. Never noticed that it mentioned diethyl 4,4′-azoxydicinnamate.

4 Liquid crystals have been known since 1888. Spurts of interest over the years. Recent spurt in the early '70s. Digital displays for watches and calculators, thermometry. Review paper appeared in *Eastman Organic Chemical Bulletin* in 1973. Marketplace answered us with the message that we could serve best by offering certain intermediates which other manufacturers finish into commercial liquid crystal products. But hardly by the tank car.

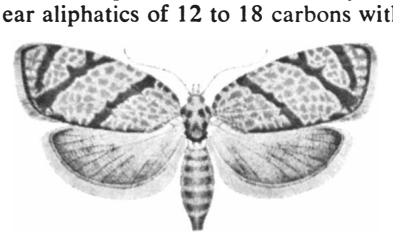
Illustrations courtesy of NYS Agricultural Experiment Station.

5 Tennessee is where we make many intermediates in tank-car quantities for many sectors of the chemical industry, including manufacturers of organic pesticides.

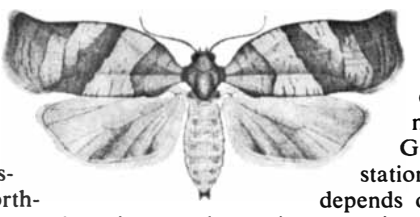
6 If all production of pesticides were to cease forthwith, more people would soon be going to bed with empty tummies. (Or at least meat consumption would rise. Insect meat, that is. Many dislike it.) But one of the unintended effects of pesticides is the way they speed up evolution of resistant insect strains.

7 Hope has arisen of reducing populations of insects that rob us of food, fiber, and lumber by disrupting the reproductive cycle through deception with their own sex attractants. Each species must have its own exclusive signals from female to male. That's what has made the existence of the species possible. From the human viewpoint the perfect specificity is the beauty part. Magic bullets to hit only the bad guys. Cut 'em way back but don't quite eliminate them.

8 These pheromones are mostly linear aliphatics of 12 to 18 carbons with one or two olefinic linkages. Encoding for species often involves only position and geometry of the double bonds. Synthesis of attractants of target species worth high-grade chemical effort. But first you have to know your target's code. Many, many targets.



9 It occurred to Brenda F. Nesbitt of the Tropical Products Institute that smectic liquid crystals might differentiate isomeric solutes by gas-liquid chromatography on the basis that differences in steric configuration influence penetration into a liquid crystal lattice—an entirely different separation mechanism from GLC with isotropic stationary phase, which depends on differences in electronic properties.



10 Dr. Nesbitt's colleague Dr. Lester proved that our No. 10086 on flux-calcined diatomite could make her idea achieve quicker and better analysis of geometric and positional isomers of these compounds than reported from attempts with other chromatographic approaches.

11 Eager readers of Lester's paper learned to their disappointment that No. 10086 was no longer available.

12 It is now.

The chemical business is very interesting. Also useful to have around.

1880 Kodak 1980

SCIENCE AND THE CITIZEN

Quark Glue

Among the many whimsical terms introduced into the physics of elementary particles in the past two decades—such as "quark," "color" and "charm"—perhaps the most aptly chosen is "gluon." The gluon is conjectured to bind together the particles called quarks, thereby forming protons, neutrons, pions and all the other entities that are classified as hadrons, or strongly interacting particles. The adhesive strength of the gluon is thought to be so great that a quark cannot be extracted from a hadron no matter what force is brought to bear on it. Moreover, the gluon itself also seems to be permanently confined: just as the quarks cannot be prized apart, it is impossible to squeeze a drop of the glue out from between the quarks. In spite of this recalcitrance there is substantial evidence for the existence of quarks. Now a series of experiments has also provided preliminary evidence for the existence of gluons.

In some respects the gluon is similar to the photon, the quantum of electromagnetic radiation. Both particles are massless and move at the speed of light, and both can be regarded as the agents of fundamental forces in nature. The photon mediates the electromagnetic force. For example, the attraction between the electron and the proton in a hydrogen atom can be visualized as resulting from the continual exchange of photons, which are emitted by one particle and absorbed by the other. Only particles that have an electric charge can emit or absorb a photon, and so only those particles are subject to electromagnetic interactions.

The force transmitted by the gluon is less familiar than electromagnetism, and its effects are more complicated. It is called the strong force or the color force, and it is governed by the property of quarks called color. (The quark colors, of course, have nothing to do with the colors of everyday life.) Whereas there is only one kind of electric charge, which can have positive and negative values, there are three kinds of color charge, each of which also has two possible values. Quarks are said to have the colors red, blue and green; the corresponding antiquarks are antired, antiblue and antigreen. According to a fundamental tenet of the theory of quark interactions hadrons can be made up only of certain combinations of quark colors, which are designated white (or colorless) combinations. The proton and many hadrons with a similar structure consist of three quarks, one quark in each of the colors. The pion and the particles related to it are made up of one quark and one antiquark, with colors that exactly cancel; for example, the quark might be red and the antiquark antired.

Just as electrically charged particles can be bound to each other by the exchange of photons, so the quarks in a hadron are cemented together by the exchange of gluons. There remains, however, a further complication. Although the photon carries the electromagnetic force, it is itself electrically neutral; when an electron emits a photon, the electron gives up energy and momentum but its electric charge is not altered. The gluon, on the other hand, bears a color charge. Indeed, there are eight species of gluon, with different combinations of colors and anticolors, combinations in which the color charge does not cancel. One kind of gluon, for instance, has the colors red and antiblue; another is green and antired. One consequence of these color charges is that gluons can interact with one another as readily as they do with quarks. Another consequence is that when a quark emits a gluon, the color of the quark is changed. If a red quark emits a red-antigreen gluon, then in order to keep the total color charge of the two particles unchanged the quark must be transformed into a green quark.

Because gluons, like quarks, cannot be observed directly the evidence for their existence must be inferred from observations of ordinary hadrons. The accelerator experiments from which this evidence has emerged begin with electrons and their antiparticles, positrons, which are stored in counterrotating beams and made to collide head on. Electrons and positrons are not hadrons, and they are not made of quarks, but when they collide, they are both annihilated to yield a state of pure energy, from which a quark and an antiquark can materialize. Of necessity the quark and the antiquark have opposite colors.

At low energies the electron-positron collisions most often give rise to a set of particles that fly away from the point of impact in all directions. As the energy is increased another kind of result becomes commoner. Two jets of particles, most of them hadrons, are arranged back to back around the collision point. The interpretation of these events is straightforward: as the quark and the antiquark move apart in opposite directions, other quark-antiquark pairs are created in their vicinity. The quarks and antiquarks then bind together to form hadrons, some of which can subsequently decay to yield still more hadrons and other particles. All the particles that are descendants of the original quark (or antiquark) continue moving in approximately the same direction, and so they form a well-focused jet.

Such two-pronged jets have been observed at several laboratories and have provided information about the quark structure of hadrons. The recent evidence for gluons comes from still more elaborate events with three jets.

When an electron and a positron collide violently, the quark and the antiquark are created with high kinetic energy. At least one of the particles is then likely to give up some of its energy by emitting a gluon. (The analogous electromagnetic process is called bremsstrahlung, or braking radiation, where a high-energy electron emits a photon.) The emission of the gluon changes the color of the quark, but the total color of the quark, the antiquark and the gluon still cancels to zero.

The emission of the gluon also alters the geometry of the observed event. Like the original quark and antiquark, the gluon induces the creation of quark-antiquark pairs in the surrounding vacuum, so that the gluon effectively decays to yield ordinary hadrons. Thus the decay products of the gluon form a third jet, and the overall distribution of particles has three prongs, all of which lie in a single plane.

Evidence for three-jet events has been found in the aftermath of electron-positron collisions at PETRA, a new particle-storage ring in Hamburg. Four detectors at PETRA, each detector operated by a separate group of investigators, have been employed in the search for such events. The first detailed analysis of the results has been reported by a group of 57 physicists from West Germany, the Netherlands, China and the U.S. The group is led by Samuel C. C. Ting of the Massachusetts Institute of Technology.

The electrons and positrons stored at PETRA can be made to collide with energies of up to about 30 billion electron volts, which is roughly three times the energy available in the largest of the earlier storage rings. Most of the events recorded so far, however, are at energies only a little above the threshold where gluon emission becomes likely. As a result there are comparatively few three-prong events, and in most of those the gluon diverges only slightly from the quark that emitted it. Hence the third jet cannot be clearly distinguished. What is actually observed looks much like a two-jet event, but one jet is more oblate than the other, indicating that it consists of two unresolved streams of particles. As additional data are collected the form of the events should become less ambiguous.

Along with the new results comes another whimsical entry in the physicist's lexicon. The three-pronged distribution of particles has given rise to the term

THE FIAT STRADA.
IT SQUEEZES 28 MILES* OUT OF EVERY GALLON WITHOUT SQUEEZING YOU.

It took a lot of automotive art and advanced technology to do it, but the 1979 Fiat Strada has it. The room of a compact and the mileage of a subcompact.

28 EST MPG / 41 EST HWY

*EPA estimates. Remember: This estimate is for comparison purposes. Your mileage may be different depending on your speed, trip length, and weather. Highway mileage will probably be less. Estimates vary in California.

That's more miles per gallon than the gas-engine Rabbit, Honda Accord, Omni or Horizon give you. And you get more room in a Strada than they give you, too. Along with front-wheel drive, a gas-saving 5th gear standard, and the kind of comfort you'd never expect to get in an economy car.

And the Strada gives you something else none of those other cars do. A 24 month/24,000 mile limited power train warranty.† One twice as long as most economy cars.

24 MONTH → 24 THOUSAND MILE WARRANTY
Power Train Warranty

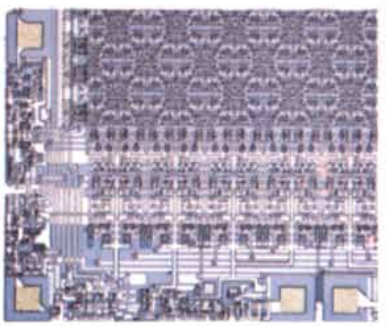
For the name of the Fiat dealer nearest you, call toll-free (800) 447-4700 or in Illinois (800) 322-4400 and test-drive the Strada today.

FIAT

STRADA. ANOTHER ITALIAN WORK OF ART.

†There are certain limitations and exclusions. See your dealer for details.

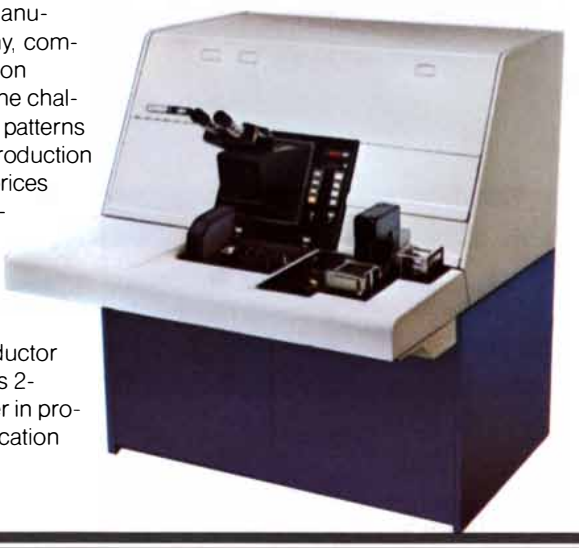
High throughput and yield in integrated circuit manufacture



The crux of microcircuit manufacture is the printing of tiny, complicated electrical patterns on photosensitive materials. The challenge is to reproduce these patterns crisply and accurately, in production quantities, at competitive prices This is exactly what Perkin-Elmer Micralign mask aligners do.

The newest Micralign, the Model 200 Series, is responsive to the semiconductor industry's needs. It achieves 2-micron geometries or better in production, distortion/magnification tolerance of .25 micron and four percent uniformity of illumination.

Like other Micralign systems, the Model 200 Series uses the 1:1 optical projection concept, pioneered by Perkin-Elmer, to focus light from a mask. The mask contains many repeats of a single pattern. In one fast scan the Model 200 Series exposes all of these patterns onto a silicon wafer coated with a photosensitive material. After development, the wafer is further processed—etched, doped, and recoated. By repeating this procedure a number of times circuits are built, layer on layer, on a single wafer. Finally the wafer is cut to separate each individual circuit.



Because dust, heat and vibration are major enemies of precise projection, the 200 houses its optics in a quiet, clean, wear-free world of their own. Vibration is minimized by two frames, one inside the other. All vibrating components have been mounted to the outer frame, thereby isolating the projection optics from all sources of vibration.

The 200 Series has a built-in environmental chamber. External air is blown through the top of the unit. The air is filtered and temperature regulated. A positive-pressure, Class 100 environment is carefully controlled to better than 1°F.

Options include automatic wafer loading and a revolutionary automatic alignment system. Deep UV coatings for even smaller geometries will be available soon.

"Sputtering" thin films onto microcircuits quickly, uniformly, economically

The thin metal films used in making semiconductors are deposited by sputtering systems. Evacuate a chamber; fill it with argon; place a high-potential cathode on the ceiling of the chamber and a lazy susan loaded with silicon wafers on the floor; attach to the cathode a target of the material to be deposited; turn on the current.

That, in essence, is sputtering. The current ionizes the argon, and the argon ions in their efforts to reach the cathode bombard the target material, knocking off atoms which settle, uniformly dispersed, on the wafers circling below. Layer by layer, the film builds up to the desired thickness.

Perkin-Elmer sputtering equipment has long set industry standards worldwide. Our latest unit, for example, the 4410, features a novel delta-shaped DC magnetron cathode and microprocessor controller. Both

play an important role in producing the high quality films and high throughput essential for the economic production of semiconductors.

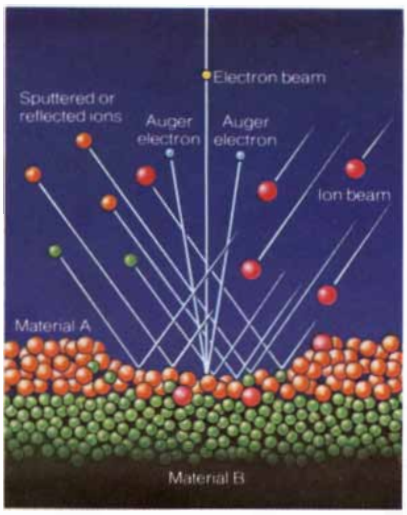
The Delta™ cathode deposits aluminum alloy and other metallic films at very high rates. This is important for building dependable microcircuits today. It's even more important for the high density semiconductor devices of the not-too-distant future.

The microprocessor automatically controls all process variables and eliminates human error. It thus assures run-to-run repeatability for high yield and throughput.

All of which help make the 4410's return-on-investment better than that of any competitive system.

Probing semiconductor quality layer by atomic layer

The leading tool for analyzing micro-electronic surfaces is the Perkin-Elmer 590 Scanning Auger (pronounced Oh-djay) Microprobe, familiarly known as SAM.



Three-dimensional materials characterization using Auger analysis with ion sputter-etching.

Typical tasks for SAM include evaluating the integrity of a semiconductor's thin film interfaces and checking to be certain impurities were not introduced during manufacture. Or detecting such imperfections as microscopic aluminum spikes between semiconductor layers which can cause short circuiting. Or scrutinizing bonding pads for trace contaminants which cause poor bonding adhesion.

To do such jobs, SAM bombards a specimen with a beam of electrons, causing the emission of X-ray photons and chemically specific Auger electrons. These Auger electrons originate in the topmost two or three atomic layers of a specimen surface. By measuring the kinetic energy and number of Auger electrons emitted, SAM provides a quantitative as well as qualitative identification of surface constituents. If the electron beam is scanned, SAM can map the distribution of chemical elements over a selected area.

SAM incorporates an ion bombardment gun that continuously erodes an area of the surface so that the microprobe can analyze downward, layer by atomic layer. This makes possible a true three-dimensional analysis of thin films.

Where applications require, SAM can be expanded into a multiple-technique instrument through addition of ESCA (Electron Spectroscopy for Chemical Analysis), which measures electrons released due to X-ray stimulation of a surface, and SIMS (Secondary Ion Mass Spectroscopy), which detects ions emitted after primary ion bombardment.

For more information on these products, please write: Corporate Communications, Perkin-Elmer, Main Ave., Norwalk, CT 06856.

PERKIN-ELMER
Responsive Technology

Mercedes events, after the well-known trademark and hood ornament of the Mercedes-Benz automobile.

CT by NMR

The introduction of computerized tomography (CT) within the past five years has given physicians an important new degree of access to what is going on inside a patient's body. In conventional radiography X rays from a single source diverge to cast on film a shadowgraph of all the structures along the path of the rays; overlapping structures and those with similar densities are not well distinguished. A CT scan, on the other hand, is a reconstruction from a large number of projections. The X-ray source and detector move around the body; in effect hundreds of X-ray pictures are made in a few minutes, but instead of being cast onto film the information is processed by a computer to yield a two-dimensional outline of discrete structures along a single plane through the body. The CT scan is invaluable for the diagnosis of brain tumors and many other pathologies, but it has one disadvantage: it depends on ionizing radiation, too much of which can be harmful to the patient. Now a new technology may be on the horizon that promises to provide images comparable to those of the CT scanner but without ionizing radiation. It is nuclear-magnetic-resonance (NMR) imaging, which depends on the magnetic properties of certain atomic nuclei in a tissue rather than on the tissue's attenuation of X rays.

NMR spectroscopy has become a well-established technique for investigating the structure and composition of matter in the past 30 years. It exploits the behavior of spinning protons (the nuclei of hydrogen atoms) and certain other nuclei when they are placed in a magnetic field. The nuclei, acting like tiny magnets, line up with the field; more precisely, they precess around the direction of the field much as a spinning top wobbles around a gravitational axis. The alignment of the nuclei with respect to the field is perturbed by the application of a radio frequency such that the nucleus is raised to a higher energy level; in effect the nucleus flips over into a direction opposite to that of the field. When the applied radio frequency is turned off, the nucleus "relaxes" back into the direction of the field, emitting energy, which is recorded as a sharp peak in a spectrum. The frequency of the energy emitted is characteristic of the nucleus and is modified by its molecular environment, yielding information on the structure and environment of complex molecules.

The shape, time of appearance and duration of the peak are further modified by the nature of the surroundings—in the case of a tissue, by its density or water content, for example. To produce an image of a tissue, inhomogeneous magnetic fields are imposed and varied to introduce a directional component, so that information comparable to that of many NMR spectra can be collected, synthesized and analyzed by a computer program. In principle this information can provide data in three dimensions, which can be projected in any desired two-dimensional plane.

The availability of high-strength magnets has already made it possible to extend NMR spectroscopy from the analysis of physical and chemical structure to the tracking of biochemical reactions (see "Science and the Citizen," September), and at the Harvard Medical School a group headed by Eric T. Fossel and Joanne S. Ingwall has been applying the technique to study such reactions in isolated organs of experimental animals. Several groups in Britain, notably those at the University of Nottingham and at the University of Aberdeen, have been developing imaging methods and applying them in studies of living animals and parts of the human body.

The Red and the Blue

The measurement of Doppler shifts in the spectra of celestial objects is one of the astronomer's most powerful tools. A shift of spectral lines to longer wavelengths, or toward the red end of the spectrum, tells him that the observed object (or dust or gas) is receding along his line of sight. A shift toward the blue end of the spectrum indicates the source is approaching. On the grand scale astronomy has been dominated by red shifts. Thus the steady increase in red shift with distance exhibited by galaxies bespeaks an expanding universe. Only a few nearby galaxies show blue shifts. Within our own galaxy stars exhibit a random mixture of small red and blue shifts. The Doppler shift is particularly useful for detecting binary star systems. When the spectrum of a source shifts periodically from slightly blue to slightly red, one can infer that the source is moving in an orbit alternately toward and away from the solar system.

Within the past year astronomers have been fascinated by the discovery of a peculiar object within our galaxy that exhibits a blue shift more than 100 times larger than any previously known, a shift comparable, although opposite in sign, to the red shifts of the nearer quasars, extragalactic objects generally thought to be enormously distant. The mystery presented by the new source is only slightly diminished by the observation that the blue shift in its spectrum is simultaneously matched by a red shift of equal magnitude. Therefore the source itself is neither approaching nor receding with the velocity that would be indicated by an isolated shift toward the blue or the red. The most plausible explanation for the dual shift is that the source is a neutron star, probably a kind of pulsar, that is emitting two beams of particles at high velocity in opposite directions.

The beam emitter was first noticed as a peculiar star of the 14th magnitude in the mid-1960's and was subsequently designated SS433. It lies about 11,000 light-years away in the constellation Aquila, four degrees north of the celestial equator. In the 1970's SS433 was identified as a source of both radio waves and X rays. As part of a program for studying X-ray-emitting stars Bruce Margon of the University of California at Los Angeles and a group of six colleagues began spectroscopic observations of SS433 with the three-meter Shane reflector at the Lick Observatory in August, 1978. "The first few spectra we took," says Margon, "told us there was something extremely bizarre." In a series of spectra made on eight consecutive nights one emission line shifted toward the red in virtually smooth steps from a wavelength of 7,400 angstrom units to one of 7,620 angstroms even as a second line shifted toward the blue from 6,120 angstroms to 5,970. The changes in wavelength correspond respectively to an increase in recession velocity of 9,000 kilometers per second and an increase in approach velocity of 7,400 kilometers per second. The emission lines are those of hydrogen and helium atoms at a temperature of about 10,000 degrees Kelvin.

Subsequent studies indicated that the emissions arise from oppositely directed rotating beams in which particles are accelerated to 81,000 kilometers per second, or 27 percent of the velocity of light. The beams sweep out two conelike patterns in a period of about 164 days. The common axis of the two cones intersects the line of sight from the solar system at an angle of 78 degrees. The vertex angle of each cone is about 34 degrees, and it seems to vary by a few degrees. Because the beams are viewed from the side the peak observed velocities along the line of sight are much less than the true velocities: about 32,000 kilometers per second toward the solar system for the primarily "blue" beam and about 54,000 kilometers per second away from the solar system for the "red" beam.

During a small part of each 164-day cycle the blue beam is aimed slightly away from a right angle to the line of sight and is therefore briefly red-shifted. Similarly, the red beam is aimed slightly forward of a right angle for the same length of time. Where one might expect the red beam to be blue-shifted in that part of its cycle it continues to exhibit a red shift because of relativistic time dilation, which gives rise to what is called a transverse Doppler shift. At the velocities at which the particles are moving they show a constant red shift equivalent to a velocity of 11,000 kilometers

per second. This offset of 11,000 kilometers per second applies, of course, to both beams at all times and so more than cancels the expected blue shift of the red beam.

It has been suspected that SS433 is a binary star system. Recently David Crampton, Anne Cowley and John Hutchings of the Dominion Astrophysical Observatory in Canada have confirmed this suspicion by refined Doppler measurements. The source of visible light is in orbit around a normal star too faint to be photographed. SS433's visible, X-ray and radio emissions evidently arise from an accretion disk around a neutron star, probably a supernova remnant. Such a disk consists of matter being transferred from a normal member of the binary pair to a highly collapsed member: the neutron star. I. S. Shklovsky of the Institute of Space Research of the Academy of Sciences of the U.S.S.R. has suggested that mass is being transferred to the neutron star faster than it can be absorbed. By some mechanism the neutron star, perhaps a rapidly spinning young pulsar, repels part of the infalling mass outward in two narrow streams, giving rise to the observed beams.

Homo erectus in the Pyrenees

For a decade French prehistorians have worked at Arago Cave, a site 20 kilometers northwest of Perpignan near the village of Tautavel in the eastern Pyrenees. They have unearthed an abundance of stone tools and animal bones and the partial remains of at least 10 members of the genus *Homo:* three children, one old woman, two young men and four other adults whose sex is not known. The remains are some 450,000 years old, which makes them the oldest hominid fossils found in Europe.

With the recent opening of a museum in Tautavel to display the results of the Arago Cave work the chief investigator, Henry de Lumley of the University of Aix-Marseille, has summarized the results. His report identifies one of the excavated specimens, Arago XXI, as a fossil of *Homo erectus,* the predecessor of *Homo sapiens.* The specimen is a face and forehead. The wear pattern of the teeth and the condition of the suture joining the frontal and parietal bones indicate that the skull belonged to a young male between 18 and 25 years old.

Arago XXI has a projecting upper jaw and heavy brow ridges; the lower jaw was not found. De Lumley estimates the capacity of the cranium as being 1,050 cubic centimeters. (The average for modern man is 1,360 c.c.) He finds that the specimen is anatomically distinct from specimens of *Homo erectus* found in Africa and Asia. He has suggested that Arago XXI be assigned to a new subspecies: *Homo erectus tautavelensis.*

The great age of Arago XXI is ap-

Chevrolet Monte Carlo for 1980. You've never seen the likes of it before.

An impressive statement, you say?

Slip behind the wheel of the new Monte Carlo for 1980. Run your hand along its rich new look of vinyl wood-grain. Take in its elegant new upholsterings.

Then turn the key. Feel yourself propelled by a new standard 3.8 Liter (229 Cu. In.) V6 engine* with greater power output than last year's standard engine. A sensation you can multiply with the new available turbocharged V6 for 1980.

And, as you drive, remind yourself that automatic transmission and power steering are standard this year. And that it all comes together in a Monte Carlo with impressive EPA estimates for 1980.

CHEVROLET MONTE CARLO	EPA EST. MPG	EST. HWY. MPG
Std. 229 Cu. In. V6	19	26
Avail. V6 Turbo*	18	25

REMEMBER: Compare these estimates to estimated MPG of other cars. Your mileage may vary depending on speed, trip length and weather. Actual highway mileage will probably be less than the estimated highway fuel economy. Monte Carlo is equipped with GM-built engines produced by various divisions. See your dealer for details.

*Not available in California where a 231 Cu. In. V6 is standard. California EPA estimates lower.

The Chevy price makes it look even better.

Which is perhaps the most exciting aspect of Monte Carlo's impressive new features and driving delights for 1980.

They all come to you for the price of a Chevrolet.

See your Chevy dealer about buying or leasing the new Monte Carlo.

You've never seen a Monte Carlo quite like it before.

Monte Carlo's new Special Custom interior available for 1980. And underneath that, the quiet comfort of a cushion-mounted full perimeter frame.

The rich new look of vinyl wood-grain accents on the instrument panel to make Monte Carlo's new easier-to-read instruments look even better to you.

TURBO

New available 3.8 Liter turbocharged V6 engine for passing acceleration on demand.

No other car looks quite like our Chevrolet Monte Carlo. Impressive new grille, new dual rectangular headlights continue that distinction for 1980.

Underneath it all, a highly tuned sport suspension system for a solid feel of the road and nimble agility around curves and in and out of traffic.

GM

Chevrolet

RCA INTRODUCES
The next step in video

This is the one. SelectaVision 600. Now, the video cassette recorder you've been waiting to buy has more of what you want most.

Like more time. The 600 has a new third speed that allows you to record up to six hours on a single cassette. That means fewer cassette-changing interruptions or missed endings when you're recording with the automatic timer. And more shows on one cassette means you save money on tape too.

Third speed gives 6 hours recording on a single VK-250 cassette.

But 6 hours is just one reason to consider the new SelectaVision 600. A 7-day programmable timer lets you record your favorite shows even when you're away. By touching a few buttons, you can program your selections into the timer and the rest is automatic. The 600 will turn itself on at kickoff time, silently record the game, then turn itself off. When it's time for your second selection, the 600 turns itself to the proper second channel and starts recording again – automatically. The entire schedule is preset by you up to a whole

7-day memory timer lets you program shows up to a week in advance.

SELECTAVISION 600. cassette recording.

6 hours recording. 7-day memory timer.

week in advance – as many as four different shows or even the same program for seven straight days. And with 6 hours of recording time, you shouldn't have to change cassettes as often.

6 hours recording – 7-day memory timer. Those are the highlights of the new 600, but there's more.

A new remote pause/channel control not only lets you edit out unwanted programming from 20 feet away but also lets you change channels – even if your present television set isn't equipped for remote control.

An ingenious tape indexing system lets you skip to the beginning of the next show by just pushing the fast forward control. The 600 stops automatically.

And then there are all the features that have helped make RCA SelectaVision one of the best selling VCR's in America. Like "touch of a button" Varactor tuning. A direct-drive motor for precise operation. Special circuitry that automatically compensates for changing signal strength. It's all there in the new 600. Go see it at your RCA Dealer.

He'll show you all the things you can do with the new SelectaVision 600. The six-hour video cassette recorder with the 7-day memory.

RCA

CAUTION: The unauthorized recording of television programs and other materials may infringe the rights of others.

Let RCA turn your television into SelectaVision®

proximately the same as that of the *Homo erectus* specimens from Peking. Their age has recently been reestimated and set at about 460,000 years. Peking Man used fire, but no evidence of fire has been found at Arago Cave. The animal bones unearthed at the site suggest that the principal game animal was the horse. Reindeer, deer, bison, cattle and goats were also eaten, and these ancient hunters even managed to kill rhinoceroses and elephants.

Grand Prix

Anyone who has mused on the behavior of drivers in countries other than his own is aware that people in different countries drive differently. Such differences are also reflected in characteristic national rates of death caused by motor-vehicle accident. The current picture is discussed in the *Statistical Bulletin* of the Metropolitan Life Insurance Company. Presenting the annual death rate from motor-vehicle accidents in three ways (per 100,000 population, per 100,000 registered motor vehicles and per 100 million motor-vehicle miles), the *Bulletin* finds that "West Germany's motor-vehicle-accident mortality was high in relation to all three measures," whereas "Japan, Great Britain and Norway had very low rates by all measures." The U.S. had the lowest fatality rate per 100 million vehicle miles and the second-lowest per 100,000 registered vehicles, but the rate per 100,000 population was moderately high.

The report compares the various rates for 13 countries in 1965–66 and 1975–76. Over that decade fatality rates declined in every country by the measures of registered vehicles and miles driven; they increased slightly per 100,000 population in France and Norway and decreased somewhat in the other countries. (The customary way to measure the hazard from motor vehicles has been in terms of the death rate per 100,000 population, but that measurement fails to take into account either the extent of motor-vehicle ownership or the degree to which motor vehicles are driven.)

The most remarkable declines by all three measures were in Japan: from 14.2 deaths per 100,000 population to 9.2; from 89.7 to 27.1 per 100,000 registered vehicles, and from 23.4 to 4.8 per 100 million vehicle miles. The figures for the U.S. in 1975–76 were 21.7 fatalities per 100,000 population, 33 per 100,000 registered vehicles and 3.4 per 100 million vehicle miles. In West Germany, which was at or near the upper extremes, the figures were respectively 24.1, 68.1 and 7.9. Other countries that were on the high side were France (25.5, 54.8 and 8.6) and Italy (17, 58.5 and 7.2).

Mushroom News

Walter Litten, the SCIENTIFIC AMERICAN correspondent on poisonous mushrooms, sends the following intelligence:

A myth with lethal consequences has long persisted. Many people still believe poisonous mushrooms blacken silver boiled with them and can thereby be told from harmless species. When told that this is not so, they ask for a more reliable test and are disheartened to learn there is no quick and easy way to make the distinction among the hundreds of species of mushroom that can be encountered.

Since the pioneering studies of Theodor Wieland and his collaborators at the Max Planck Institute for Medical Research in Heidelberg it has been generally accepted that the deadly ingredients of the deadly mushroom species are cyclopeptides: rings of amino acid units. The chromatographic procedures for demonstrating the presence of cyclopeptides are scarcely feasible for mushroom hunters inexperienced in analytical chemistry. Now, however, Axel Meixner has published in *Zeitschrift für Mykologie* an extremely simple test for

amatoxins, cyclopeptides of eight amino acid units that attack the liver and kidneys of careless eaters of wild mushrooms by inhibiting an RNA polymerase enzyme. The procedure is as follows: (1) With a garlic press squeeze a drop of juice from the fresh mushroom onto a piece of newsprint. If only a small piece is available for testing, enough juice may be absorbed into the paper merely by mashing. Mark the wet spot with a pencil. (2) Dry the spot by gentle heating. (3) Put a drop of concentrated hydrochloric acid on the dried spot. (4) If amatoxin is present, a blue color will appear within one minute to 20 minutes, depending on whether the amatoxin content is high or just a trace.

High-quality cellulose paper will not work, because the test somehow depends on an acid-catalyzed reaction with the lignin left in inexpensive pulp paper. With less dangerous mushrooms containing tryptamine, such as *Amanita citrina,* a red color appears immediately and turns bluish in half an hour. The toxicology committee of the North American Mycological Association warns that a negative result does not constitute proof of edibility. The test does not yield a blue color for phallotoxins, rings of seven amino acids that may also be present. These cyclopeptides almost instantaneously destroy liver cells when they are injected into the peritoneum of an experimental animal, but they are now thought to play no part in mushroom poisoning by mouth.

Thadée Staron and Michel Courtillot of the French National Center for Agronomic Research believe that in mushroom tissue the cyclopeptides are themselves linked in rings around a central polysaccharide of high molecular weight and that the resulting structures in turn are joined to form even larger aggregates. Their toxicological assays convince them that these giant molecules are more toxic weight for weight than the separated cyclopeptides. They postulate that the polysaccharide by itself is not toxic but serves to hold the entire structure together until the cyclopeptides break away in the acid environment of the stomach, as they do under chemical manipulation. According to the French workers giant molecules of elevated toxicity re-form when the separated components enter the neutral environment of the intestine. From there they presumably pass into the circulatory system.

On the basis of reports that horses and many rodents are comparatively tolerant to the deadly mushrooms, Barbara Courtillot has incubated an extract of *Amanita virosa* (known as "The Destroying Angel") with homogenate of stomach muscle from horses and mice. She has found a reduction in toxicity, leading to the suggestion that something in the gastric mucosa of resistant animals binds the cyclopeptides of the dissociated ultratoxic molecule and leaves the polysaccharide to pass on to the intestine, where it is degraded. Subsequently, in this view, the cyclopeptides are gradually released from the gastric mucosa and eliminated almost harmlessly.

Staron and Michel Courtillot have tested their model by purifying the polysaccharide by chromatography and incubation with a homogenate of rat stomach tissue to remove the remaining cyclopeptide. Injected into donor animals, this nontoxic polysaccharide acts as an antigen to create antibody that can immobilize the separated polysaccharide and make it unavailable for reconstituting the ultratoxic giant molecule. To some extent serum from these animals can protect other animals. Against a dose of toxin only slightly above the minimum lethal level, protection by the purified polysaccharide can last for as long as 15 months. Protection can also be conferred by giving the serum up to 15 hours after the toxin. Conceivably in cases where human amatoxin poisoning is suspected soon enough a serum injection could make the difference between life and death.

If all you this is all

Thinsulate.® Very simply, the best synthetic insulating material ever developed for outdoor wear. An innovation of the 3-M Company that offers all the wetproof warmth of the early synthetics, without their bulk and weight. In fact, Thinsulate is almost as light as goosedown.

Insulation values, compared. One centimeter of Thinsulate vs. other materials.

THINSULATE

DOWN

POLYESTER

PILE

WOOL FELT

"Thinsulate" is a reg. TM of 3M Company

Last year, Sierra Designs was the first manufacturer to offer Thinsulate-lined clothing designed for backpackers, campers, and hikers.

Not surprisingly, our line of slim, lightweight jackets, parkas and vests was an immediate success. Also not surprisingly, our lead was immediately followed by other manufacturers. Today, there are dozens of Thinsulate-lined garments available, under a variety of labels.

want is warmth, you need.

And a variety of quality standards.

So even though we're no longer your only source of Thinsulate, we think it still makes sense to look for the Sierra Designs label. Because it's the only way to be sure you're getting Sierra Designs' tailoring, attention to detail, and meticulous stitching.

For example, every stitch in our Thinsulate-lined garments is locked in place, so it will never pull out. And every seam is sewn eight to ten stitches to the inch, because we've found that fewer stitches weaken the seam, while more stitches weaken the fabric.

We apply the same rigid standard of tailoring to the inside of a garment as we do to the outside. Not because it looks better, which it does. But because it's stronger and longer lasting.

And we inspect our Thinsulate-lined garments as carefully as we sew them. Three separate times. Before you ever try one on, every seam will be checked and trimmed. Every snap will be snapped and unsnapped. Every zipper zipped and unzipped.

These are some of the things we do that most other manufacturers don't. But we also do something no other manufacturer does.

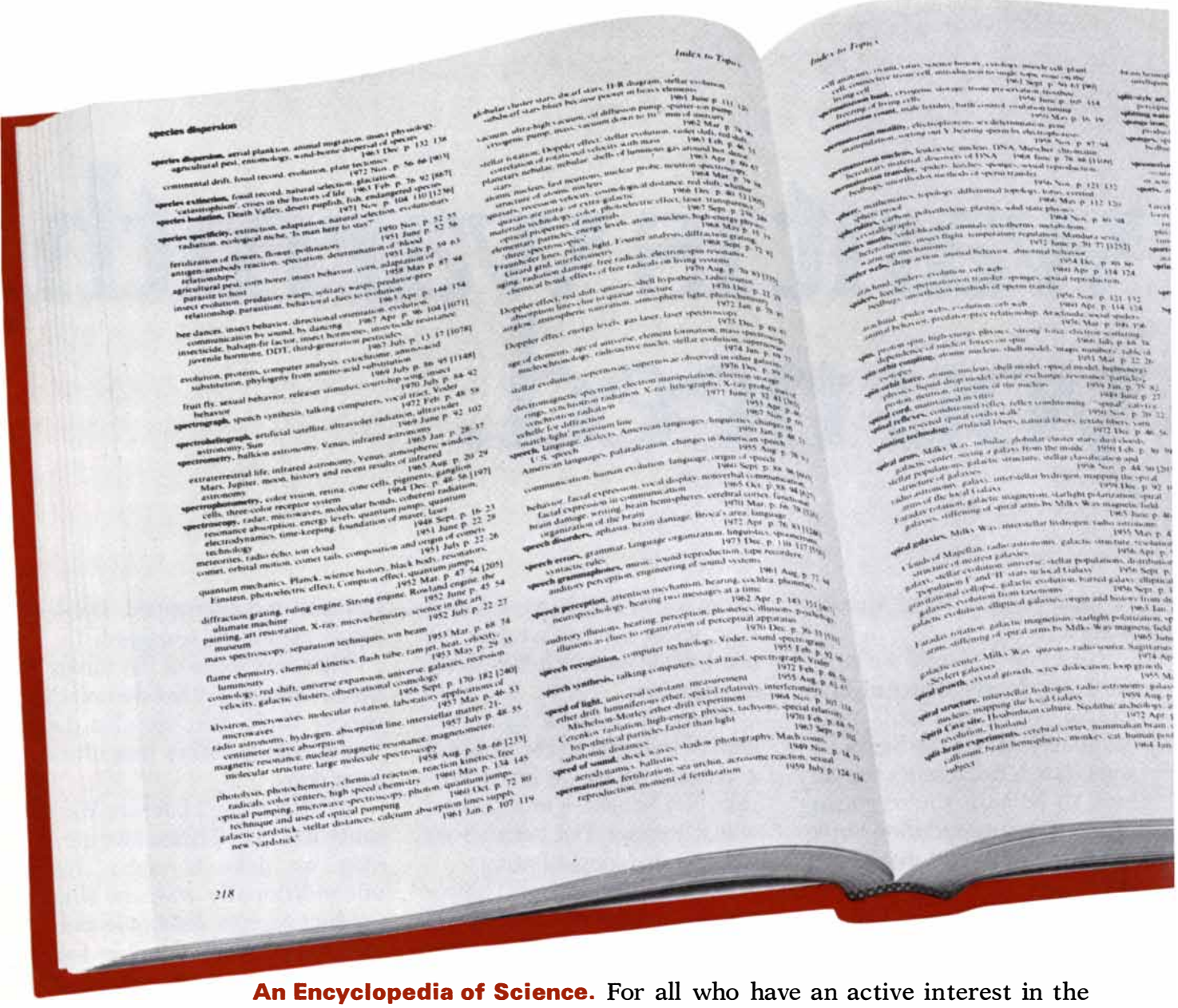
We attach a Full Lifetime Warranty to every Thinsulate garment we make. It reads: "We unconditionally warrant this product against defects in materials and workmanship for your lifetime. If any stitching ever wears out, tears loose, or breaks, return the product to us. We will repair or replace it, free of charge."

All of this means just one thing: you may not be warmer in a Sierra Designs Thinsulate-lined garment.

But you should be warm a lot longer.

For a list of dealers near you, call 800-227-2082, toll-free. Or write us at 247 Fourth Street, Oakland, California 94607.

Sierra Designs



An Encyclopedia of Science. For all who have an active interest in the work of science, the new Index transforms the collected issues of SCIENTIFIC AMERICAN into an encyclopedia of the substance and history of modern science. It is actually eight different indexes in one:

Index to Topics—a rotated key-word index of the subject matter of all articles and Science and the Citizen features (see above).

Index to Proper Names—the names of all persons mentioned in the articles and Science and the Citizen and of places and institutions featured in a primary role.

Listing of Tables of Contents—a chronological listing of the complete contents of all 362 issues that provides ready identification, by titles and authors, of the articles cited in the two foregoing indexes.

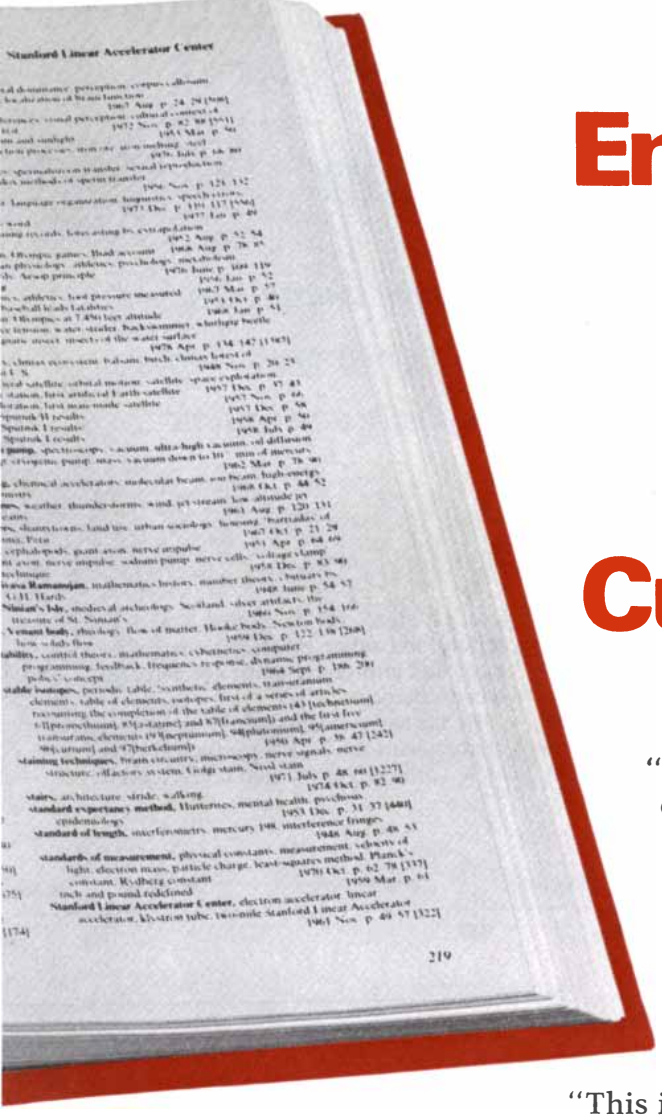
Index to Authors—an alphabetical listing of authors of all articles.

Index to Titles of Articles—an alphabetical listing of all articles by the first key word in each title, and by the other key words in the title.

Index to Book Reviews—a listing of the longer book reviews in three parts: authors of books, titles and reviewers.

Index to Mathematical Games—a listing of the puzzles, games and diversions presented in this department since its inception in January 1957, under the editorship of Martin Gardner.

Index to The Amateur Scientist—a listing of all the projects, experiments and demonstrations presented in this department from April 1952 through February 1976 under the editorship of the late C. L. Stong and from July 1977 through June 1978 under the editorship of Jearl Walker.



Enjoy direct entry to 30 years of SCIENTIFIC AMERICAN in the new Cumulative Index 1948-1978

"This kind of meticulous cross-referencing inspires confidence. Only a very persistent gremlin could lose something from SCIENTIFIC AMERICAN down a black hole . . ."

—**The New Library World**, London, June 1979

"You people sure know how to organize an index!"

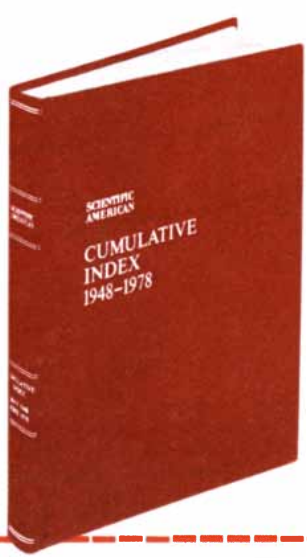
—**Donn P. Brazier, Director,**
St. Louis Museum of Science and Natural History

"This index is an invaluable complement to thirty years of SCIENTIFIC AMERICAN. Aside from the usual author and title indexes, there is a topic index—a kind of permuted key-word index—which takes the reader-researcher below the level of title index."

—**New Technical Books,**
The New York Public Library, July 1979

SCIENTIFIC AMERICAN

415 Madison Avenue, Dept. CI, New York, N.Y. 10017



Yes, send me _____ copies of the new SCIENTIFIC AMERICAN Cumulative Index to the 362 issues from May 1948 to June 1978, at the price of $45 each plus $2.50 for shipping. If I enclose payment, you will absorb the shipping charge.

☐ **My check for $_____ is enclosed, saving me $2.50 shipping charges.**

☐ **Bill me for $45 each copy, plus shipping.**

Name_____

Address_____

City_____

State_____ Zip_____

(Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York.)

11/79

The Neutral Theory of Molecular Evolution

It holds that at the molecular level most evolutionary change and most of the variability within a species are caused not by selection but by random drift of mutant genes that are selectively equivalent

by Motoo Kimura

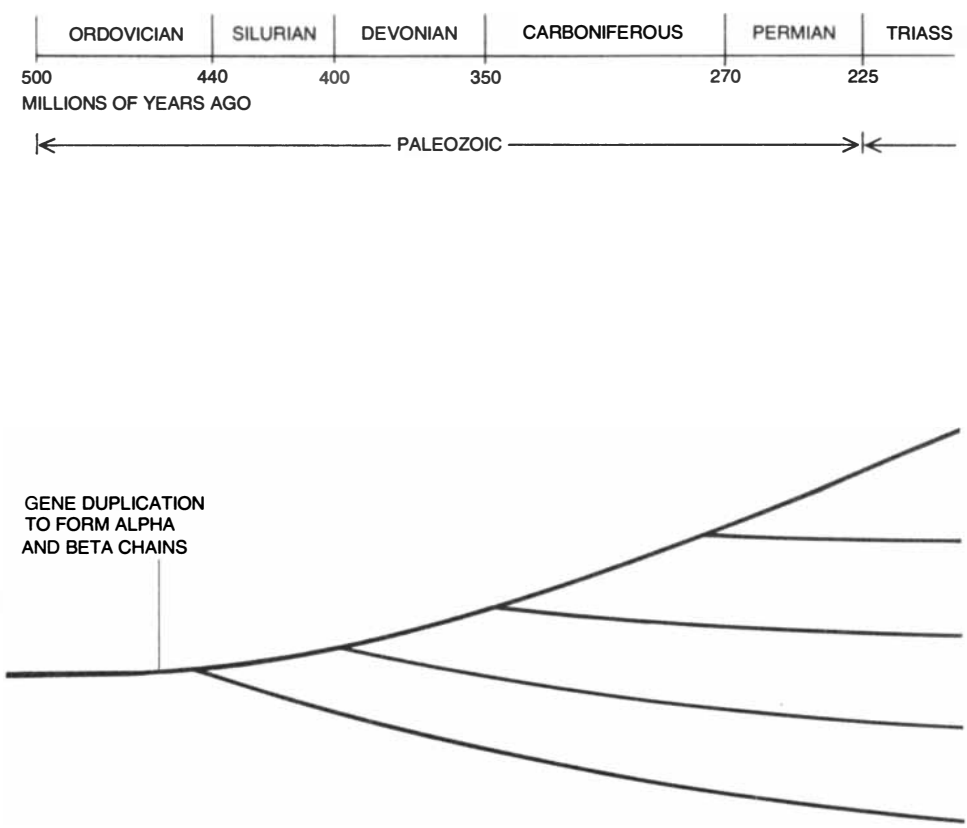
The Darwinian theory of evolution through natural selection is firmly established among biologists. The theory holds that evolution is the result of an interplay between variation and selection. In each generation a vast amount of variation is produced within a species by the mutation of genes and by the random assortment of genes in reproduction. Individuals whose genes give rise to characters that are best adapted to the environment will be the fittest to survive, reproduce and leave survivors that reproduce in their turn. Species evolve by accumulating adaptive mutant genes and the characters to which those genes give rise.

In this view any mutant allele, or mutated form of a gene, is either more adaptive or less adaptive than the allele from which it is derived. It increases in the population only by passing the stringent test of natural selection. For more than a decade now I have championed a different view. I believe most of the mutant genes that are detected only by the chemical techniques of molecular genetics are selectively neutral, that is, they are adaptively neither more nor less advantageous than the genes they replace; at the molecular level most evolutionary changes are caused by the "random drift" of selectively equivalent mutant genes.

The Evolution of Darwinism

The controversy between the neutralist view and the "panselectionist" assumption arises from the way the modern "synthetic" theory of evolution has itself evolved. When Darwin formulated his original theory, the mechanisms of inheritance and the nature of heritable variations were not known. With the rise of Mendelian genetics in this century the way was opened for efforts to supply a genetic base for Darwin's insights. This was achieved largely through the elucidation by H. J. Muller of the fundamental nature of the gene and through the methods of population genetics developed mainly by R. A. Fisher, J. B. S. Haldane and Sewall Wright. On their foundation subsequent studies of natural populations by Theodosius Dobzhansky, paleontological analyses by George Gaylord Simpson, the "ecological genetics" of E. B. Ford and his school and other investigations built a large and impressive edifice of neo-Darwinian theory.

By the early 1960's there was a general consensus that every biological character could be interpreted in the light of



PHYLOGENETIC TREE displays the evolutionary relations among seven vertebrates and shows how and when their lineages have diverged from one another over geologic time. The table at the right shows the extent to which an important protein, the alpha chain of hemoglobin, differs in the seven animals; specifically it gives the number of differences in the sequence

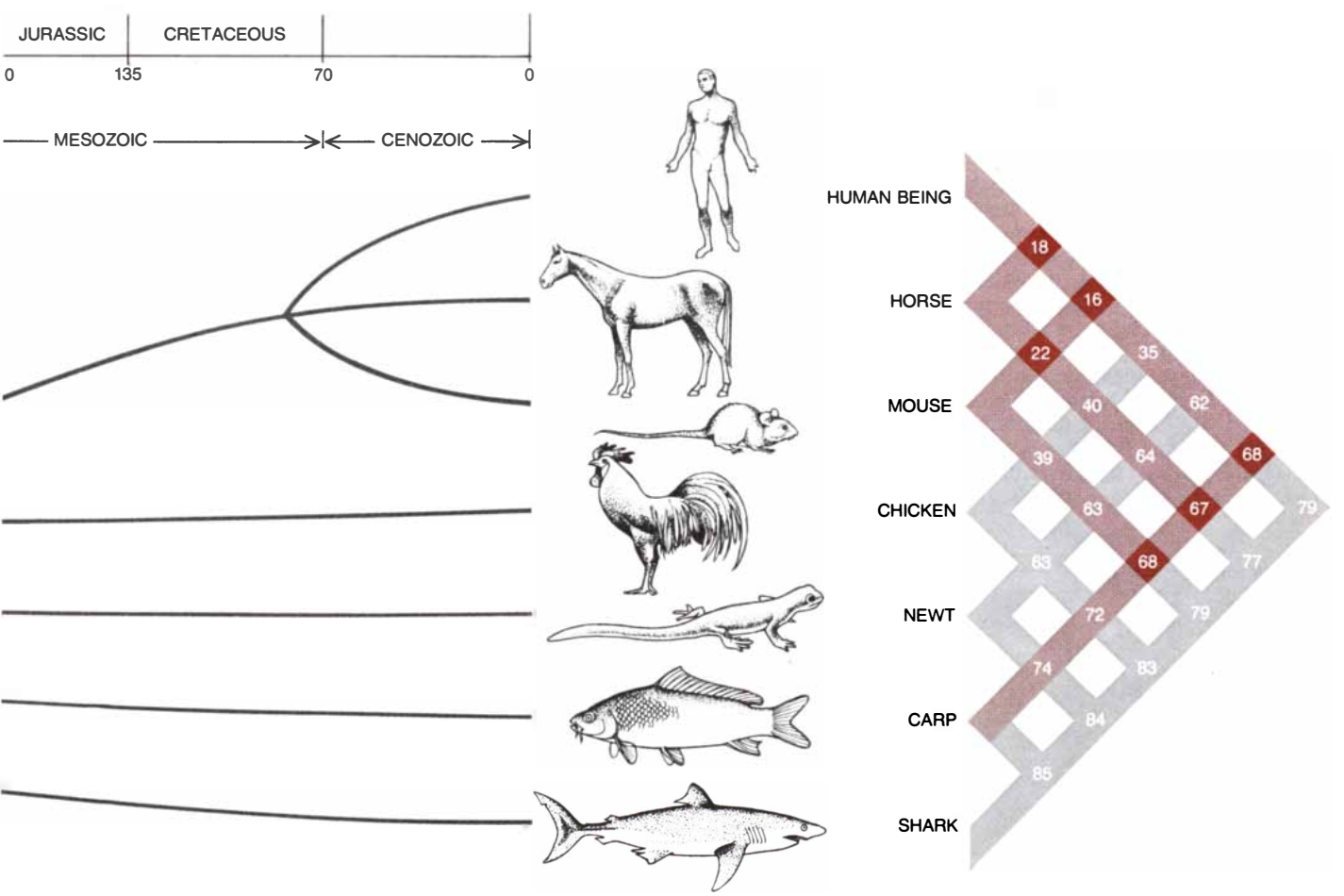
adaptive evolution through natural selection and that almost no mutant genes were selectively neutral. As Ernst Mayr stated the case in 1963, "I consider it ... exceedingly unlikely that any gene will remain selectively neutral for any length of time." A great deal was said by many workers about how genes interact, how gene pools of species are organized and how gene frequencies in populations change in the course of evolution. These conclusions, however, were necessarily inferences based on observations at the phenotypic level: the level of the form and function arising from the operation of genes. There was no way of knowing what actually goes on in evolution at the level of the internal structure of the gene.

Meanwhile the mathematical theory of population genetics was becoming quite sophisticated (which is rather unusual in biology). Particularly noteworthy was the theoretical framework provided by the manipulation of partial differential equations called diffusion equations. Diffusion models enable one to describe the behavior of mutant alleles by considering the random changes resulting from random sampling of gametes (germ cells) in reproduction as well as the deterministic changes caused by mutation and selection. Although the diffusion-equation method involves approximation, it yields answers to important but difficult questions that are inaccessible by other methods, such as: What is the probability of fixation for a single mutant appearing in a finite population and having a certain selective advantage, that is, what is the probability that it will eventually spread through the entire population?

The applicability of this method to gene changes in evolution, however, remained rather limited for some time. The reason is that population genetics deals with the concept of gene frequencies (the relative prevalence of various alleles within a population), whereas conventional studies of evolution were conducted at the phenotypic level, and there was no direct way of connecting the two sets of data unambiguously. That obstacle was removed with the advent of molecular genetics. It became possible to compare, in related organisms, individual RNA molecules (the direct products of genes) and proteins (the ultimate products) and so to estimate the rate at which allelic genes are substituted in evolution. It also became possible to study the variability of genes within a species. At last the time was at hand for applying the mathematical theory of population genetics to find out how genes evolve. One might have expected that the principle of Darwinian selection would prove to prevail at that fundamental level. Indeed, many evolutionary biologists found what they expected to find, and they have tended to extend panselectionism to the molecular level.

The Neutral Theory

The picture of evolutionary change that actually emerged from molecular studies seemed to me, however, to be quite incompatible with the expectations of neo-Darwinism. One of my salient findings with regard to evolution was that in a given protein the rate at which amino acids (the subunits of proteins) are substituted for one another is about the same in many diverse line-



of amino acids that constitutes the chain. The hemoglobin molecule has two alpha chains and two beta chains, which originated through the duplication of a single gene some 450 million years ago. The table reflects the approximate uniformity (predicted by the neutral theory) of the rate of evolution of a given protein in very different organisms. The number of amino acid differences is roughly 20 when any of the three mammals are compared with one another, and it is approximately 70 when the carp is compared with any of the three mammals.

TYPE OF CHANGE	HUMAN ALPHA V. HUMAN BETA	CARP ALPHA V. HUMAN BETA
NO CHANGE	62	61
ONE-NUCLEOTIDE	55	49
TWO-NUCLEOTIDE	21	29
GAP	9	10
TOTAL	147	149

NUMBER OF DIFFERENCES between the amino acid sequences of the alpha chain and the beta chain of human hemoglobin is compared with the number of differences between the sequences of the alpha chain in the carp and the human beta chain. The column at the left categorizes the amino acid sites according to whether there is no change, a change due to a minimum of at least one nucleotide substitution or at least two substitutions in the genetic code at each site, or a "gap": an addition or a deletion of an amino acid. The numbers are similar whether one compares the chains in the same species or in the two species, suggesting that alpha chains in two lineages have accumulated mutations at about same rate for 400 million years.

ages. Another finding was that the substitutions seem to be random rather than having a pattern. A third finding was that the overall rate of change at the level of DNA, the actual genetic material, is very high, amounting to the substitution of at least one nucleotide base (DNA subunit) per genome (total genetic complement) every two years in a mammalian lineage. As for the extent of variability within a species, electrophoretic methods for detecting small differences among proteins suddenly disclosed a wealth of genetic variability; the proteins produced by a large fraction of the genes in diverse organisms were found to be polymorphic, that is, they were present in the species in variant forms. In many cases the protein polymorphisms had no visible phenotypic effects and no obvious correlation with environmental conditions.

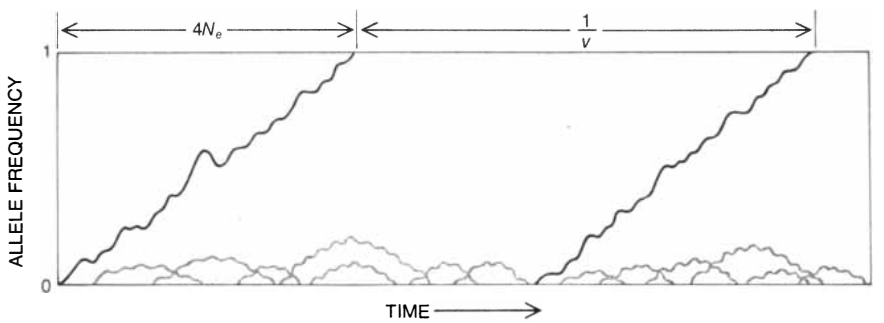
In 1967, as I considered these puzzling observations, I decided they suggested two things. One was that a majority of the nucleotide substitutions in the course of evolution must be the result of the random fixation of neutral or nearly neutral mutants rather than the result of positive Darwinian selection. The other was that many protein polymorphisms must be selectively neutral or nearly so and must be maintained in a population by the balance between mutational input and random extinction. I presented these thoughts at a meeting of the Genetics Club in Fukuoka in November, 1967, and in a short paper in *Nature* the following February. In 1969 strong support came from a paper in *Science* by Jack Lester King, now of the University of California at Santa Barbara, and Thomas H. Jukes of the University of California at Berkeley. They had arrived at the same ideas on molecular evolution (although not on protein polymorphisms) independently, and they presented cogent supporting data from molecular biology.

The papers suggesting a neutral theory were severely criticized by evolutionists who believed the new molecular data could be understood in the light of orthodox neo-Darwinian principles, and the neutralist-selectionist controversy continues today. The essential difference between the two schools of thought can be appreciated by comparing their differing explanations of the evolutionary process by which mutant genes come to be substituted in a species. Every substitution involves a sequence of events in which a rare mutant allele appears in a population and eventually spreads through the population to reach fixation, or a frequency of 100 percent. Selectionists maintain that for a mutant allele to spread through a species it must have some selective advantage (although they admit that an allele that is itself neutral may occasionally be carried along by "hitchhiking" on a gene that is selected for and with which it is closely linked, and may thus reach a high frequency).

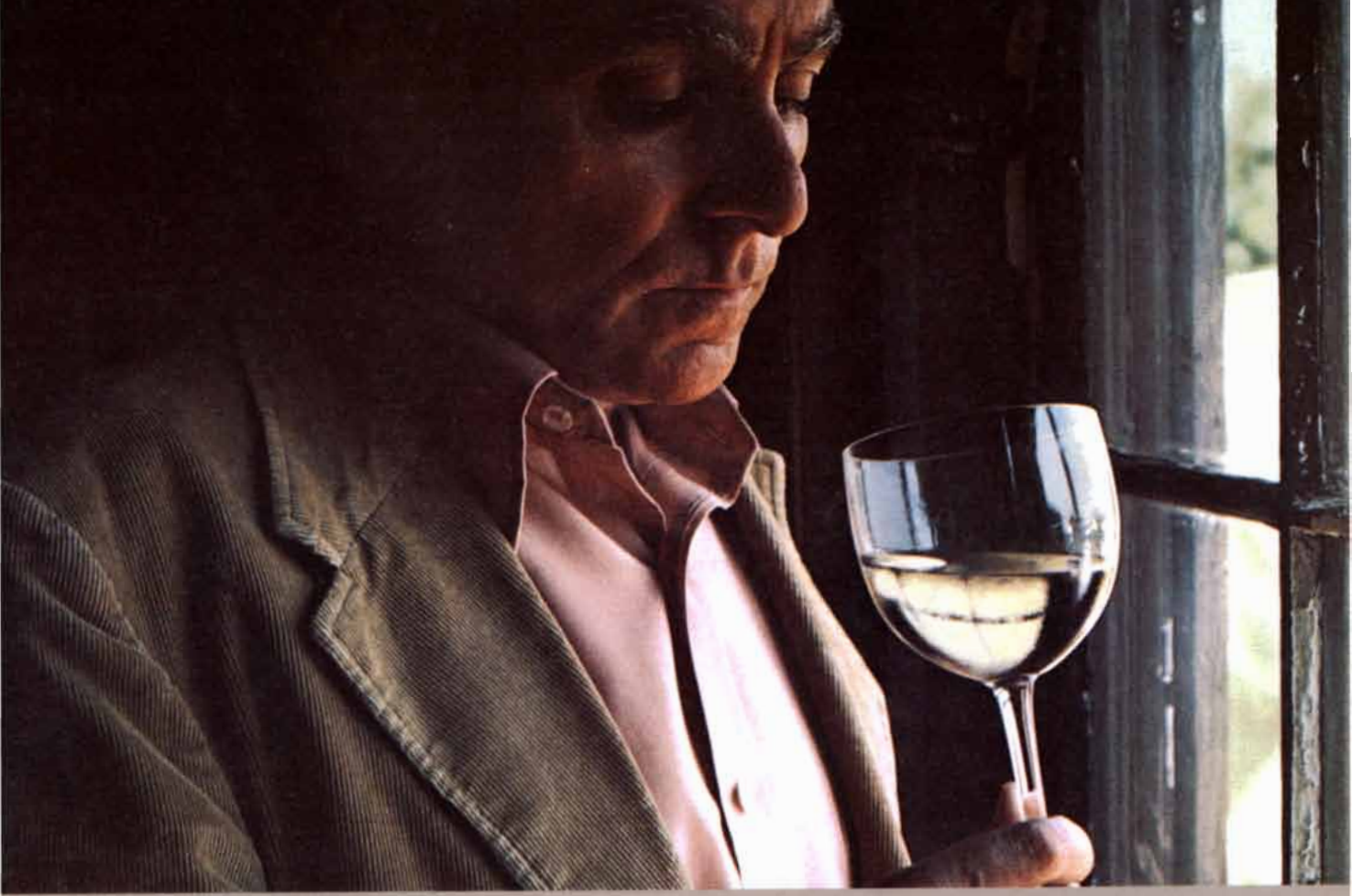
Neutralists, on the other hand, contend that some mutants can spread through a population on their own without having any selective advantage. If a mutant is selectively equivalent to preexisting alleles, its fate is left to chance. Its frequency fluctuates, increasing or decreasing fortuitously over time, because only a relatively small number of gametes are "sampled," out of the vast number of male and female gametes produced in each generation, and are therefore represented in individuals of the next generation [*see illustration on page 102*].

In the course of this random drift the overwhelming majority of mutant alleles are lost by chance, but a remaining minority of them eventually become fixed in the population. If neutral mutations are common at the molecular level and if the random drift is continuous over a long time (say millions of generations), the genetic composition of the population will change significantly. For any neutral mutant that appears in a population the probability of eventual fixation is equal to its initial frequency. The average length of time until fixation (excluding alleles that are lost) is four times the "effective" population size, or $4N_e$. (The effective size of a population is approximately equal to the number of breeding individuals in one generation, and it is usually much smaller than the total number of individuals in the species.)

The neutral theory, I should make clear, does not assume that neutral genes are functionless but only that various alleles may be equally effective in promoting the survival and reproduction of the individual. If a mutant allele encodes variant amino acids in a protein, the modified protein need function only about as well as the original form; it need not be precisely equivalent. In higher organisms particularly, homeostasis counteracts external environmental changes just as it does internal physiological changes; fluctuations in the environment do not necessarily imply



MUTANT ALLELES (variant genes) arise in a population at random. Their frequency fluctuates; in time most of them disappear (*gray lines*), but some of them spread through the population to fixation: a frequency of unity, or 100 percent (*black lines*). Population-genetic studies show that for a neutral allele that is destined for fixation the average number of generations until fixation is four times the effective population size, or $4N_e$. The average number of generations between consecutive fixations is equal to the reciprocal of the mutation rate v.



BECAUSE the delicate bouquet of this young wine foretells its noble future, our winemaker can patiently and carefully guide the promising youth of our Johannisberg Riesling.
Every step we take, we take with care because THE WINE REMEMBERS

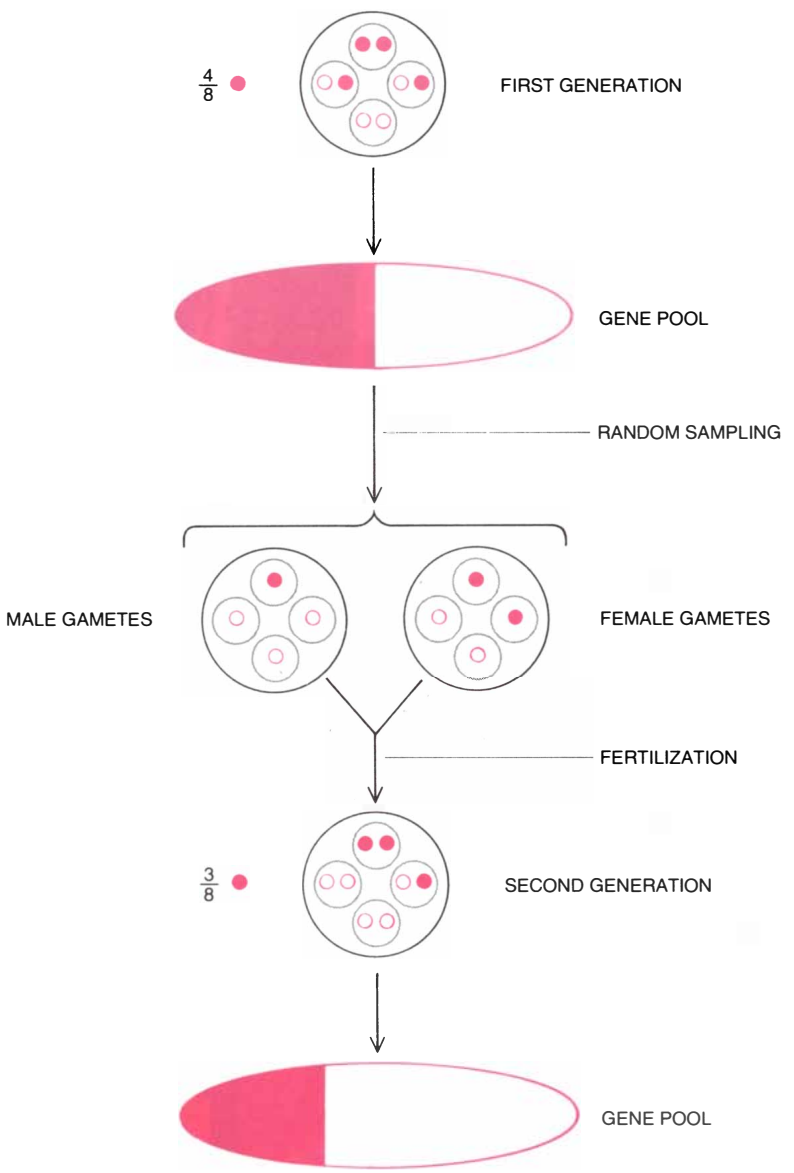
THE WINERY OF
ERNEST & JULIO
GALLO

Ernest & Julio Gallo, Modesto, CA

comparable fluctuations in the Darwinian fitness of mutant genes.

Some criticisms of the neutral theory arise from an incorrect definition of "natural selection." The phrase should be applied strictly in the Darwinian sense: natural selection acts through—and must be assessed by—the differential survival and reproduction of the individual. The mere existence of detectable functional differences between two molecular forms is not evidence for the operation of natural selection, which can be assessed only through investigation of survival rates and fecundity. Moreover, a clear distinction should be made between positive (Darwinian) selection and negative selection. The latter, which Muller showed is the commoner form, eliminates deleterious mutants; it has little to do with the gene substitutions of evolution. A finding of negative selection does not contradict the neutral theory. Finally, the distinction between gene mutation in the individual and gene substitution in the population should be kept in mind; only the latter is directly related to molecular evolution. For advantageous mutants the rate of substitution is greatly influenced by population size and degree of selective advantage (as I shall show below) as well as by the mutation rate.

Two major findings with regard to molecular evolution demonstrate particularly clearly that its patterns are quite different from those of phenotypic evolution and that the laws governing the two forms of evolution are different. One is the finding, alluded to earlier, that for each protein the rate of evolution in terms of amino acid substitutions per year is approximately constant and about the same in various lineages. The other is that molecules or parts of a molecule subjected to a relatively small degree of functional constraint evolve at a higher rate (in terms of mutant substitutions) than those subjected to stronger constraints do.



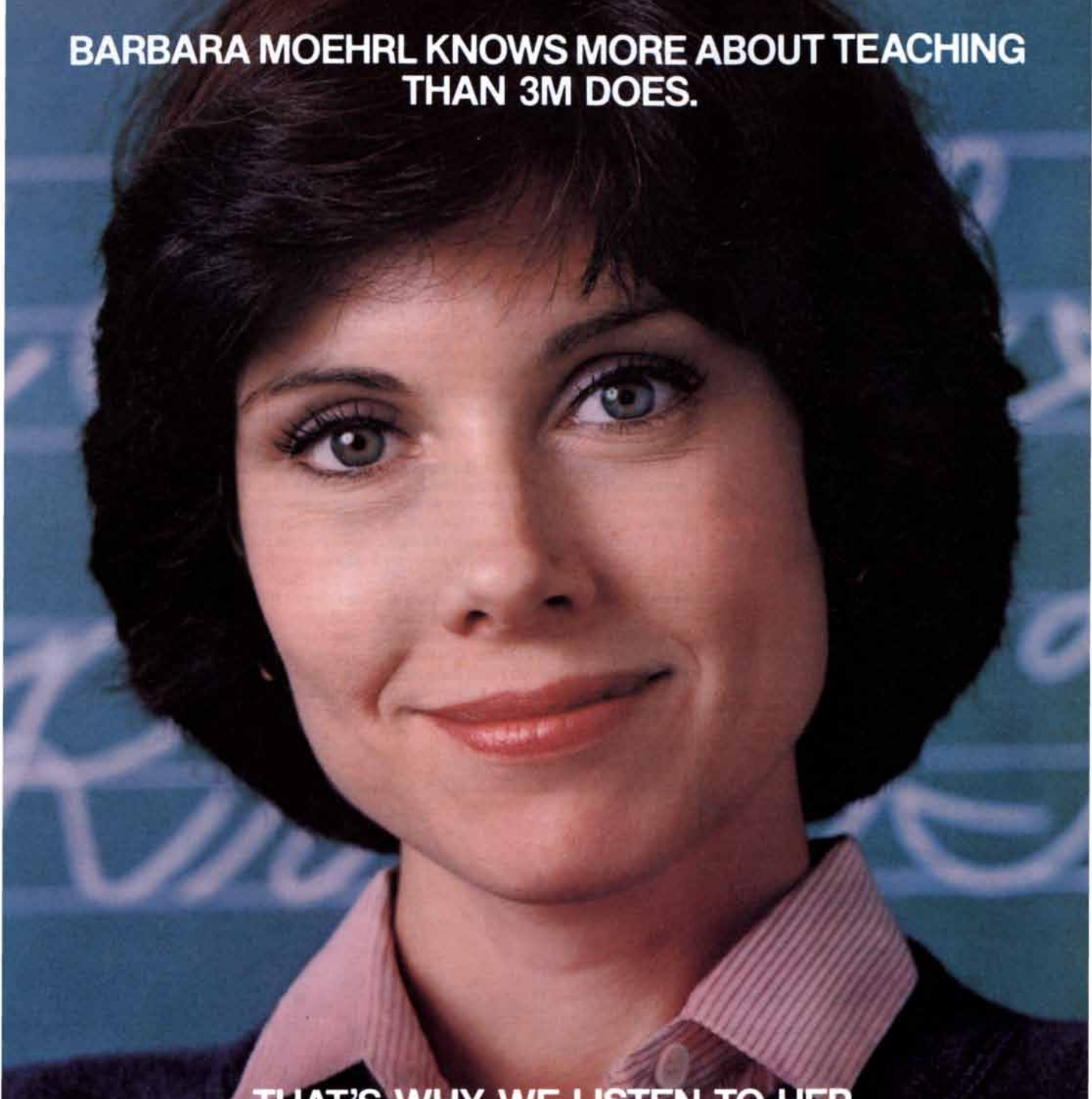
RANDOM CHANGES in gene frequency arise from random sampling of gametes (germ cells) in reproduction, as is shown here for a hypothetical population of four individuals (*gray circles*), each of which has two homologous genes (*solid color and open color*) inherited from the male and female parents. In the first generation the frequency of the "solid" allele is 4/8, and so the gene pool is 50 percent solid. Of the many gametes produced by a generation only a few are sampled, at random, in reproduction; here only one solid allele happens to be present in the four first-generation male gametes that engage in reproduction, so that the frequency of the solid allele changes to 3/8 in the second generation of individuals and hence in their gene pool.

Molecular Evolution

The constancy of the evolutionary rate is apparent in the molecule of hemoglobin, which in bony fishes and higher vertebrates is a tetramer (a molecule with four large subunits) consisting of two identical alpha chains and two identical beta chains. In mammals amino acids are substituted in the alpha chain, which has 141 amino acids, at the rate of roughly one substitution in seven million years. This corresponds to about one substitution in a billion years (or 10^{-9} substitution per year) per amino acid site. The rate does not appear to depend on such factors as generation time, living conditions and population size. The approximate constancy of the rate is evident when the number of amino acid differences between the alpha chains of various vertebrates is charted together with the phylogenetic tree showing the relations among the vertebrates and the times when they diverged from one another in evolution [*see illustration on pages 98 and 99*].

The alpha and beta chains have essentially the same structure, are about the same length and show roughly the same rate of evolutionary amino acid substitution. They arose through gene duplication some 450 million years ago and became differentiated as they accumulated mutations independently. If one compares the divergence between the alpha and the beta chain of man with the divergence between the alpha chain of the carp and the beta chain of man, it is evident that in both cases the alpha and beta chains differ from each other to roughly the same extent. Because the alpha chain of man and that of the carp differ from each other in about half of their amino acid sites this suggests that the alpha chains in two distinct lineages, one leading to the carp and the other to man, have accumulated mutations independently and at practically the same rate over a span of about 400 million years. Moreover, the rate of amino acid substitution observed in these comparisons is very similar to the rates observed in comparisons of the alpha chains in various mammals.

My assertion of constancy of the evo-



BARBARA MOEHRL KNOWS MORE ABOUT TEACHING THAN 3M DOES.

THAT'S WHY WE LISTEN TO HER.

When Barbara Moehrl talks about teaching, it's in our interest at 3M to listen. Because Barbara's commitment to classroom innovation and contemporary teaching methods dates from the start of her career in Marshalltown, Iowa, in 1968.

Today, media-conscious teachers like Barbara reach far and wide for communications techniques. Like 3M Overhead Projectors for visual impact. And Wollensak Cassette Recorders for excitement in sound. And when new aids are required, Barbara Moehrl will be among the first to know.

By listening to her, 3M can create more ingenious products for practical use. And that's our job. At 3M, listening is more than just good philosophy. It's vital to our future.

3M HEARS YOU

3M

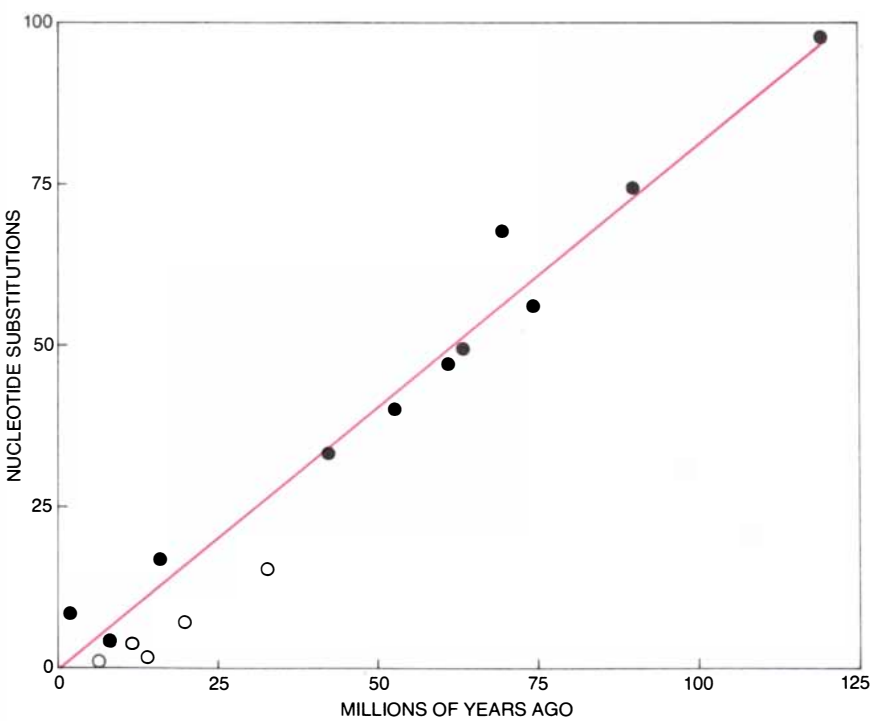
lutionary rate at the molecular level has been criticized by, among others, Richard C. Lewontin of Harvard University, who called the asserted constancy "simply a confusion between an average and a constant" and "nothing but the law of large numbers." The remarks reveal a misunderstanding of the nature of molecular evolution. One is attempting here to compare intrinsic rates of evolution in different lineages. The death rates characteristic of man and of an insect do not become equal by merely being averaged over a long period of time or over a large number of individuals; there is no reason to expect two averages to converge on each other unless the intrinsic factors shaping them are the same. My point is that intrinsic evolutionary rates are essentially determined by the structure and function of molecules and not by environmental conditions.

Evolutionary rates are, to be sure, not precisely constant in the sense that a rate of radioactive decay is constant. My colleague Tomoko Ohta and I showed in 1971 that the variance (the squared standard deviation) of the evolutionary rate observed for hemoglobins and for the protein cytochrome *c* in different mammalian lines is about 1.5 to 2.5 times larger than the variance to be expected if it were due only to chance. Charles H. Langley and Walter M. Fitch did a more elaborate analysis at the University of Wisconsin School of Medicine, combining data for the alpha and beta hemoglobin chains, cytochrome *c* and fibrinopeptide *A;* they found a variation in rates of mutant substitution about 2.5 times larger than the expected variance due to chance fluctuations, and they took this as evidence against the neutral theory. Yet they also showed that when the estimated number of substitutions between diverging branches of a phylogenetic tree is plotted against the corresponding time of divergence, the points fall on a straight line, which suggests the substantial uniformity of the evolutionary rates. It seems to me to be wrong to emphasize local fluctuations as evidence against the neutral theory while neglecting to inquire into why the rate remains essentially constant.

Rates of Evolution

Turning now to the quantitative relations that determine rates of evolution, consider first the nucleotides constituting a genome: a single (haploid) set of chromosomes. For a human being the number of nucleotides is very large, on the order of 3.5 billion. Because the mutation rate per nucleotide site is low (perhaps 10^{-8} per generation, or one mutation per 100 million generations) one can assume that whenever a mutant appears it is at a new site. This assumption is called the "infinite site" model in population genetics.

Let v represent the mutation rate per



NUCLEOTIDE SUBSTITUTIONS estimated from the total number of amino acid differences observed in seven proteins in 16 pairs of mammals are plotted against the time since members of each pair diverged. Except for lineages involving primates (*open circles*) the points fall close to a straight line, again suggesting approximate uniformity of a protein's molecular evolutionary rate. Data are from Walter M. Fitch of University of Wisconsin School of Medicine.

There is just one hatchback in all the world  with front-wheel  drive, transverse-mounted engine, rack and pinion steering,  low drag front disc  brakes,  over 40 cu. ft. of load space, and  Pontiac's kind of flair. 

1980 PONTIAC PHOENIX

24	37	336	518
EPA EST. MPG	HWY. EST. MPG	MILES EST. RANGE	MILES HWY. EST. RANGE

Remember: Compare the circled *estimated MPG* with those of other cars. Your mileage and range will vary by speed, distance and weather. Actual highway mileage and range will probably be less than the highway estimates. Mileage and range lower in California.

MORE PONTIAC EXCITEMENT TO THE GALLON

Range estimates were obtained by multiplying EPA estimates by the car's 14 gallon fuel tank capacity. Phoenix is equipped with GM-built engines produced by various divisions. See your dealer for details.

GM

MARK OF EXCELLENCE

The brain behind...

Look at that face. She's beautiful. And Polaroid's SX-70 Sonar Land camera caught her beauty with technology unmatched by any camera in the world.

Sonar automatic focusing. With it, you get perfect focus at the touch of a button. In milliseconds the brain reacts. Sound waves travel to your subject and echo back. The distance is measured, and the lens rotates automatically to perfect focus.

The versatility of the camera allows you to shoot at will. Take a long shot. Come in for a close-up. Capture an expression before it becomes self-conscious. Take automatic exposures up to 14 seconds. Or shoot as often as every 1.5 seconds. Sonar readjusts the focus every time. Instantly and automatically.

The beauty.

But then, virtually everything about the SX-70 Sonar camera is automatic.

It deals automatically with shutter speeds.

It automatically adjusts f-stops.

It enables you to shoot from as close as 10.4 inches to infinity, never requiring a change of lens.

It is the world's smartest camera.

For the beauty, look to the photographs. The close-up of the child is a portrait of subtle shadings and nuances of color.

Look at the richness and depth of color in the kimono. These are the photographs the SX-70 Sonar can deliver. Instantly.

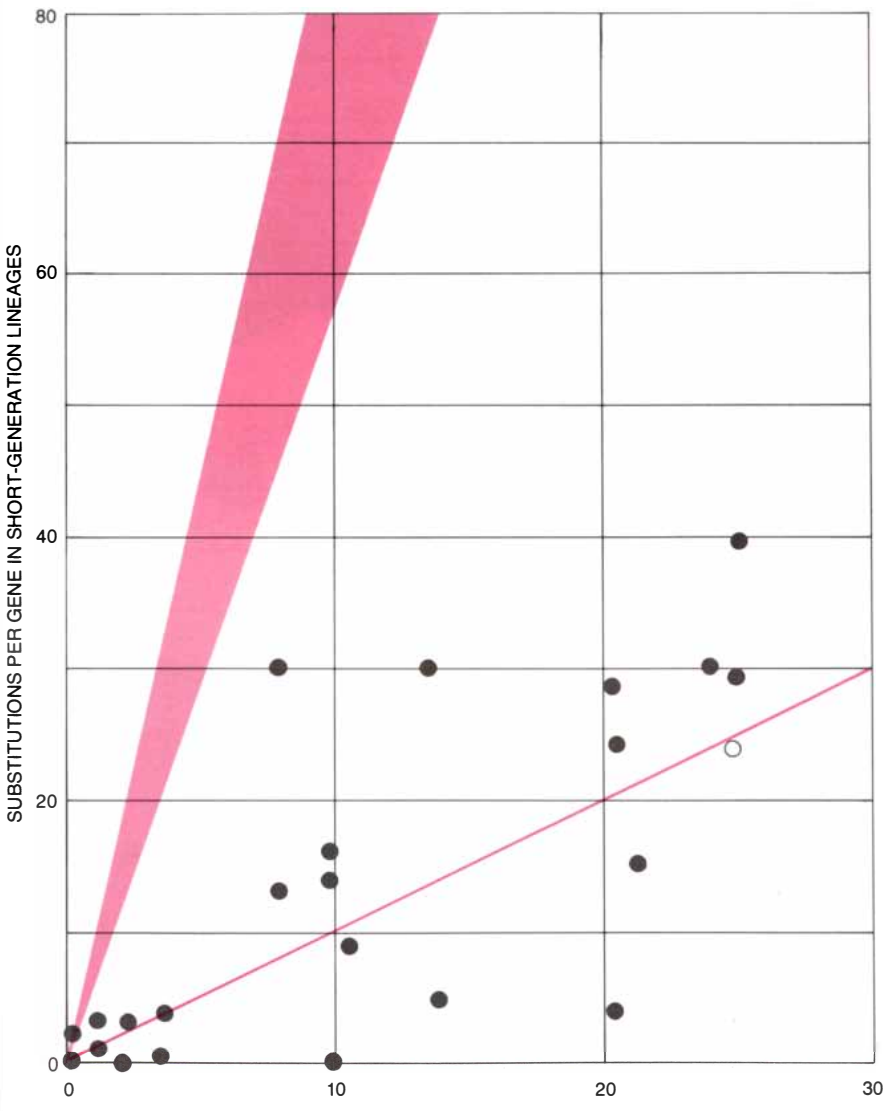
Because behind every beauty the camera hands you, is the remarkable SX-70 Sonar brain.

POLAROID'S SX-70 SONAR

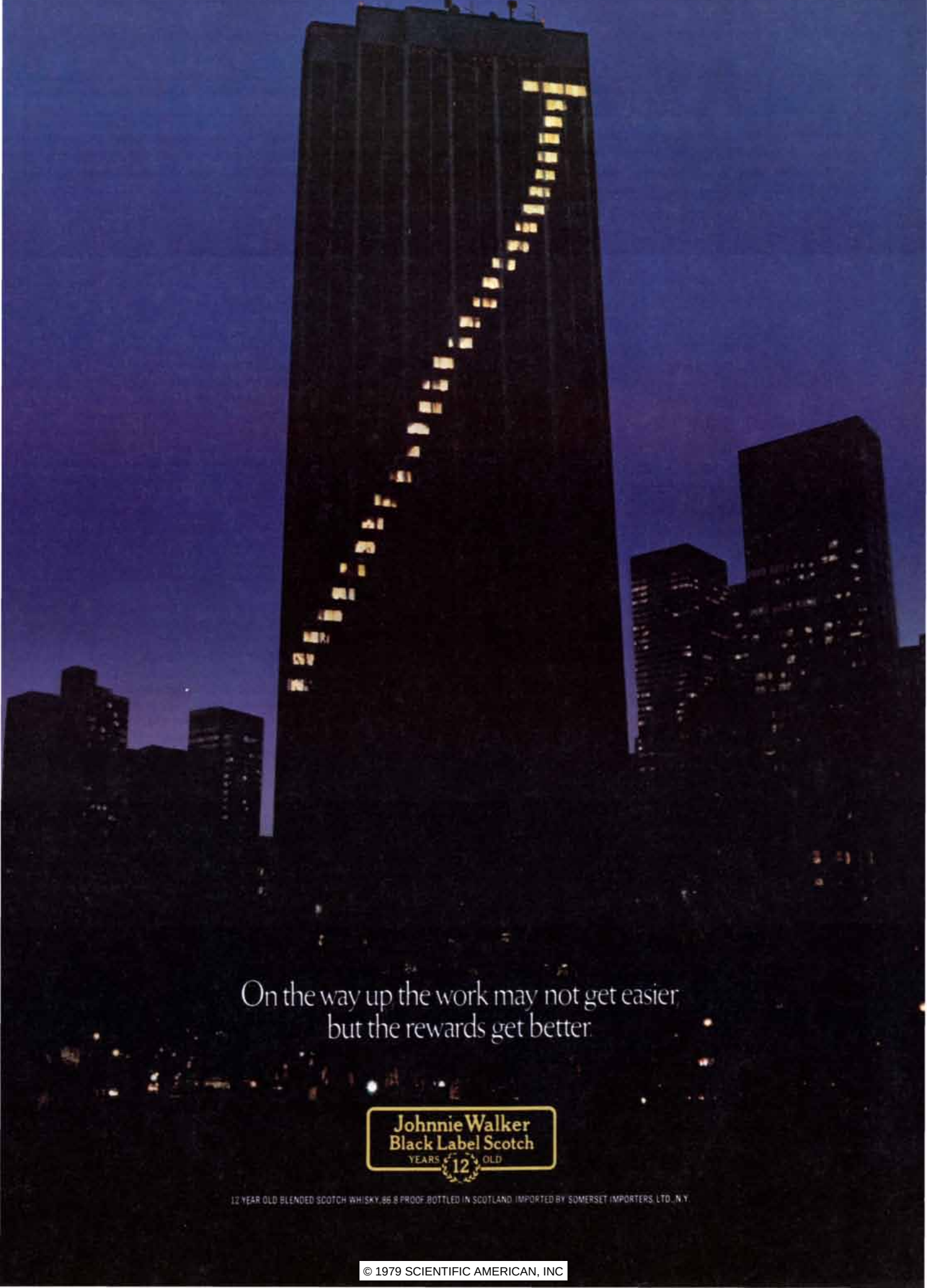
The world's smartest camera.

gamete per unit time (generation). Since each individual has two sets of chromosomes, the total number of new mutants introduced into a population of N individuals in each generation is $2Nv$. Now let u be the probability that a single mutant will ultimately reach fixation. Then, in a steady state in which the process of substitution goes on for a very long time, the rate k of mutant substitution per unit time is given by the equation $k = 2Nvu$. That is, $2Nv$ new mutants appear in each generation, of which the fraction u eventually reach fixation, and k represents the rate of evolution in terms of mutant substitutions. The equation can be applied not only to the genome as a whole but also, with good approximation, to a single gene consisting of several hundred nucleotides or to the protein encoded by a gene.

The probability u of ultimate fixation is a well-known quantity in population genetics. If the mutant is selectively neutral, u equals $1/(2N)$. The reason is that any one of the $2N$ genes in the population is as likely as any other to be fixed, and so the probability that the new mutant will be the lucky gene is $1/(2N)$. (This assumes that the process is being viewed over a very long period of time, since the average time for a neutral gene to sweep through the population is $4N_e$.) Substituting $1/(2N)$ for u in the equation for the rate of evolution ($k = 2Nvu$), one gets $k = v$. That is, the rate of evolution in terms of mutant substitutions in the population is simply



MOLECULAR EVOLUTIONARY RATE in mammals that have a short generation time was compared with the rate in mammals having a long generation time by Allan C. Wilson and his colleagues at the University of California at Berkeley. Each point represents the ratio of the implied nucleotide substitutions in the two animals of a pair since the two diverged from a common ancestor; the open circle, for example, is for the beta chain of hemoglobin in the elephant (*abscissa*) and in the mouse (*ordinate*). If the rate of change per year were identical in both animals of a pair, the points would fall on a line (*solid color*). Actually the points are close to that absolute-time line and far from sector predicted for generation-time effect (*colored area*). Apparently molecular evolutionary rate is roughly constant per year, not per generation.



equal to the rate of mutation per gamete, independent of what the population size may be.

This remarkable relation applies only for neutral alleles. If the mutant has a small selective advantage s, then u equals $2s$ with good approximation, and the equation for the rate of evolution becomes $k = 4Nsv$. That is, the rate of evolution for selectively advantageous genes depends on the size of the population, the selective advantage and the rate at which mutants having a given selective advantage arise in each generation. One should expect, in this case, that the rate of evolution would depend strongly on the environment, being high for a species offered new ecological opportunities but low for those kept in a stable environment. It is highly unlikely, I think, that the product Nsv should be the same in diverse vertebrate lineages, in some of which phenotypic evolution has been very rapid (as in the line leading to man) and in others of which phenotypic evolution has long since practically ceased (as in the line leading to the carp). And yet the observed rates of molecular evolution show remarkable constancy. It seems to me that this constancy is much more compatible with the expectation of the neutral theory, that is, with the equation $k = v$, than it is with the selectionist relation $k = 4Nsv$.

Even more striking than the constancy of the rate of evolution is the second major feature of molecular evolution: the weaker the functional constraint on a molecule or a part of a molecule, the higher the evolutionary rate of mutant substitutions. There are regions of DNA between genes, for example, and in the case of higher organisms even within genes, that do not participate in protein formation and must therefore be much less subject to natural selection; some recent research has indicated that nucleotide substitutions are particularly prevalent in such regions of DNA.

Functional Constraint

This relation between a relative lack of selective constraint and a relatively high rate of molecular evolution has been well established for certain proteins. Among the proteins so far investigated the highest evolutionary rate has been found in fibrinopeptides, which appear to have little function, if any, after they become separated from fibrinogen to yield fibrin, a protein that plays a role in blood clotting. The same effect is observed in the case of the C chain of the proinsulin molecule, a precursor of insulin. The C chain, which is cleaved from the precursor to form active insulin, evolves at a rate several times higher than that of the active molecule. The effect of constraint on the evolutionary rate has also been noted for different parts of the hemoglobin molecule. The precise structure of the surface of the

HOW THE NEW EATON PROFITS FROM CHANGE

We're easing the squeeze in the Suez Canal with air traffic control technology.

Eaton electronic traffic management systems. They help direct traffic into busy airports. Soon one will keep traffic moving in the Suez Canal.

That means 22,000 vessels a year – including 30% of the world's tanker fleet – will get the information they need to steer clear every crowded mile between Suez and Port Said.

We also profit from change in hydraulics for mining...

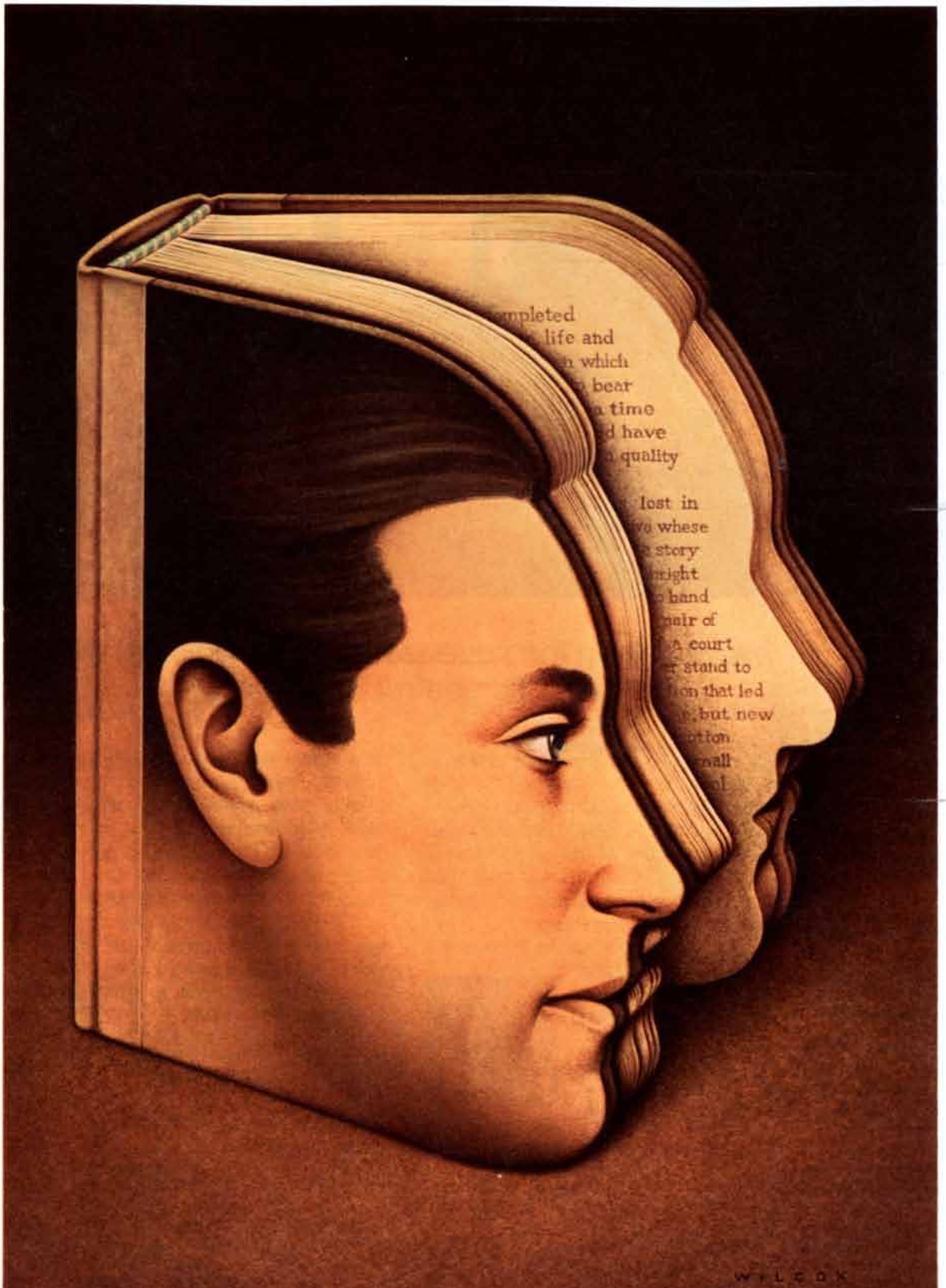
Designed and built by Eaton's AIL Division, this Vessel Traffic Management System combines ultra-high resolution radar with sophisticated communications, a precise ship-locating system, and a computerized operations center. The result: a new level of efficiency and safety in busy ports from Alaska to San Francisco, from New York to the Mideast.

clutches and adjustable speed drives...

and sophisticated hoses for industry and farm.

Technological leadership like this has helped Eaton grow to an annual rate of $3 billion in sales. For the complete story, write: Eaton Corporation, 100 Erieview Plaza, Cleveland, Ohio 44114.

EATON



Painting by David Wilcox

The future is coming. And with it will come great benefits for mankind. And a whole new set of problems. Because we are a forest products company, and plant seeds that take up to 50 years to become mature trees, Champion International has to think a lot about the future. We'd like to share some of the things we've learned with you—to help you make intelligent choices in the years to come. Here is something you might want to think about.

In the future, electronic information gathering and analysis techniques could be used to find out more about you, and what you think and feel, than you may want others to know.

How will we prevent our heads from turning into open books?

Are your thoughts and feelings *your* property? Or do they belong to any organization that has the skill, technology and *intent* to collect whatever personal data they want about you?

This is a major question concerning many of the world's leading lawmakers today. And with advancing technology, it will become even more of an issue in the future.

Futurist David Goodman believes that ours could become "the most snooped-on society in history" unless we take steps to prevent it.

The issue of personal privacy is a very complex one.

First of all, there is no commonly accepted definition of privacy.

Is privacy "the right to be let alone"? Or is it "the right of the individual to determine when, how, and to what extent there should be a disclosure of information about himself"?

To dramatize how difficult it is to find an appropriate definition, after two years of study and debate a government commission called The Privacy Protection Study Commission finally concluded that *privacy* was not even a good label for many of the issues they were considering, and settled, for lack of a better term, on "fair informational practices."

Many legislators understand that while personal privacy must be meaningfully protected, restrictions on the gathering and analysis of personal data could seriously hamper our society's ability to make important decisions.

As our society becomes more complex, we need to rely on the accurate large-scale gathering and analysis of information to help us with rational decision-making about important public issues like allocation of resources, housing and food supply and health care needs.

Accurate information gathering would also be invaluable in such critical personal decisions as choice of careers, habitats—even mates.

Obviously, information will be one of our most important products in the future. Already, one job in two in the U.S. is related to the handling of information.

The Privacy Protection Study Commission has recommended that Congress establish an independent Federal Privacy Board within the government to protect personal privacy in The Age of Information, with its omnidirectional microphones, its lasers, its incredible data banks, etc.

Obviously, there are legitimate needs on both sides of the privacy issue. What we now do about the use of personal information can have a profound effect on our future. It's something to think about.

But if you'd like to do more than just think about it, write to:

Champion International Corporation, Box 10126, 1 Landmark Square, Stamford, Connecticut 06921.

Champion, a forest products company with its roots planted firmly in the future.

We are in the forest products business. We plant trees and harvest trees. We produce wood and paper. And we make things out of wood and paper.

Because we make our living from the forest, our success depends, in one way or another, on the future. These are our operating divisions that are planning for the future:

CHAMPION TIMBERLANDS
CHAMPION BUILDING PRODUCTS
CHAMPION PAPERS
CHAMPION PACKAGING (HOERNER WALDORF)

Champion

Champion International Corporation

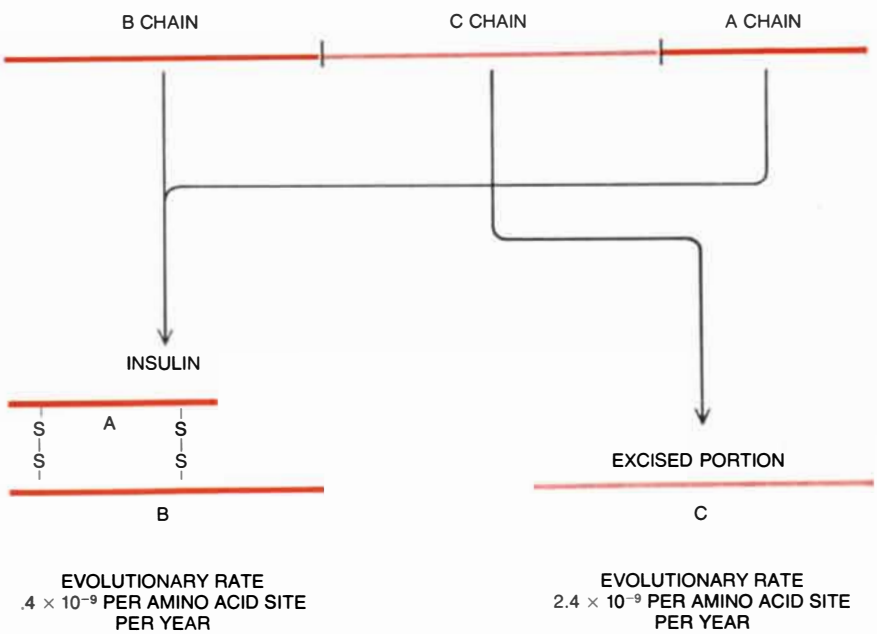
Planting seeds for the future

PROTEIN	EVOLUTIONARY RATE
FIBRINOPEPTIDES	9.0
PANCREATIC RIBONUCLEASE	3.3
HEMOGLOBIN CHAINS	1.4
MYOGLOBIN	1.3
ANIMAL LYSOZYME	1.0
INSULIN	.4
CYTOCHROME *c*	.3
HISTONE IV	.006

PROTEINS VARY WIDELY in their rate of evolution. Here the rate is given for several proteins in terms of the number of amino acid substitutions per amino acid site per billion years. The rate is particularly high for fibrinopeptides, which appear to have little function after they are cleaved from a precursor molecule to yield the active blood-clotting protein fibrin. Proteins such as fibrinopeptides are subject to less functional constraint, and evolve faster, than proteins whose precise shape is significant and hence subjects them to stronger constraint.

molecule is presumably less significant than the structure of the pockets, in the interior of the molecule, that hold the iron-containing heme groups. Ohta and I have estimated that the regions of both the alpha and the beta chain that are at the surface of the protein evolve about 10 times faster than the regions forming the heme pocket.

The genetic code is based on groups of three nucleotides, with each triplet "codon" in a strand of RNA specifying a particular amino acid of the protein chain encoded by the RNA. For example, the codon *GUU* (the letters stand for particular nucleotide bases) specifies the amino acid valine. So does *GUC*, however; the genetic code is "degenerate," with most amino acids being designated by two or more synonyms, which typically differ only in the third position of the triplet. As a result a large fraction (perhaps 70 percent) of all random nucleotide substitutions at the third position are synonymous changes and do not lead to amino acid replacements. There is growing evidence that evolutionary nucleotide substitution goes on at a particularly high rate at the third position. Michael Grunstein of the University of California at Los Angeles and his colleagues compared the RNA sequences encoding the protein histone IV in two species of sea urchin. They found that although the protein has maintained a practically unchanged amino acid se-



C CHAIN OF PROINSULIN (*light color*) is cleaved from the precursor molecule and discarded, and the A and B chains (*dark color*) are joined by disulfide bridges (*S–S*) to form the active insulin molecule. In keeping with the C chain's relative freedom from functional constraint, it has been found to evolve at a much higher rate than the two chains of the active molecule.

This is the computer...

TRS-80 Computer Consultant
Ralph Lipe of Fort Worth, Texas

that taught this boy...

to write this program! TRS-80™... from $499* (Level I, 4K)

Fourteen-year-old Ralph Lipe taught himself to program a TRS-80 Microcomputer System with Radio Shack's "plain language" instruction manual. Then he developed a program for completing the IRS 1040-A tax form that took top honors at his school and regional science fairs. *That was last year.*

This year Ralph is in high school and working part-time as a computer consultant. Everyone from grade schoolers to graduate students can use the world's best-selling microcomputer, the TRS-80. It doesn't do your thinking for you — it helps you learn to think. Think about it!

Meet TRS-80's Big Brother!

New TRS-80 Model II

Systems from $3450

We've added a bigger, more powerful "brother" to the TRS-80 family. It's TRS-80 Model II, a completely new microcomputer for business applications. Its capabilities begin where Model I approaches upper limits. Storage capacity from ½ megabyte to 2 megabytes! For earliest delivery, order now from any of our 7000 locations.

Radio Shack®

The biggest name in little computers™

A Division of Tandy Corporation • Over 7000 Locations in 40 Countries

***Retail prices may vary at individual stores and dealers.**

Apply for this Card

The more you travel and the more you entertain, the more you need the American Express® Card. Successful business trips and successful vacations are often due to what the Card can do for you.

The American Express Card can do far more than charge things. It almost goes without saying that the Card is welcome at better restaurants, hotels and shops of the world. Not to mention airlines and car rentals.

But the travel services and security that go along with it are what distinguish the Card. And there are more than 900 Travel Service Offices of American Express, its subsidiaries and Representatives. At home and abroad. To help you with replacing a lost Card, with emergency money or simply to help with a tangled itinerary. No other card has anything like them.

So if you want to carry the world's most respected travel and entertainment card, here's a perfect opportunity to apply.

If the application is missing, just phone 800-528-8000 and we'll send you another one.

AMERICAN EXPRESS

The American Express Card. Don't leave home without it.℠

quence for about a billion years, a number of synonymous nucleotide differences are found in the RNA sequences of the two species. On the basis of their data and paleontological evidence on the length of time since the species diverged, I have estimated that the rate of nucleotide substitution has been roughly 3.7×10^{-9} per year at the third position, a very high rate. What is remarkable is that there have been so many synonymous mutant substitutions in the histone-IV gene in spite of the very low rate of amino acid changes in the corresponding protein.

These observations can be explained simply and consistently by the neutral theory. Suppose a certain fraction f_0 of all molecular mutants are selectively neutral and the rest are definitely deleterious. Then the mutation rate v for neutral alleles is equal to the total mutation rate v_T multiplied by f_0, so that the overall rate k of mutant substitution becomes equal to $v_T f_0$. Now assume that the probability that a mutational change is neutral (not harmful) depends strongly on functional constraints. The weaker the constraint is, the larger will be the probability f_0 that a random change is neutral, with the result that the rate of evolution k increases. The maximum evolutionary rate is attained when f_0 equals 1, that is, when all the mutations are neutral. In my opinion the high evolutionary rates observed at the third position of the codon are rather near this limit.

The neutral theory, then, predicts that as functional constraint diminishes, the rate of evolution converges to the maximum value set by the total mutation rate. Confirmation of such a convergence, or plateauing, of molecular evolutionary rates by further studies would be strong evidence in support of the neutral theory. This interpretation of the molecular data will not make sense to selectionists. In their view molecules or parts of a molecule that are evolving rapidly in terms of mutant substitutions must have some important but as yet unknown function and must be undergoing rapid adaptive improvement by accumulating beneficial mutants. And they will see no reason to believe the upper limit of the evolutionary rate is related to the total mutation rate.

Polymorphism

Neutralists and selectionists also have diametrically opposed explanations for the mechanisms by which genetic variability is maintained within a species, particularly in the form of protein polymorphism: the coexistence in a species of two or more different forms of a protein. Neutralists maintain that polymorphisms are selectively neutral and are maintained in a population through mutational input and random extinction; in every generation a number of neutral mutants arise and in time either become fixed in the population or are lost, and in the process they contribute to genetic variability in the form of polymorphisms. In the neutral view polymorphism and molecular evolution are not two distinct phenomena; polymorphism is simply one phase of molecular evolution.

Selectionists maintain that polymorphisms are actively maintained by some form of "balancing selection," notable among which are heterotic selection, or "heterozygote advantage," and frequency-dependent selection. At one time the former was enthusiastically proposed as the main agent maintaining polymorphisms. There are instances in which individuals that are heterozygous for a particular gene (that carry a different allele of the gene on each of their two chromosomes) are fitter than individuals that are homozygous for either allele (that carry one or the other allele on both chromosomes). Selection will then tend to preserve both alleles in the population as a balanced polymorphism. In 1973, however, Roger D. Milkman of the University of Iowa found abundant polymorphisms in the bacterium *Escherichia coli,* which is a haploid organism: it has only one set of genes. Heterozygote advantage cannot explain such polymorphisms.

Nowadays many selectionists explain polymorphisms as being the result of frequency-dependent selection, in which the fitnesses of two alleles vary with their relative frequency. This was first proposed by the late Ken-Ichi Kojima of the University of Texas at Austin and his colleagues, who obtained results indicating marked frequency-dependent selection affecting the genes for the enzymes esterase-6 and alcohol dehydrogenase (ADH) in the fruit fly *Drosophila melanogaster;* Bryan Clarke of the University of Nottingham reported that he had confirmed Kojima's results in the case of ADH. On the other hand, experiments done by Tsuneyuki Yamazaki, now of Kyushu University, failed to show any such selection for esterase-5 alleles in *Drosophila pseudoobscura.* A group led by Terumi Mukai of Kyushu carried out extensive studies of selection for several enzymes in *D. melanogaster* and found no evidence for a difference in the fitness of variant forms of the enzymes. And recent careful, large-scale experiments by Mukai and Hiroshi Yoshimaru have failed to find any frequency-dependent selection for ADH in *D. melanogaster.*

If selection is not responsible for maintaining polymorphisms, what neutralist explanation is there for the fact that some proteins are more often polymorphic than others? Recently Richard K. Koehn of the State University of New York at Stony Brook and W. F. Eanes, now of Harvard, and also Masatoshi Nei's group at the University of Texas at Houston, have shown that in various *Drosophila* species there is a significant correlation between the genetic variability (or polymorphism) of proteins and the weight of their molecular subunits. This is easy to explain according to the neutral theory because the larger the size of a subunit is, the higher its mutation rate should be. Harry Harris of the University of Pennsylvania

	24	25	26	27	28	29	30	31	32	33	34
L. PICTUS MESSENGER RNA	GA **U**	AAC	AUC	CAA	GG **A**	AU **A**	AC **U**	AAA	CCG	GC **A**	AUC
S. PURPURATUS MESSENGER RNA	GA **C**	AAC	AUC	CAA	GG **U**	AU **C**	AC **G**	?	?	GC **U**	AUC
HISTONE IV AMINO ACID SEQUENCE IN BOTH SPECIES	Asp	Asn	Ile	Gln	Gly	Ile	Thr	Lys	Pro	Ala	Ile

NUCLEOTIDE SEQUENCES of the messenger RNA encoding the protein histone IV in two sea-urchin species, *Lytechinus pictus* and *Strongylocentrotus purpuratus,* were compared by Michael Grunstein of the University of California at Los Angeles. There are four RNA nucleotides (*A, G, U* and *C*); three-nucleotide codons specify the various amino acids that constitute a protein. Most amino acids are specified by two or more synonymous codons, which usually differ only at their third position. In this short stretch of RNA coding for amino acid sites 24 through 34 of histone IV there are five synonymous differences (*color*) in third-position nucleotides. That is, there has been a high rate of nucleotide substitution at the unconstrained third position, leaving the amino acid sequences unaffected.

If we didn't live in Maryland, this gang of mine would probably set a new record for hospital bills.

A Hospital Services Cost Review Commission keeps their rates down.

These days, having a healthy family is more of a blessing than you may realize. In recent years, the cost of health care has risen faster than the rate of inflation.

It's lucky for all of us that states like Maryland are blazing some cost-cutting trails. They've set up hospital cost review commissions that study hospital budgets a year in advance and recommend changes that hold down cost increases.

And look what Maryland accomplished. Since 1976, cost increases have been about 16% below the national average; a $189 million saving for all patients. And the results are as impressive in Connecticut, Massachusetts and other states.

Figures like that speak for themselves. With a little effort, and a lot of involvement, imagine what we could accomplish if every state had a commission.

We hope you're as impressed with this cost cutting as we are. As a major part of the nation's private health insurance system, which provides general health care coverage for 179,000,000 Americans and protection against major medical expenses for nearly 140,000,000, keeping costs controlled is a top priority with us. Because keeping costs down keeps premiums down, too. And that makes good health care affordable to more people.

That's why we're doing our part to help. For instance, many of our companies are cutting costs by paying for pre-hospitalization testing and second opinions for surgery.

If you'd like to know more about state hospital cost review commissions, we'll send you our free booklet, *Cutting Health Care Costs with Hospital Budget Review.* Write us at Health Insurance Institute, Department 17, 1850 K Street, N.W., Washington, D.C. 20006.

THE HEALTH INSURANCE COMPANIES IN AMERICA

Let's Keep Health Care Healthy.

and his colleagues could not find the same correlation when they investigated human polymorphisms, but they did find that single-subunit enzymes are more polymorphic than multiple-subunit ones, something that Eleftherios Zouros of Dalhousie University had earlier reported in *Drosophila*. One finding by Zouros and Harris fits the neutral theory particularly well: multiple-subunit enzymes that form hybrid molecules by combining with enzymes encoded by other genes have a clearly reduced level of polymorphism. The precise interaction among subunits required to form such enzymes would increase the degree of functional constraint and so reduce the possibility that a mutation will be harmless, or neutral.

Neutralists, in other words, consider molecular structure and function to be the major determinants of protein polymorphisms. Selectionists consider environmental conditions to be the major determinants. They have maintained that there should be a correlation between environmental variability and genetic variability. They predicted, for example, that organisms living at the bottom of the deep sea would generally be found to display little genetic variability because their environment is stable and homogeneous, whereas organisms living in the intertidal zone would display a great deal of genetic variability because their environment is a changeable one. The prediction was logical and plausible, but it failed: genetic variability has been found to be generally extremely high among organisms living at the bottom of the oceans and to be very low among organisms living in the intertidal zone.

Models

In order to carry out quantitative studies based on population genetics one needs mathematical models for the mutational production of new alleles. The first such model was proposed in 1964 by James F. Crow of the University of Wisconsin and me. It is based on the fact that each gene consists of a large number of nucleotides, so that a practically infinite number of alleles can arise; the model therefore assumes that any new mutant arising represents a new allele rather than a preexisting one. The model predicts that variability within a species, in terms of the average heterozygosity H per gene, will be determined essentially by the product of the effective population size N_e and the mutation rate v per gene per generation, rather than by N_e and v separately. Specifically, H equals $4N_ev/(4N_ev + 1)$. For example, if the mutation rate is 10^{-6} and the effective population size is 10^5, the average heterozygosity per gene will be about .286, that is, 28.6 percent of the individuals are heterozygous at each gene locus on the average. The larger either the population size or the mutation rate per gene per generation is, the closer the average heterozygosity will be to unity (or 100 percent).

The model assumes that alleles are identified at the level of the gene in terms of actual nucleotide substitutions. Most observations of variability depend, however, on the electrophoresis of proteins, which has much less resolving power and is far from revealing all nucleotide substitutions (or even all amino acid changes), so that the observed heterozygosity is less than the true amount. Even when the model is modified to take account of this problem, very large populations should, according to the neutral theory, display nearly 100 percent heterozygosity. When observations suggest otherwise, the theory is subject to criticism. For example, Francisco J. Ayala of the University of California at Davis has reported that in the neotropical fruit fly *Drosophila willistoni,* for which he estimates a very large effective population size of 10^9, he has found an observed heterozygosity of roughly 18 percent. He points out that even assuming a very low rate of neutral mutations per generation, 10^{-7}, the predicted heterozygosity is practically 100 percent.

There are at least two ways to deal with this apparent inconsistency. First of all, it is possible that the effective population size of *D. willistoni* has not been as large as 10^9 even if the apparent present size of the population is enormous. One can show mathematically that the genetic variability due to neutral alleles can be greatly reduced by a population "bottleneck" from time to time, after which it takes millions of generations for the variability to build up again to the theoretical level characteristic of a very large population maintained constantly over a long period. In this sense such neotropical species as *D. willistoni* may still show the effects of the bottleneck imposed by the last continental glaciation, between some 30,000 and 10,000 years ago. In addition the local extinction of colonies of a species, which may be fairly frequent, must reduce the effective population size.

A second possibility is that, as Ohta first proposed in 1973, the majority of "neutral" alleles may not actually be strictly neutral but rather may be very slightly deleterious. Adopting Ohta's idea, but retaining room for truly neutral mutations also, I have considered a model in which the selection coefficient s' against the mutant follows a particular distribution (the gamma distribution) [*see illustration on page 124*]. The mutation rate for variants whose negative selection (s') value is smaller than the reciprocal of two times the population size, or $1/(2N)$, can be considered the effectively neutral mutation rate v_e. It can be shown that this effectively neu-

Main intersection at a Hollywood oil field.

"This is the set where they shot 'Hello, Dolly!' What you can't see is us pumping oil behind it."

"One of the strangest places I ever produced oil," says Gulf Production Superintendent Eddie Culbertson, "was right here in Los Angeles, in the back lot of the 20th Century-Fox studio.

"They found a big oil field right under a permanent set being used for shooting the movie 'Hello, Dolly!' in 1969.

The indirect approach

"How do you get at the oil without disturbing the set, the streets, or the buildings? With a technique called directional drilling. The oil is as far as one mile from the surface location of the drill sites, which were hidden by the scenery.

"We showed them how they could have their oil and their movies."

"Directional drilling has other advantages. You can drill 20 or 30 wells from just one drill site. Or you can drill for some offshore oil without putting the drilling rig right in the water.

"Behind the scenery are 22 producing oil wells."

Keep America beautiful

"Sometimes it's a real challenge getting the oil without spoiling the landscape. Directional drilling is one way to do it. There are many more.

"Getting the oil out and preserving the environment is a challenge, but I'd say we were meeting it very well."

Gulf

Gulf people: meeting the challenge.

Gulf Oil Corporation.

THE AUDI 5000:

"...IT WOULD BE HARD TO FIND A BETTER PERFORMANCE, COMFORT AND FIVE-

W.A.W., to insiders, is WARD'S AUTO WORLD, a highly regarded trade journal read primarily by engineers and executives who work in the automotive manufacturing industry.

In the November 1978 issue, WARD'S AUTO WORLD rated 24 imported cars on the basis of driving evaluations by W.A.W. editors and writers. (The Audi 5000 used in this road test was a 1978 model equipped with automatic transmission.)

Here are excerpts from the W.A.W. report:

"A winner since the day it entered this country, Audi's front-drive 5-cyl. 5000 sedan is believed by many to be the best buy in the import luxury class."

"Not surprisingly, the 5000 grabbed a clear

Suggested retail price $9,725 P.O.E. Transportation, local taxes and dealer delivery charges additional. Fully equipped "S" model $11,360.

COMBINATION OF RIDE, ROADHOLDING, PASSENGER ROOM AT ANY PRICE."

—W.A.W.'S 4TH ANNUAL IMPORTED CAR EVALUATIONS

first-place in this year's evaluations, drawing an excellent mark in exterior design and finish and above average in all else." [i.e., engine, transmission, handling, brakes, noise, interior-front, interior-rear, restraint, controls, ride and overall rating.]

"...it would be hard to find a better combination of ride, roadholding, performance, comfort and five-passenger room at any price."

Why is the Audi 5000 the choice of so many insiders? Test-drive the 1979 model at your Porsche + Audi dealer and see for yourself.

PORSCHE+AUDI

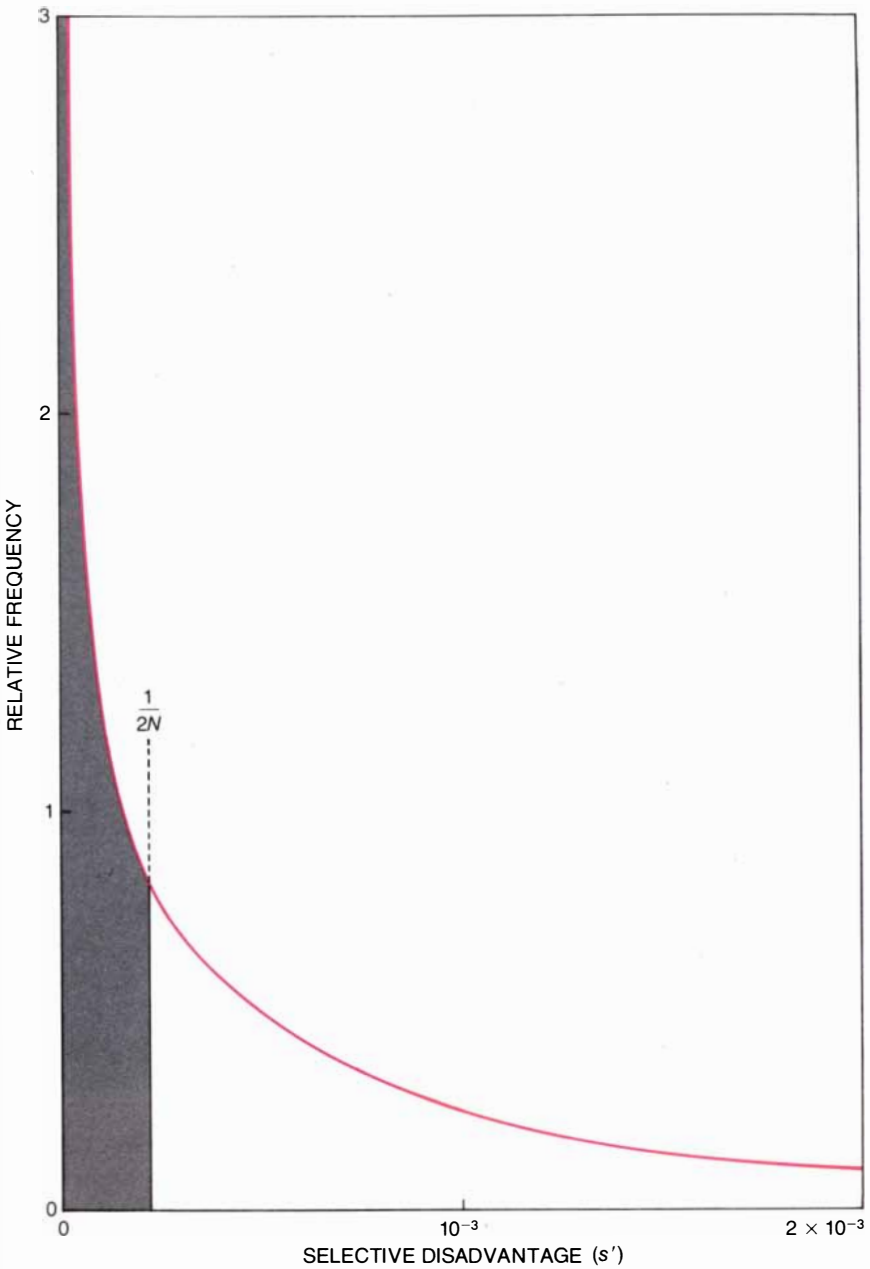
Nothing Even Comes Close

For information on leasing and overseas delivery, see your Porsche + Audi dealer.

tral mutation rate decreases as the population increases; in the case illustrated the rate is proportional to 1 divided by the square root of the population size. In this model the level of heterozygosity increases only slowly as the population increases. Moreover, given a realistic assumption about generation time, the rate of evolution in terms of mutant substitutions would be roughly constant per year for various lineages if the mutation rate per generation is constant. Note that although this explanation invokes natural selection, it is quite different from the selectionist explanation.

A Quantitative Approach

The neutral theory of molecular evolution and polymorphism that I have developed in collaboration with my colleagues Ohta and Takeo Maruyama is



THEORETICAL MODEL assumes that most "neutral" alleles are actually slightly deleterious and therefore subject to a small coefficient of negative selection s'. The model shows the frequency distribution (*color curve*) of the value of s' for mutations at various sites in a gene. Mutations whose selective disadvantage is less than 1 divided by two times the population size N are effectively neutral. The area under the curve (*gray*) that is occupied by such mutants will decrease as the population increases. As the effectively neutral fraction of the mutants thus decreases, and as more mutants are consequently exposed to negative selection, the rate of molecular evolution will be reduced. In this model level of heterozygosity (a measure of variability) increases slowly with population size, bringing predictions of neutral theory into line with observation.

distinguished from most selectionist approaches—in particular from the approach of ecological genetics—in that it aims at a quantitative description of molecular evolution, which we attempt to achieve by manipulating diffusion equations. It is a venture in what might be called molecular population genetics. Nei and his associates at Houston have contributed greatly to this effort, in particular by connecting theoretical predictions with actual observations. They have shown, for example, that the variance of heterozygosity for particular enzymes within a species can be predicted fairly accurately by the neutral theory on the basis of observations of mean heterozygosity.

Because our theory is quantitative it is testable and therefore much more susceptible to refutation when it is wrong than are selectionist theories, which can invoke special kinds of selection to fit special circumstances and which usually fail to make quantitative predictions. To test the neutral theory, however, it is necessary to estimate such quantities as mutation rates, selection coefficients, population sizes and migration rates. Many evolutionary biologists maintain that such population-genetic quantities can never be accurately determined and that consequently any theory dependent on them is a futile exercise. I, on the other hand, believe these quantities must be investigated and measured if the mechanisms of evolution are to be understood. Surely astronomers and cosmologists cannot eschew theories set forth in terms of astronomical quantities simply because such quantities are hard to estimate accurately.

Darwinian selection acts mainly on phenotypes shaped by the activity of many genes. Environmental conditions surely play a decisive role in determining what phenotypes are selected for; Darwinian, or positive, selection cares little how those phenotypes are determined by genotypes. The laws governing molecular evolution are clearly different from those governing phenotypic evolution. Even if Darwin's principle of natural selection prevails in determining evolution at the phenotypic level, down at the level of the internal structure of the genetic material a great deal of evolutionary change is propelled by random drift. Although this random process is slow and insignificant in the time frame of man's ephemeral existence, over geologic time it makes for change on an enormous scale.

People have told me, directly and indirectly, that the neutral theory is not important biologically because neutral genes are not involved in adaptation. My own view is that what is important is to find the truth, and that if the neutral theory is a valid investigative hypothesis, then to establish the theory, test it against the data and defend it is a worthwhile scientific enterprise.

JBL'S NEW L150: ITS BOTTOM PUTS IT ON TOP.

JBL's new L150 takes you deeper into the low frequencies of music without taking you deeper into your budget.

This short-tower, floor-standing loudspeaker system produces bass with depth, power and transparency that comes incredibly close to a live performance.

A completely new 12" driver was created for the L150. It has an innovative magnetic assembly, the result of years of research at JBL. It uses a stiff, heavy cone that's been coated with an exclusive damping formulation for optimum mass and density.

And it has an unusually large 3" voice coil, which aids the L150's efficiency and its ability to respond to transients (peaks, climaxes and sudden spurts) in music.

There's even more to the L150's bottom—a 12" passive radiator. It looks like a driver but it's not. We use it to replace a large volume of air and contribute to the production of true, deep bass. Bass without boom.

If you're impressed with the L150's lows, you'll be equally impressed with its highs and mids. Its powerful 1" high-frequency dome radiator provides wide dispersion throughout its range. And a 5" midrange transducer handles high volume levels without distorting. The maximum power recommended is 300 watts per channel.

The L150's other attributes include typical JBL accuracy—the kind that recording professionals rely on. Maximum power/flat frequency response. High efficiency. And extraordinary time/phase accuracy.

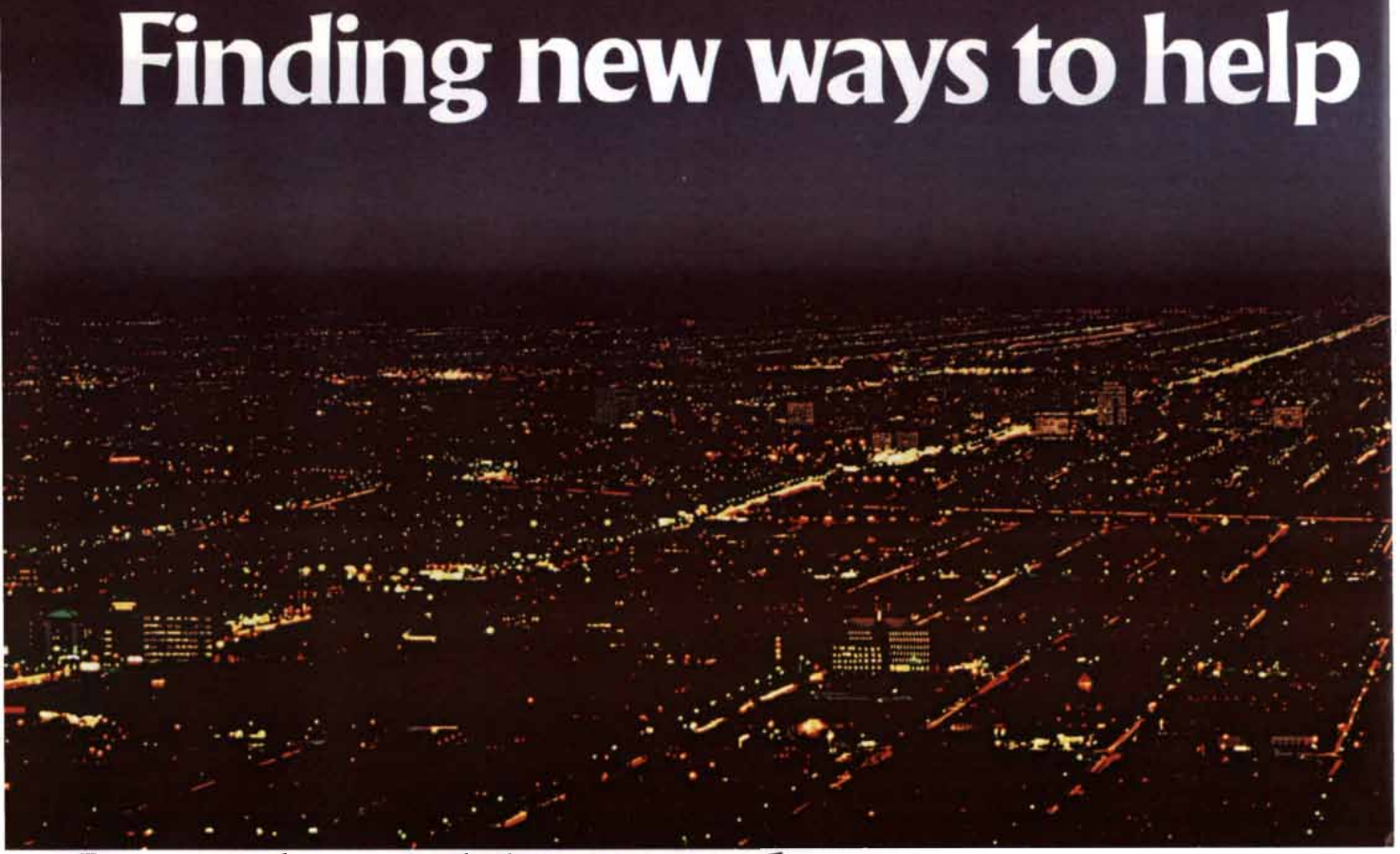
Before you believe that you can't afford a floor system, listen to an L150. While its bottom is tops, its price isn't.

James B. Lansing Sound, Inc., 8500 Balboa Boulevard, Northridge, CA 91329.

JBL

FIRST WITH THE PROS.

Finding new ways to help



Lockheed knows how.

Taking power from the oceans.

The first sea-based generating plant fueled by the warm and cold waters of the ocean is now producing electricity off the island of Hawaii.

Barge-mounted, it is a small-scale version of the free-fuel concept called Ocean Thermal Energy Conversion (OTEC). Basically, the principle uses a flow of warm surface waters and cold deep waters to drive a heat-exchanging engine, which in turn powers electric generators.

The small Hawaiian OTEC installation, for which Lockheed has major responsibility, produces only about 50 kilowatts. But it foretells huge OTEC platforms, each producing enough power for an onshore population of 100,000. In groups, the floating plants could generate all the electricity for major cities, using only the free ocean waters as fuel.

A proposed giant OTEC plant.

Harnessing the wind.

Atop a 100-foot steel tower in Clayton, New Mexico, two giant Lockheed-built blades are being wind-driven to furnish about 15% of the town's electric power.

The 60-foot-long blades are connected to a 200-kilowatt generator and rotate when the wind exceeds about 12 miles per hour, which it does most of the time.

A joint project of the Department of Energy and NASA, the Clayton wind turbine was the first of such installations. A second is now at work on Culebra Island, Puerto Rico, and a third will be operating on Block Island, off Rhode Island, this year.

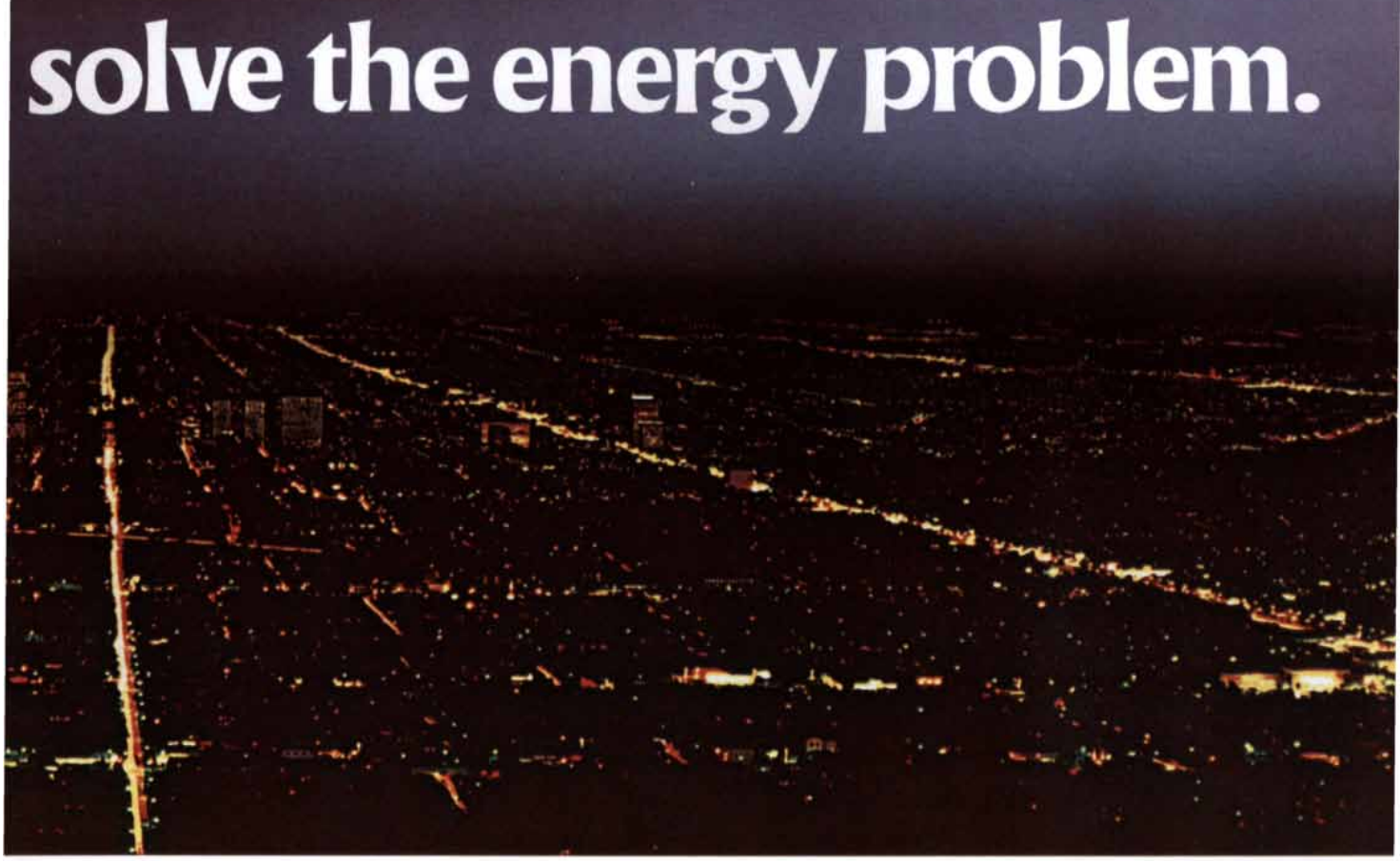
Wind-harnessing is a fast-growing activity. And experts estimate that, by the year 2000, wind turbines could supply as much as 10% of the nation's electric energy.

Polishing up a power plant.

That future outer-space missions will demand huge amounts of electrical power is certain. So Lockheed is readying a large-scale, photovoltaic generating plant for its first test flight on board NASA's Space Shuttle in the early 1980s.

The project is being developed for Marshall Space Flight

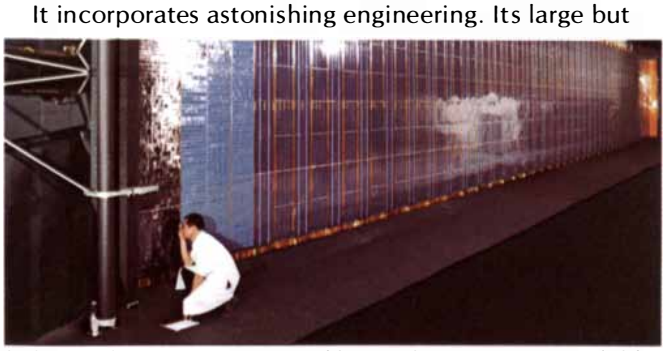
solve the energy problem.



Center. In final operating form, it will be a giant, foldable, winglike array—13½ feet wide by 105 feet long—paved with 41 modules of tightly packed solar cells that convert the sun's energy into a hefty 12½ kilowatts of power.

The great, single array—largest ever in size and specific power output—may foreshadow large orbiting stations that transmit electric power from space to Earth.

It incorporates astonishing engineering. Its large but



lightweight mass, positioned by a telescoping mast, folds into a spacecraft storage area 13½ feet long, only 2½ feet wide, and a mere 5 inches deep!

Making jetliners more fuel-efficient.

Now in many L-1011 TriStars is an exclusive fuel-saving system that's been called "the most significant advance since the autopilot."

This computer-controlled Lockheed Flight Management System automatically calculates altitude, speed and other factors. Then it acts physically, without

Control panel for Flight Management System.

pilot attention, to control engine throttles for optimum fuel-use efficiency throughout flight.

The system isn't just for new L-1011s. It can easily be retrofitted into existing planes as well.

And the savings? Big. They could total millions of dollars over the life of each L-1011 TriStar.

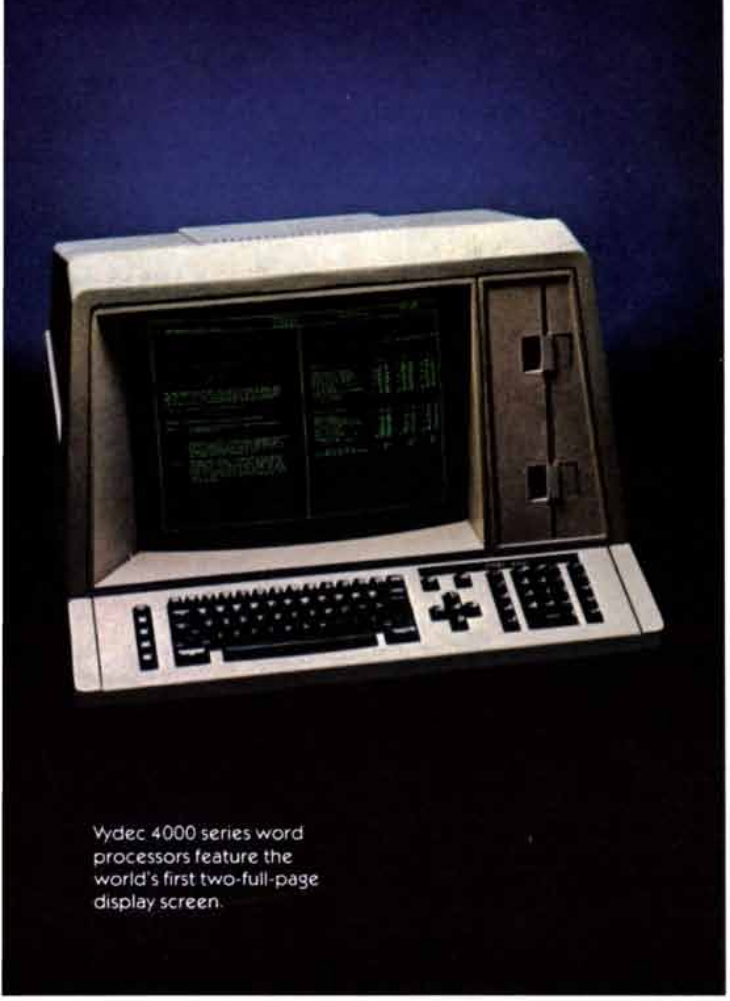
Working to keep up our energy.

Lockheed scientists and engineers are hard at work on other energy projects.

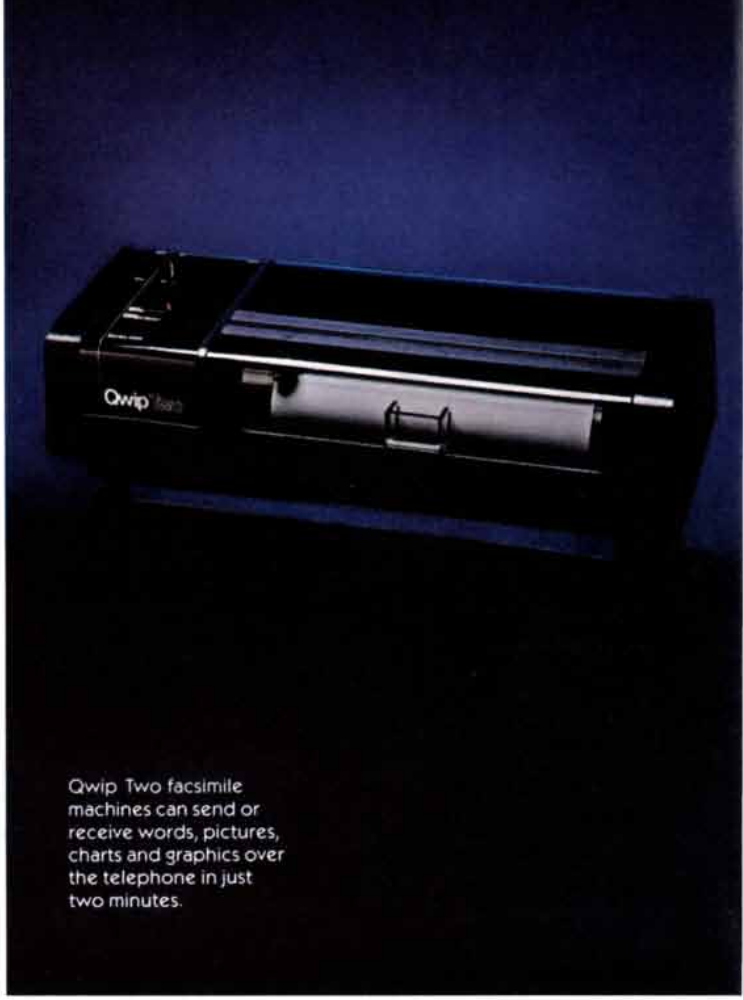
They're investigating advanced fuel-cell power systems. They're studying ways to make liquid hydrogen the fuel for future hypersonic aircraft. And they're constantly looking at new applications concepts for solar energy.

When it comes to solving complex problems by using advanced technologies, Lockheed knows how.

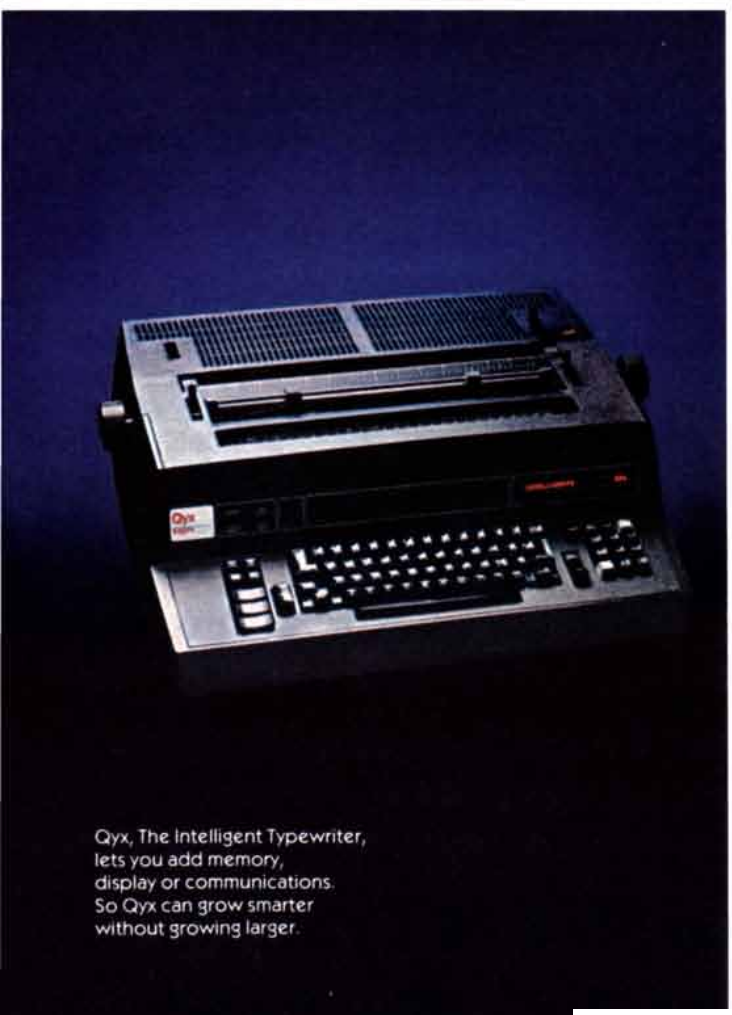
Lockheed



Wydec 4000 series word processors feature the world's first two-full-page display screen.



Qwip Two facsimile machines can send or receive words, pictures, charts and graphics over the telephone in just two minutes.



Qyx, The Intelligent Typewriter, lets you add memory, display or communications. So Qyx can grow smarter without growing larger.

Zilog MCZ-1/70 microcomputers can support five terminals, giving multiple users access to the same information.

WE'RE MAKING IT EASIER FOR YOU TO DO BUSINESS.

We are Exxon Information Systems companies. Vydec, Qwip, Qyx, and Zilog.

Into our products go the most advanced technologies in information processing, communications and storage. Out of them come easier, faster, more efficient ways for you to do business.

Vydec® for example. Vydec, Inc. just introduced the Vydec 4000 series... the first 19-inch two-full-page display word processor. The 4000 lets you work with up to 6,000 characters on the screen at one time.

Qwip® facsimile machines speed paperwork across town or cross-country in just two minutes. Because Qwip is used with your telephone, sending copies from "here" to "there" is as simple as dialing your destination.

Qyx,® The Intelligent Typewriter,® has electronic modules that let you add memory, display or communications. So Qyx gets smarter without getting bigger.

Zilog, Inc. makes highly flexible microcomputer systems that can use five high-level languages. So Zilog® microcomputers have the versatility for a broad range of business and industrial applications at a very low cost.

To learn more about our products and how they can make it easier for you to do business, just call 800-631-8181 toll free. In New Jersey, call 800-452-9300.

EXXON Information Systems

Qwip and Qyx are divisions of Exxon Enterprises Inc.

Primeval Galaxies

The first galaxies to form after the "big bang" have not been seen, but there is good reason to believe they could be. The characteristics of older galaxies suggest what those of younger ones would be like

by David L. Meier and Rashid A. Sunyaev

Since the speed of light is finite, the astronomer can in principle look back in time almost to the creation of the universe by observing objects so distant that their light has taken nearly 16.5 billion years, the current best estimate for the age of the universe, to reach the solar system. In particular there is the intriguing prospect that at those great distances one could observe galaxies in the process of formation. Since it is believed most galaxies formed the same way, the investigation of such primeval galaxies should help to explain how our own galaxy came into being 13 billion years ago. The search for primeval galaxies has not yet been successful, partly because they would be difficult to distinguish from other faint objects such as quasars. Moreover, they should be so far away that many may be too faint to be detected with current telescopes.

The situation should improve in 1985, when the superpowerful space telescope is scheduled to be launched by the Space Shuttle. With 10 times the resolution of the best optical instrument now available, the space telescope should be able to detect stellar objects 100 times fainter than any that have been detected so far. Even diffuse objects such as distant galaxies should be visible at much greater distances than they are at present. Our own work has centered on determining the properties of a primeval galaxy, for example its spectrum and the nature of its photographic image, so that astronomers will be able to identify one when they see it.

According to prevailing cosmological theory the universe began with an explosion from a superdense state in which the rate of expansion increases with the distance from the observer. The wavelength at which electromagnetic radiation from a distant object reaches the earth is increased by the velocity of recession of the object with respect to the observer. This is the well-known red shift, so named because if the radiation is in the visible region of the spectrum, it is made redder. The amount of the red shift is a measure not only of the remoteness of the object but also, since one is looking backward in time, of its age since the "big bang."

The greater the red shift, the younger the galaxy. The nearest galaxies, comparable in age to our own, are about 20 million light-years away and are receding with a red shift of .001, or .1 percent. Above a red shift of .1 quasars enter the picture. These ancient starlike objects are quite common at red shifts near 2.5. It is generally believed quasars exist at the center of otherwise normal galaxies. This suggests that the galaxies formed before the quasars, and so primeval galaxies must have a red shift of more than 2.5. An upper limit on the red shift of a galaxy in the process of formation can be determined by extrapolating backward from the present-day universe to the time when the galaxies were touching one another. That happens at a red shift of roughly 100, which represents the earliest possible time when galaxies could have been separate entities.

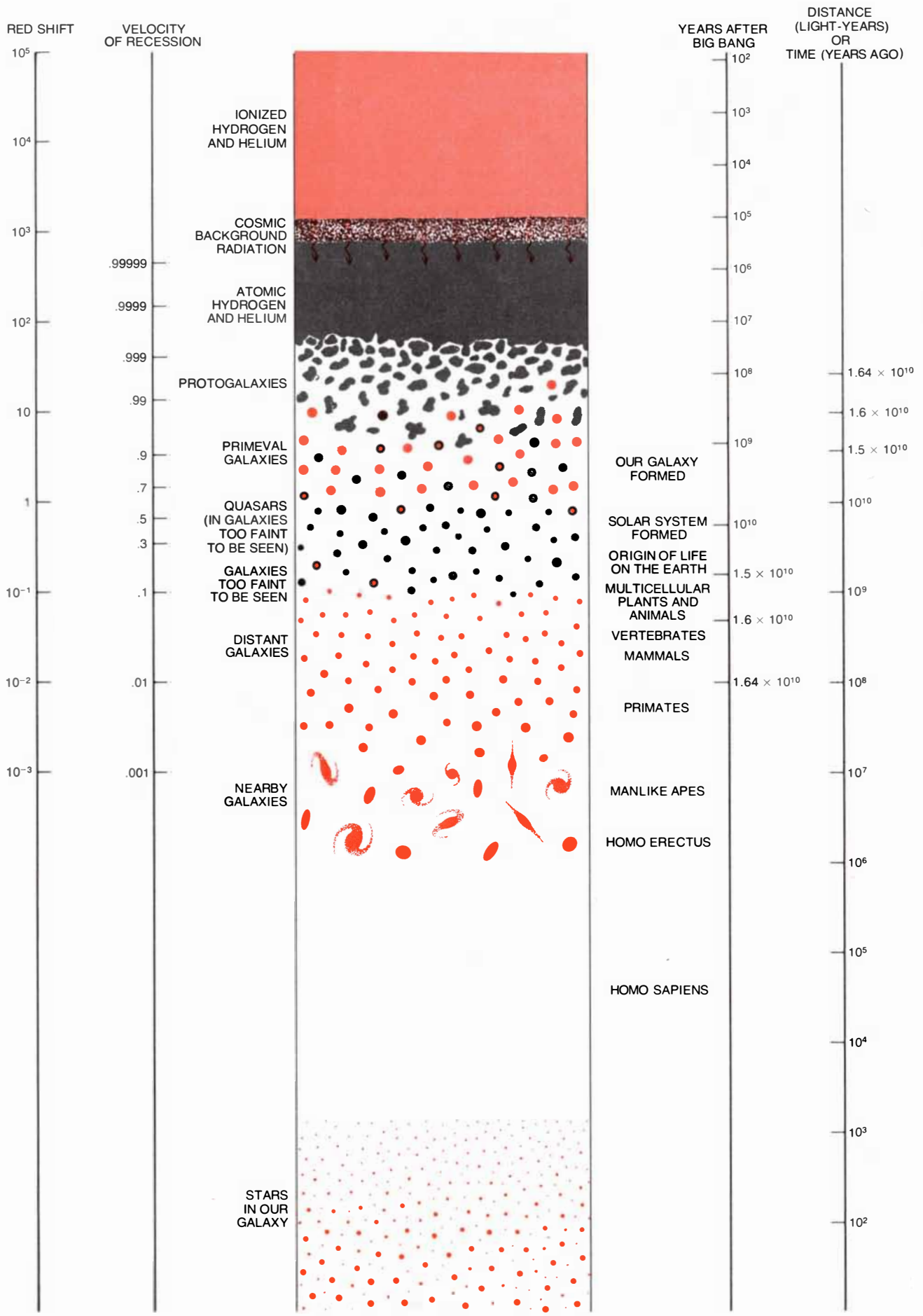
The search for primeval galaxies between red shifts of 2.5 and 100 is an extremely difficult undertaking. Indeed, no one has yet proposed a single convincing candidate for a primeval galaxy. The reason is that primeval galaxies should be quite faint at such great distances. The most distant galaxy detected so far lies at a red shift of only .75. It is even possible that the space telescope will not be able to detect normal galaxies at the high red shifts characteristic of primeval ones. There is much evidence, however, that galaxies in the process of formation have many more bright stars than they do at any other stage in their life cycle. This means that they should be much brighter than normal galaxies at the same red shift. The implication is remarkable: some primeval galaxies and perhaps all of them could be within the reach of current instruments.

The faintness of primeval galaxies nonetheless makes them difficult to distinguish from other faint objects such as quasars. Many astronomical phenomena, whatever their nature and their distance, look alike on photographic plates. Photographs that record faint objects pick up distant stars in our galaxy, distant quasars and distant normal galaxies. On sensitive photographic plates galaxies can often be distinguished from stars and quasars only by their larger and somewhat more diffuse images, and the fainter the object, the less obvious the distinction. Usually objects can be best identified by general features in their spectra: quasars have emission lines but normal galaxies do not, and galaxies have red shifts but stars do not.

None of these distinctions, however, may serve to identify a galaxy in the process of formation. The bright, hot, massive stars that are expected to exist in primeval galaxies would ionize the gas in the medium between the stars, so that the spectrum of such a galaxy would include emission lines. It is conceivable, then, that primeval galaxies might be mistaken for quasars because of their strong emission lines and large red shifts. More information is needed on the expected spectrum of a primeval galaxy and how it differs from that of a quasar.

To determine the expected spectrum we have relied on theoretical models

HISTORY OF THE UNIVERSE, presented here on a logarithmic scale, can be viewed through telescopes because of the finite speed of light. In principle the astronomer can look back in time almost to the creation of the universe by observing objects so distant that their light has taken 16.5 billion years, or the age of the universe, to reach our galaxy. The remoteness of an object is measured by its red shift: the speed of recession of the object divided by the speed of light. Here the velocity of recession is expressed as a fraction of the speed of light. To the left of the central panel are listed those astronomical objects that predominate at various red shifts. To the right of the panel, to illustrate the "look back" effect, are listed some events in the history of the earth and its biological evolution that occurred at the corresponding look-back time.

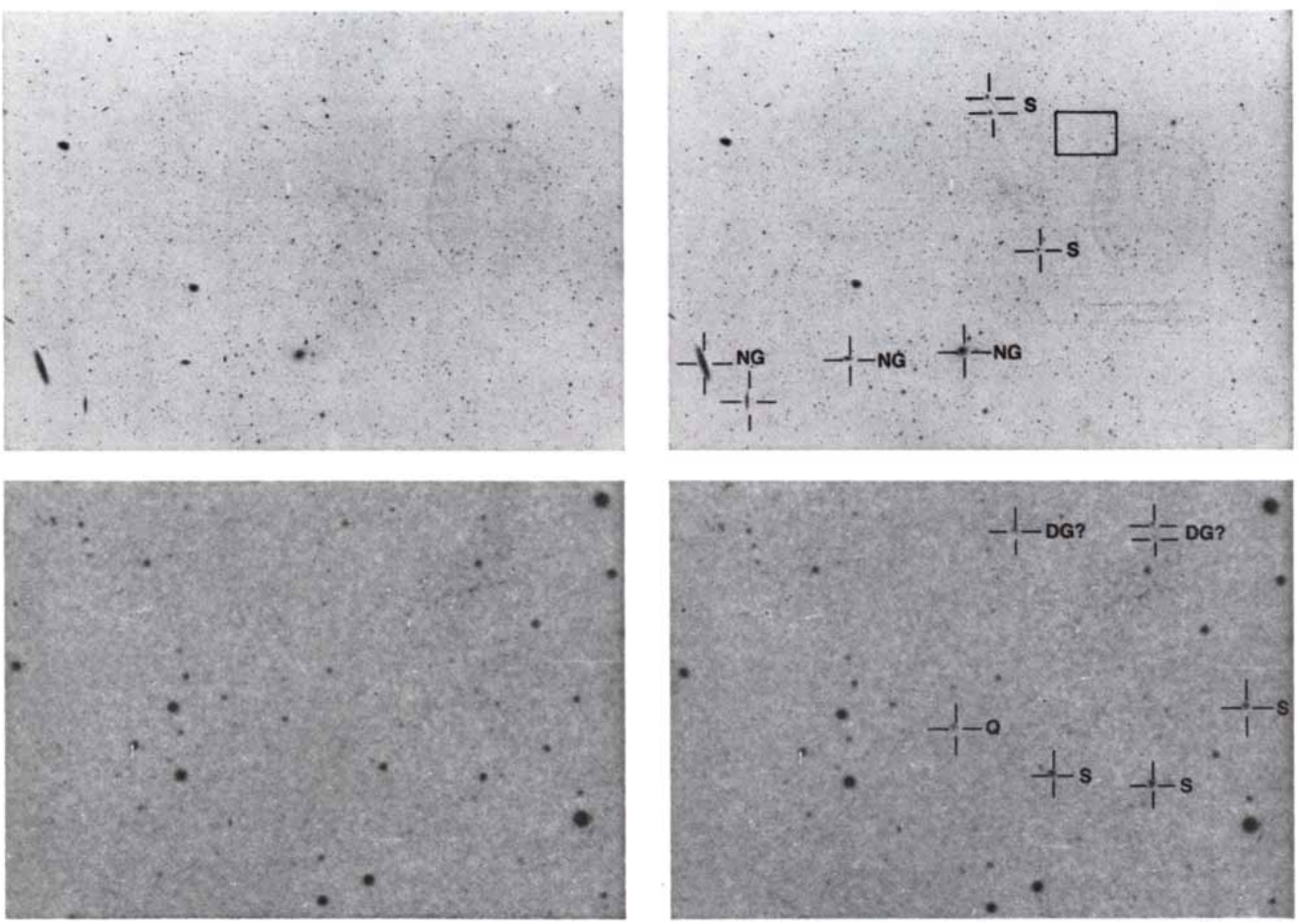


of the formation of galaxies. Because many factors in such models are unknown the models are only a crude description of the real thing. Nevertheless, they provide the only information on which astronomers can depend until the happy day when a primeval galaxy is definitively identified. Such models are based on results from many different areas of astrophysics and cosmology. They incorporate the physical processes known to govern a galaxy: the macroscopic ones of stellar and gas dynamics that determine the galaxy's overall size and shape and the microscopic ones of stellar formation, evolution and death that determine the properties of the galaxy's constituent stars. Apart from the observational resemblance between primeval galaxies and quasars the two kinds of object may be related in a deeper sense. Quasars are physically distinct from galaxies, but as we have mentioned they seem to exist in galaxies and they were more abundant in the infant universe. As a result they are probably connected in some way with the formation of galaxies.

The most widely accepted theory of galaxy formation is the theory of gravitational instability, which maintains that galaxies condensed out of the hot cosmological fluid that expanded from the big bang. If a region of the early universe happened to have a density higher than that of the surrounding regions, it would be gravitationally attracted more to itself than to the ambient material. If such a region, characterized as a perturbation, was free of gas or radiation pressure, it would contract under its own gravity and hence would increase in density. This contraction process would create "droplets" out of a universe that previously was quite homogeneous.

The scale of the perturbations probably ranged from the mass of a globular star cluster (10^6 times the mass of the sun) to the mass of a large aggregate of galaxies (10^{15} solar masses). The perturbations that are the size of galaxies (10^{11} solar masses) are called protogalaxies. Initially each protogalaxy would expand with the rest of the universe, but it would do so at a slightly lower rate. After a few hundred million years it would stop expanding even though the universe continued to expand. At that point the protogalaxy would in effect be detached from the universe, free to collapse on itself and form a galaxy.

What do we mean when we say a galaxy has formed? For a protogalaxy to become a galaxy two things must happen. First, a population of stars must form from the protogalactic gas. Second, the gas and stars must come together to form the well-ordered structure of a galaxy. The second process is fairly well understood. With a high-speed computer it is straightforward to follow the collapse of a model protogalaxy by solving the equations of motion for stars and gas under the influence of their mutual gravity. Not much is known, however, about how the stars actually formed out of the gas in the course of its collapse. What little is known comes from observations of our galaxy and nearby galaxies. Such observations suggest possible events that could trigger star formation and also the



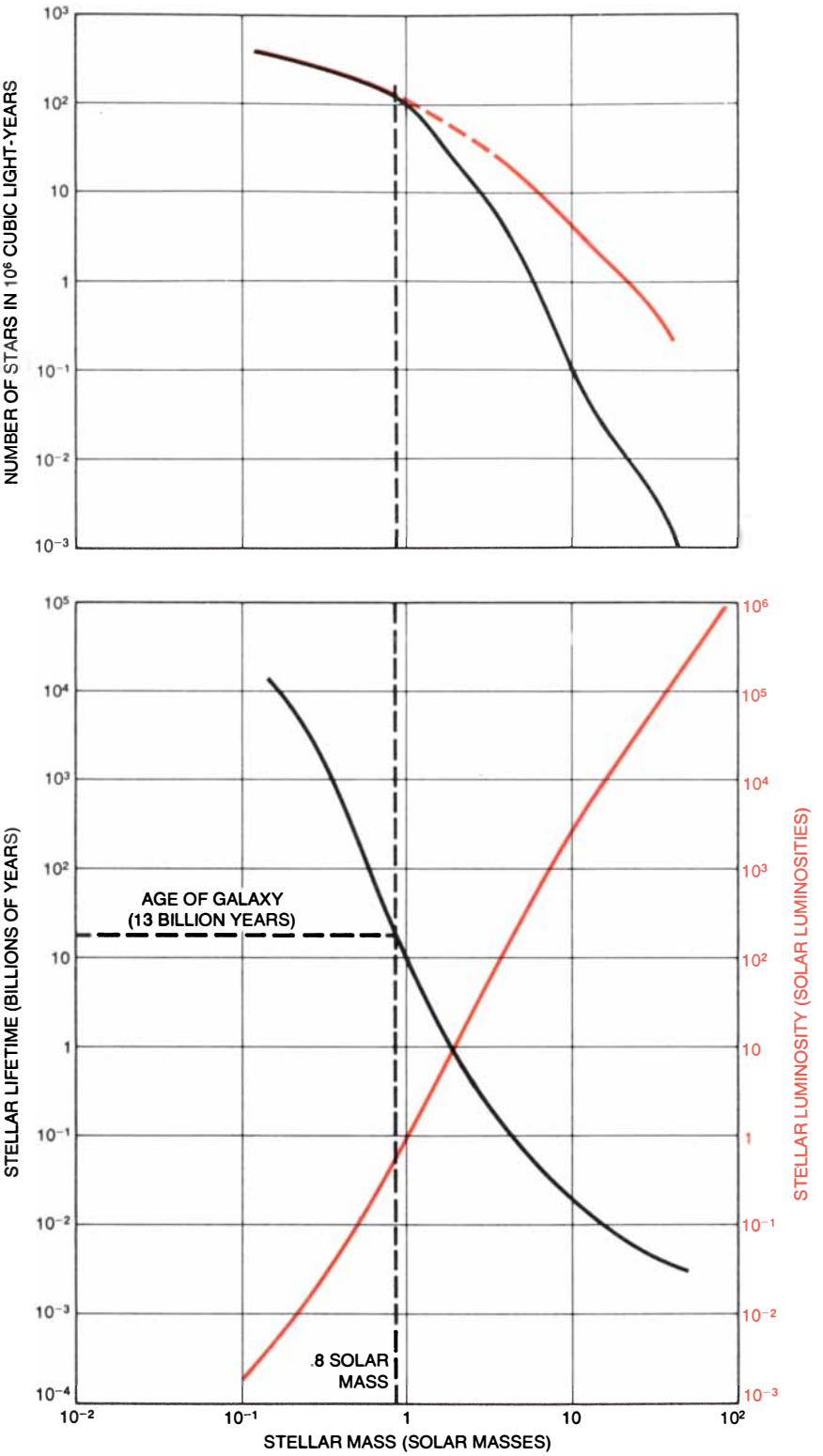
FAINTNESS OF PRIMEVAL GALAXIES makes them difficult to distinguish from other faint objects such as quasars. The negative print at the top left, made with the 48-inch Schmidt telescope on Palomar Mountain, is of a fairly large (1.5 by two degrees) region of the sky at the outer edge of the Virgo cluster of galaxies 45 million light-years from our galaxy. The positions of a few nearby galaxies (*NG*) and of some fairly bright stars (*S*) in our galaxy are marked on the duplicate of the print at the top right. The print also records a multitude of extremely faint objects. Some of them are possibly primeval galaxies. Most of the primeval galaxies, however, are probably too faint to be seen. The print at the bottom left is of the boxed region (eight by 11 minutes of arc) of the top photograph. The positions of several faint stars (*S*), a quasar (*Q*) and candidates for distant galaxies (*DG?*) in this region are marked on the duplicate at the bottom right.

proportions of stars of different masses that are created in a burst of star formation. With this information it is possible to develop a computer-generated semiquantitative description of galaxy formation.

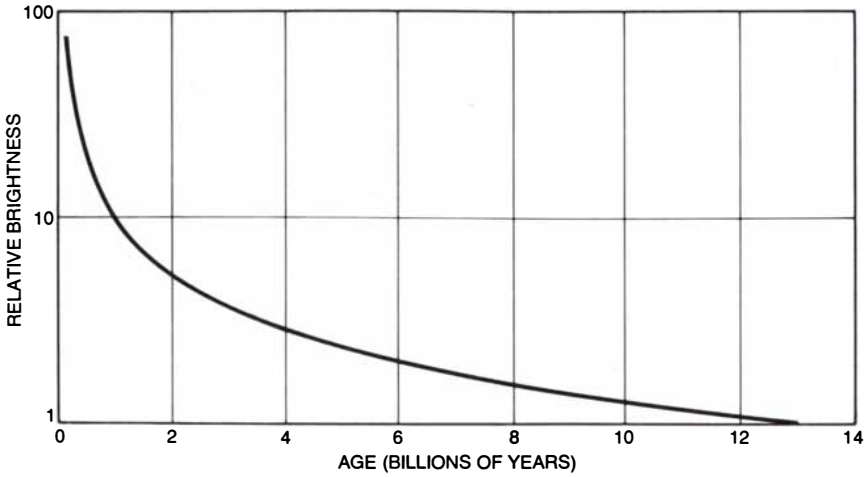
Like galaxies, stars probably form from the interstellar gas because of gravitational instabilities. Unlike galaxy formation, however, star formation is generally not a spontaneous process because the pressure of the interstellar gas is usually sufficient to prevent perturbations as small as stars from collapsing. If the gas is compressed by some violent phenomenon, the density can increase to the point where gravity overcomes the gas pressure and the perturbations contract into stars. In our galaxy the violent phenomena include a spiral density wave of the kind that is responsible for the pinwheel structure of disk galaxies and also shock waves from supernova explosions and from the expanding regions of ionized gas that surround hot massive stars. In other galaxies star formation may also have been triggered by collisions or near collisions with neighboring galaxies. In a protogalaxy the violent collapsing motion itself is probably the chief process that causes stars to form. A protogalaxy can be regarded as a system of gas clouds racing in orbit and giving birth to stars when the clouds collide.

Observations of the stars in the vicinity of the sun have made it possible to estimate the distribution of stars of different masses that will be created in any burst of star formation. Such a distribution is termed the initial mass function. The present-day local mass function—the mass distribution of stars in the neighborhood of the sun—includes many stars weighing less than .8 solar mass but only a few weighing more than that. More massive stars are scarce because such stars burn faster and hence do not live as long. The sun will live for 10 billion years, but a star with only 10 times the sun's mass would shine with 20,000 times the luminosity and would live for only 10 million years, or a thousandth as long. Stars weighing more than .8 solar mass have lifetimes that are shorter than the age of the galaxy, and so such stars that formed when the galaxy was created 13 billion years ago have died out.

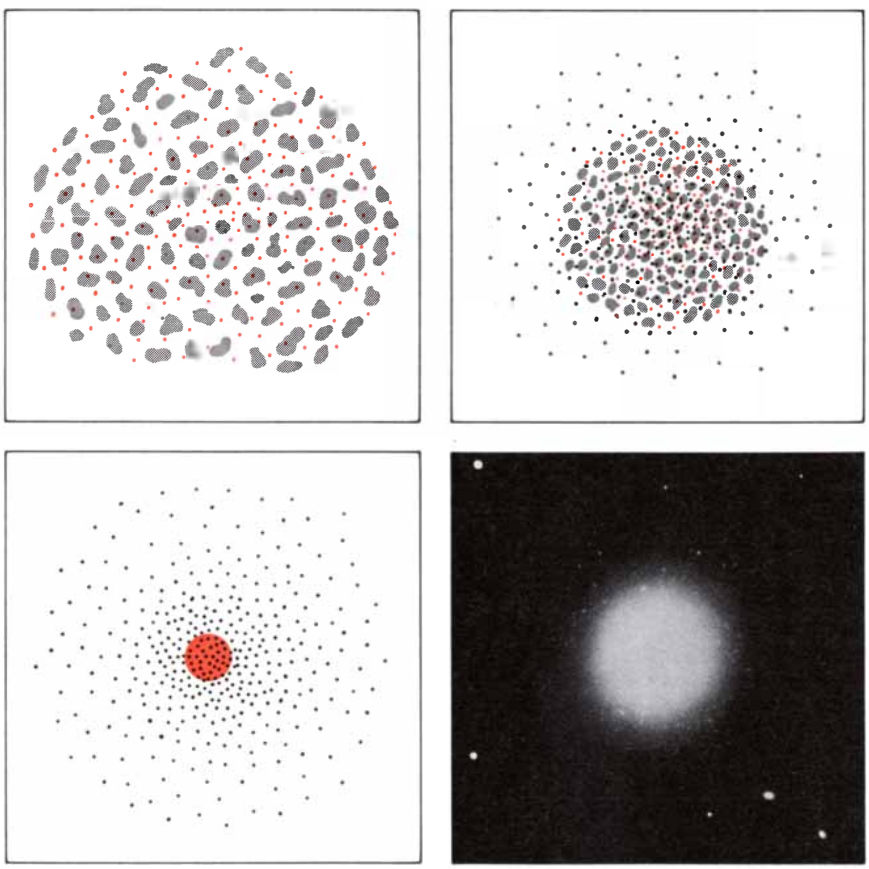
Stars weighing less than .8 solar mass have lifetimes that are longer than the age of the galaxy. The stars in this lower mass range have been increasing in number in the course of many bursts of star formation, and so the relative mass distribution of the stars must be identical with that of the stars created in any one burst, assuming of course that stars have always formed the same way. In other words, below .8 solar mass the present-day local mass function is



INITIAL MASS FUNCTION, the distribution of stars of different masses that will be created in a burst of star formation in a galaxy, is shown in the top diagram. The black curve represents the current mass distribution of stars in the neighborhood of the sun. Stars with a mass of .8 times the mass of the sun have lifetimes that are longer than the age of the galaxy, and so the mass distribution of these stars is the same as it was initially. The more massive the star is, however, the faster it burns (*bottom black curve*). Stars with a mass of more than .8 solar mass have lifetimes that are shorter than the age of the galaxy, and so the stars that formed 13 billion years ago when the galaxy was created have died out. At masses above .8 solar mass the initial mass function (*top colored curve*) can be determined by extrapolating backward in time, replacing the dead stars with live ones. Although the number of stars decreases with increasing mass, the decrease is not sufficient to overcome the increase in stellar luminosity with mass (*bottom colored curve*). This means that the luminosity of a newly formed cluster of stars is quite high and is accounted for chiefly by the bright and short-lived massive stars belonging to the cluster.



LUMINOSITY OF A STAR CLUSTER is highest when it forms because of the multitude of bright massive stars. As these short-lived stars burn out, the luminosity of the cluster decreases rapidly. "Relative brightness" is determined by dividing the luminosity of a population of stars by the luminosity at the age of 13 billion years, the age of our galaxy. That is, the relative brightness of a stellar population is a measure of how many times brighter it once was than it is now.



GALAXY FORMATION according to some computer models begins with a cluster of small massive clouds of gas about 200,000 light-years in extent (*top left*) that collide, collapse and form a population of bright stars (*colored dots*). As the clouds fall inward into a smaller volume because of their mutual gravitational attraction the rate of star formation increases. After about 200 million years the protogalaxy is roughly 100,000 light-years across (*top right*) and the first population of stars (*black dots*) has died out. The clouds of gas continue to contract at an increasing rate. After about 300 million years the gas is packed into the center of the galaxy, bringing star formation to a peak rate (*bottom left*). The dense nucleus of stars in the galaxy is perhaps only 10,000 light-years in extent. At this stage the galaxy is one of the brightest objects in the universe. From a great distance, however, even such a luminous galaxy would resemble a faint star. After this stage star formation ceases because the gas in the galaxy has been exhausted. As brighter stars die out, leaving behind the older and fainter ones, the galaxy shines less brightly but more steadily as a giant spherical structure, such as the galaxy M87 (*bottom right*).

the initial mass function. Above .8 solar mass the initial mass function can be determined by extrapolating backward in time by replacing the dead stars with living ones. The resulting relation, which was derived by Edwin E. Salpeter of Cornell University, indicates that whenever a group of stars is created, for each increase by a factor of 10 in the mass there is a decrease in the number of stars by a factor of 22.

If the Salpeter relation holds for all episodes of star formation, it has startling implications for galaxy formation and for the search for primeval galaxies. Although the number of stars decreases with increasing mass, the decrease is not sufficient to overcome the increase in stellar luminosity with mass (a factor of about 10,000 for each increase by a factor of 10 in mass). This means that the luminosity of a newly formed cluster of stars is quite high and is due chiefly to the bright and short-lived massive stars. After a short time, however, the bright stars die out and the luminosity of the cluster rapidly decreases. As a result a primeval galaxy, in which stars would be expected to form at a much higher rate than they are forming in present-day galaxies, should be quite bright.

Richard B. Larson of Yale University has developed complex computer models of three types of galaxy—spherical, elliptical and disk—in the process of collapsing. The spherical galaxy is the brightest, and so it is the best candidate for a detectable primeval galaxy. After the protospherical galaxy reaches its maximum expansion its gas begins to contract and its small internal clouds accelerate, collide and form stars. Most of these stars will continue to burn for billions of years and their orbits will be responsible for the overall shape of the galaxy. The massive stars, however, soon explode violently, hurling into the interstellar gas the nuclei of the heavy elements synthesized by nuclear reactions in their interior.

Once the stars have formed, their dynamical behavior is different from that of the gas. The orbits of the stars simply conform to the shrinking shape of the galaxy. The gas clouds, however, tend to collide with other gas clouds, so that their orbits decay and the clouds plunge toward the center of the galaxy. This causes the entire cycle of cloud collision, star formation, stellar explosion and heavy-element enrichment to proceed even faster. Eventually the gas is packed into the center of the galaxy, bringing star formation to a peak rate. Within a few thousand light-years of the center hosts of stars are shining brightly and exploding, gas is glowing and the dust of heavy elements is radiating copiously in the infrared. At this stage the galaxy is one of the brightest objects in the universe, hundreds of times more luminous than the galaxies of today.

Then the activity of the galaxy abrupt-

Add your personal touch to our extraordinary brandy.

Personalize every bottle of our XO Rare Reserve Brandy that you give as a gift this year. Just ask us to print specially-designed neck labels for you that say, *Compliments of,* followed by your name, or your company's logotype. Or, you can order elegant labels that read, *For the Personal Enjoyment of,* followed by the name of your friends or clients. What's more, you can select other labels that wish *Happy Holidays, Happy Birthday, Happy Anniversary,* and even *Happy Father's Day.*

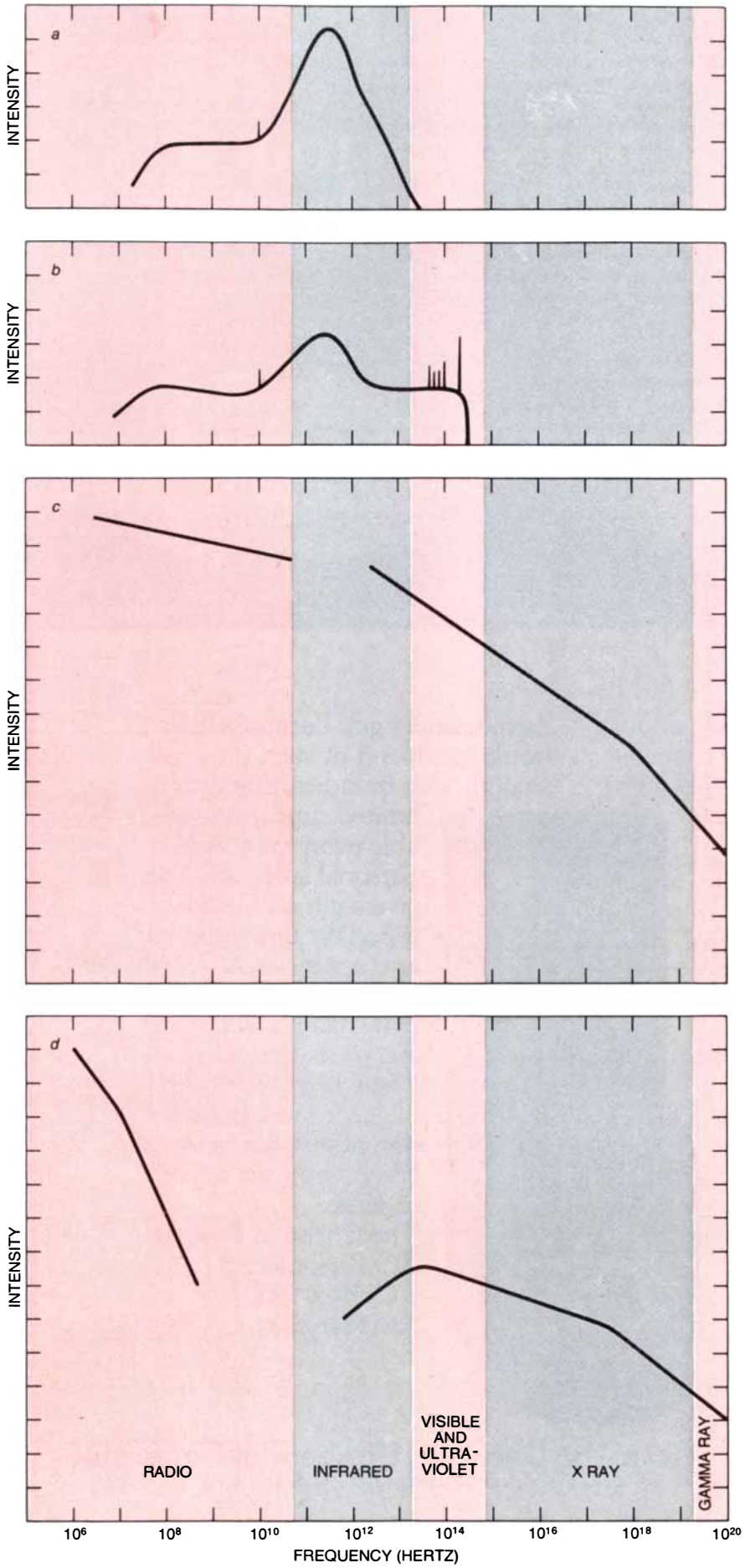
Of course, even without your personal label, XO makes an extraordinary gift. Because every bottle is a blend of such unusually smooth, rich brandies, that only a limited supply is available every year. Add a personal label, and you give a gift that will be especially remembered and appreciated.

And naturally, these personalized labels are yours free from The Christian Brothers.

For more information about our personal labels, see your retailer or write:
The Christian Brothers,
XO Personalized Labels,
P.O. Box 99607,
San Francisco, CA 94109.

XO Rare Reserve Brandy from The Christian Brothers of California.

WORLDWIDE DISTRIBUTORS: FROMM AND SICHEL, INC., SAN FRANCISCO, CALIFORNIA, U.S.A. BRANDY: 80 PROOF.



SPECTRA OF MASSIVE STARS at various stages in their life cycle will contribute to the spectrum of a primeval galaxy. Massive stars shrouded by clouds of dust emit radio waves that penetrate the dust (*a*). The dust absorbs radiation of longer wavelengths but reradiates the energy of the radiation in the form of thermal emission in the infrared. Carbon monoxide in clouds of molecules contributes a sharp emission line at 115 gigahertz (billion cycles per second). The intensity of visible and ultraviolet radiation is much greater for stars in which the shrouding dust has been dispersed or destroyed (*b*). An exploding massive star leaves a remnant that emits radio waves and X rays (*c*). Pulsars also emit radio waves (*d*).

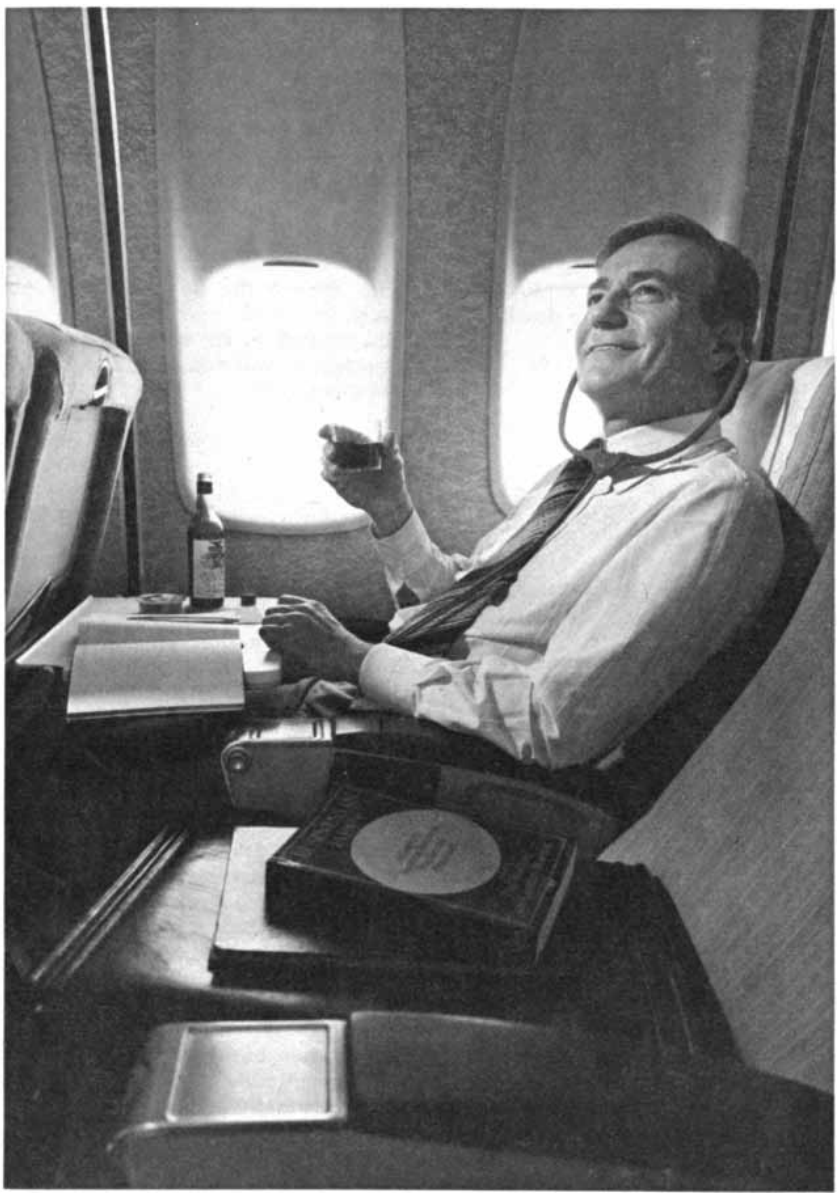
ly stops. The gas has been exhausted; there is no more left to collapse into stars. The remaining bright stars burn out quickly. As the stars with average lifetimes begin to die, leaving the older and fainter ones behind, the galaxy shines less brightly but more steadily as a giant spherical structure.

Such models of galaxy formation suggest that most primeval galaxies were born between 100 million and a few billion years after the big bang. That puts them at red shifts of between about 3 and 30, which is in accord with the earlier estimate we made on the basis of quasar red shifts. Through current telescopes most primeval galaxies with lower red shifts would be barely visible and those with higher red shifts would be completely invisible. Exceptionally bright primeval galaxies, however, should be quite apparent. The space telescope may be able to detect all primeval galaxies except those with the highest red shift.

The model also suggests that the ratio of the brightness of a galaxy's nucleus to the brightness of its outer regions is greater for a primeval galaxy than it is for a normal one. This means that the images of primeval galaxies resemble the images of quasars and stars, and so it is understandable that they have been difficult to distinguish. At a distance of 16 billion light-years the bright nucleus, although it would be thousands of light-years in diameter, would be only a second of arc in apparent diameter. The space telescope will be able to clearly resolve the structure of such an object. Indeed, thousands of primeval galaxies should eventually be detectable in a square degree of sky.

Working with Beatrice M. Tinsley of Yale University, we undertook to calculate the expected spectrum of a primeval galaxy. We were able to secure much information from our own galaxy, as had been the case with the determination of the initial mass function of the stars in the universe. In a primeval galaxy the infant stars and the objects associated with them should contribute to the spectrum of the galaxy. Although

Helpfulness. Part of JAL's Executive Service.



At Japan Air Lines we recognize the special needs of the business traveler. To help the executive planning to do business in the Orient, JAL initiated the Executive Service Program. It starts working for you the moment you book.

Even before the flight begins, JAL offers informative handbooks on doing business in the Orient. We also

offer those all-important bilingual business cards, Japanese or Chinese on one side, English on the other. You'll find both these offerings worth the nominal extra charge you pay for them.

On board, helpful hospitality. An Executive Service Section just for full-fare passengers. And to cheer the way, free earphones and free drinks. (As you might expect, not applicable to passengers traveling on economy class promotional fares.)

At the end of the flight, you'll find JAL's Executive Hotel Service offers rooms at special rates and a 6 p.m. checkout at major hotels throughout the Orient. Our Hospitality Desks are everything the name implies. And JAL's Executive Service Lounge in Tokyo's Imperial Hotel serves as your elegant office away from the office.

Otetsudai. Helpfulness. Because we never forget how important you are.

The way we are is the way we fly.

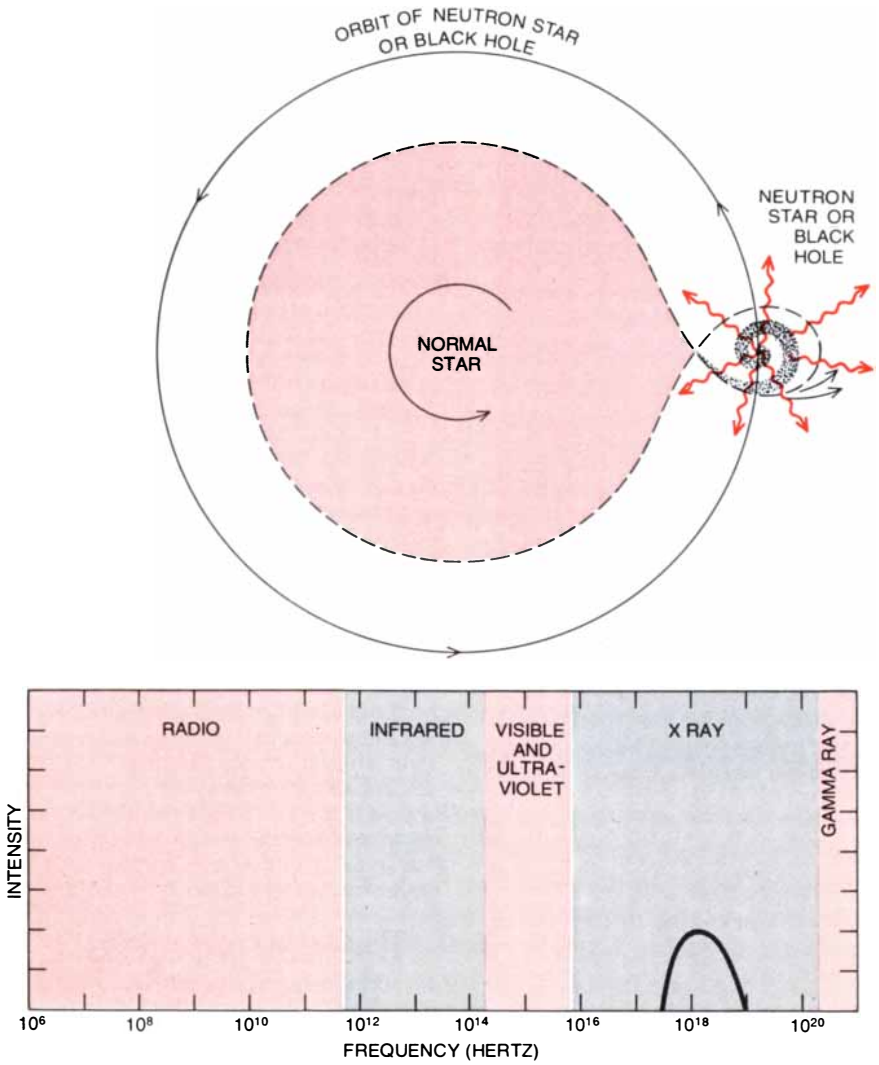
JAL

JAPAN AIR LINES

these stars and objects are distant and ancient, they are expected to be almost identical with the stars and objects in our galaxy. This close similarity is quite likely because the model suggests that heavy elements are created and distributed early in the collapse of a protogalaxy, giving an infant star the same chemical composition, and hence the same spectrum, as an infant star in our galaxy. As a result all the ingredients for determining the properties of remote and ancient primeval galaxies are present in our galaxy. To calculate the spectrum of a primeval galaxy we simply estimated the prevalence of various objects in such a galaxy, took their spectra from how they appear in our galaxy and summed the spectra according to the prevalence of the objects.

The task of determining the prevalence of objects in a primeval galaxy is simplified by the fact that most of the objects are the offspring of the massive stars that continually form and die. (The low-mass stars contribute only visible light and live too long to create any interesting objects.) Moreover, the offspring objects are as short-lived (compared with the collapse time of the protogalaxy) as the massive stars themselves, and so the number of these objects that contribute to the spectrum of the primeval galaxy is proportional to the rate of star formation. The rate of star formation in a bright primeval galaxy may be 3,000 times greater than the present rate in our own galaxy, so that we expect 3,000 times the number of massive stars and their short-lived offspring in a primeval galaxy.

The kinds of such offspring can be determined by following the evolution of a massive star from birth to death. The star is believed to form inside a cold (10 degrees Kelvin) cloud consisting chiefly of molecular hydrogen (H_2), carbon monoxide (CO) and dust grains of silicate and carbon. Molecular hydrogen does not emit much radiation, so that the only important waves emanating from the cloud are associated with the microwave-radio spectral line of carbon monoxide (at a frequency of 115 gigahertz). The star shines brightly in the visible and ultraviolet regions of the spectrum, although such radiation is absorbed by the cloud and hence cannot be seen outside it. The radiation heats the dust grains shrouding the cloud to about 30 degrees K., at which temperature the grains reradiate the energy of the star in the form of thermal emission in the infrared. A thick shell of ionized gas also forms around the shrouded star. Although the visible and ultraviolet radiation emitted by the star and gas cannot penetrate the dust grains, the radio waves can.

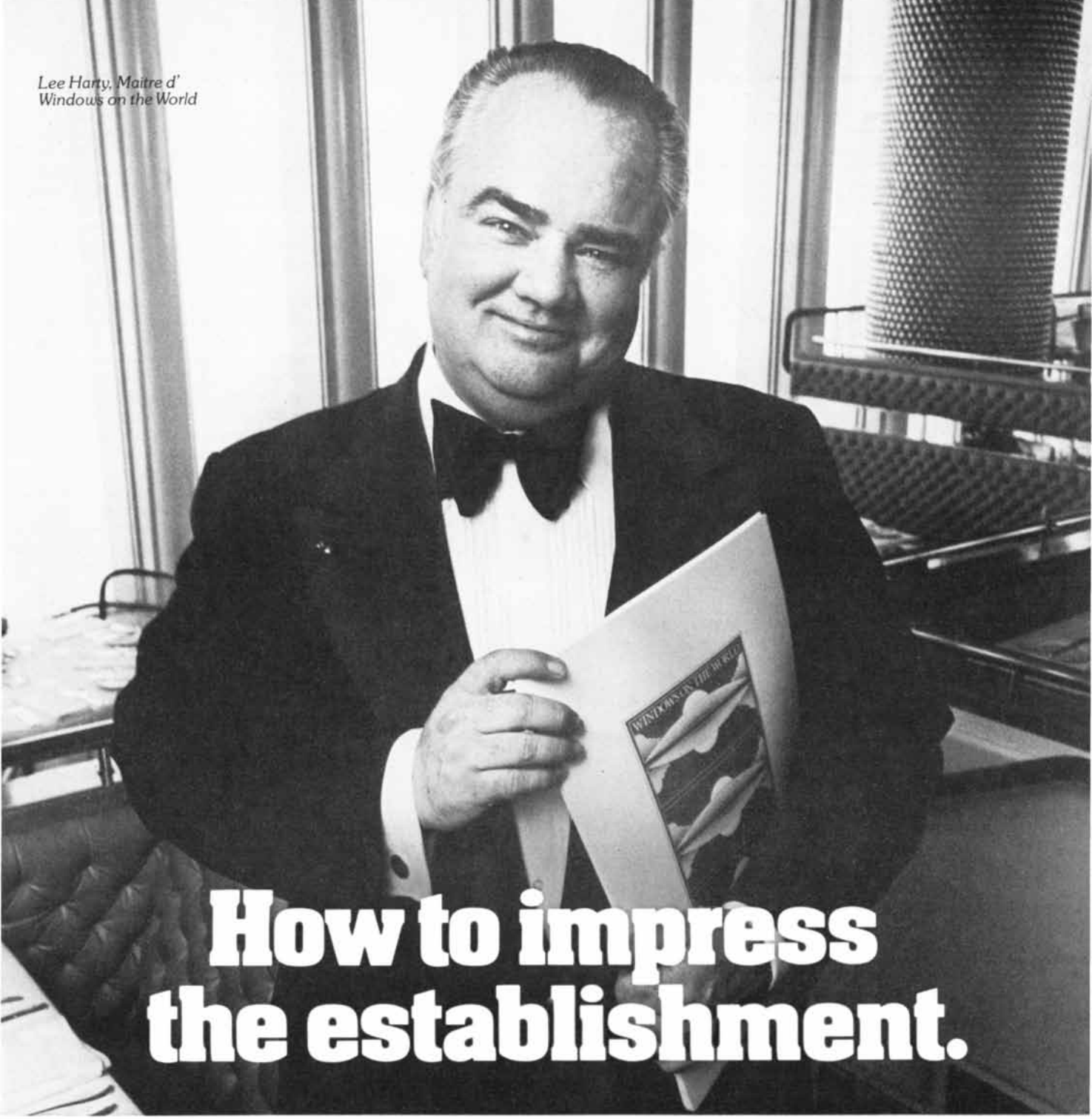


NEUTRON STARS AND BLACK HOLES that are the compact remains of supernova explosions could contribute to the spectrum of a primeval galaxy in an unusual way. If the exploded star was originally a member of a binary system in which the other star had a longer lifetime, the neutron star or black hole could be left orbiting the longer-lived star (*top*). Matter could then be transferred to this compact object by means of a stellar "wind" or gravity, the object pulling material off the star. This process would give rise to strong X-ray emissions (*bottom*).

Soon the shell of ionized gas expands, unveiling the star by dispersing or destroying most of the dust grains. As a result the intensity of the infrared radiation decreases sharply and the intensity of the visible radiation increases. (The Orion nebula in our galaxy is an example of a region of visible hot stars and ionized gas.) When the massive star reaches the end of its life cycle, it explodes violently as a supernova. A supernova explosion would be a spectacular event in a normal galaxy because of the comparative faintness of the other stars. In a primeval galaxy, however, a supernova would go unnoticed at visible wavelengths because of the abundance of other bright, massive stars.

The situation is different at other wavelengths. The compact remnant of the supernova, which could be a pulsar or a black hole, and the expanding shell of gas surrounding the remnant would appreciably increase the radio and X-ray emissions of the galaxy. If the compact remnant is a pulsating neutron star, it would strongly emit radio waves and X rays. It could also contribute X rays in another unusual way. If, as is often the case, the original star was part of a binary system with a star whose lifetime is longer, the neutron star could be left orbiting the longer-lived star. Matter could then be transferred to the neutron star by means of a stellar "wind" or gravity, the neutron star pulling material off the other star. The result would be an X-ray-emitting binary system like those that have been found in our galaxy. A similar picture would apply if instead of a neutron star the binary remnant was a black hole.

Pulsars and supernova remnants probably also contribute to the spectrum by giving rise to cosmic rays. The rapidly rotating magnetic field of the pulsar and the strong shock wave of the



How to impress the establishment.

"We may be 107 stories high, but we're really down-to-earth. You see, at Windows on the World, the whole world is at home. What impresses us? People who know how to enjoy themselves. Whether they're having a half dozen impeccable clams on the half shell with a lusty Italian white. Or rack of young lamb with a pedigreed French red.

"And time after time, it's the Diners Club member who's enjoying these delights. And that's how I know one when I see one."

Diners Club members get more out of life. So it's only right they get more out of the card they carry. Diners does more things, in more places, than any other travel and entertainment card.

The Doublecard. Takes the advantages of a credit card. And doubles them. Diners membership entitles you to two cards. To separate business and personal charges, you can put your name and title on one, and just your name on the other. Or you can give the second card to a member of your family.

The card you can cash in on. Diners Club membership offers you the Diners Cash Advantage℠ — a separate reserve of cash of up to $15,000* from the Chase Manhattan Bank, N.A.

And these are just some of the things Diners Club is constantly doing to help the Club serve its members.

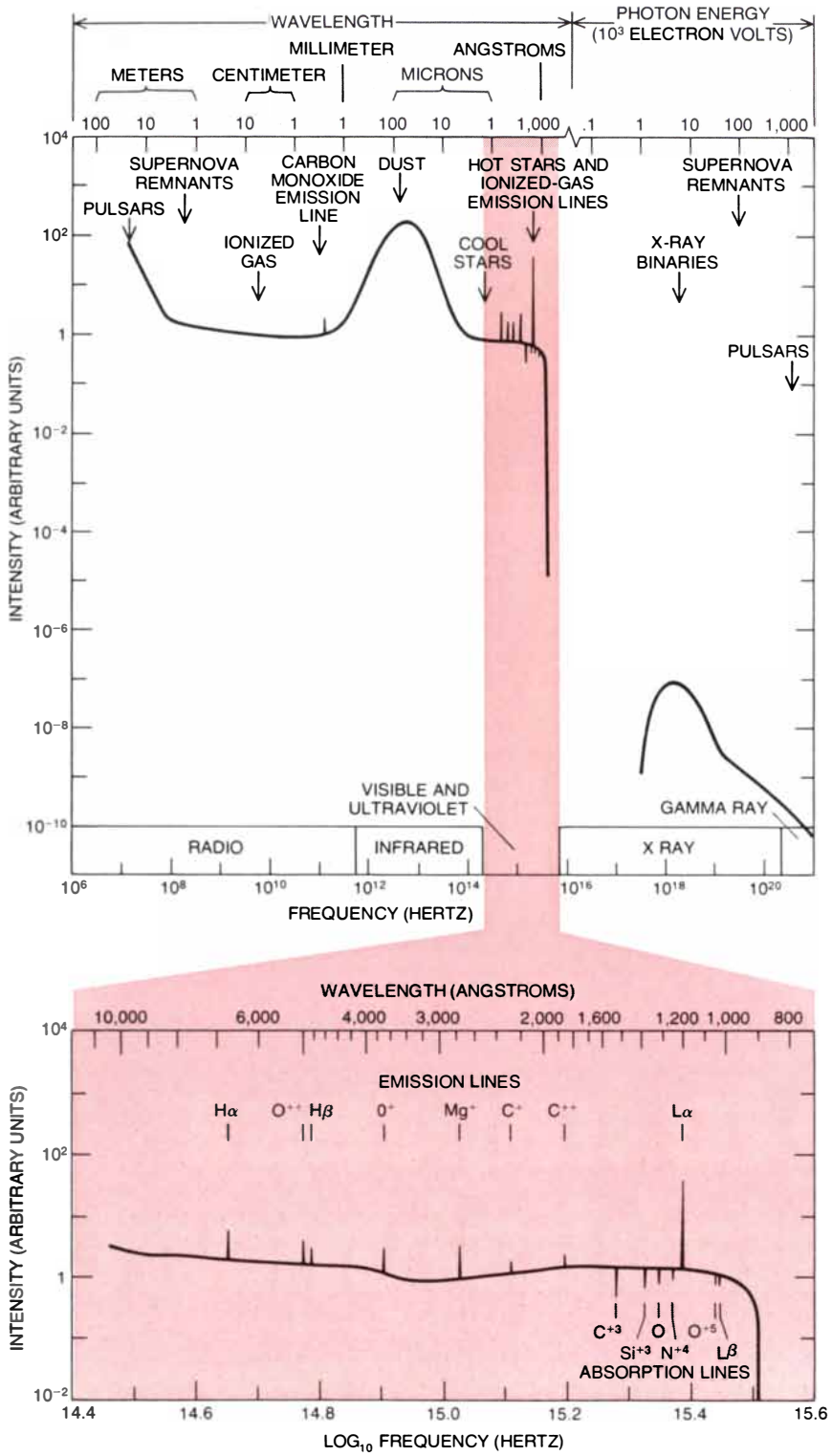
If you aren't a member, why not call our toll-free number for an application? We think you'd fit right in. After all, we know one when we see one, too.

*Minimum $2,000. Subject to legal restrictions.

THE DINERS CLUB MEMBER. YOU KNOW ONE WHEN YOU SEE ONE.

If you want to be one, call 1-800-525-7000

(In Colorado call collect 303-770-7252)



EXPECTED SPECTRUM OF A PRIMEVAL GALAXY (*top*) is obtained by adding the spectra of the contributing phenomena, such as the ones in the illustrations on pages 136 and 138. This spectrum is the intrinsic one; the measured spectrum would be shifted in frequency to the left according to the precise red shift. The greatest uncertainty in determining the spectrum comes in estimating the ratio of the number of massive stars that are shrouded by dust to the number that are not shrouded. Here the ratio is assumed to be 1 : 1. The visible and ultraviolet regions (*color*), which are shown in an expanded view at the bottom, include emission lines of hydrogen (the Balmer lines $H\alpha$ and $H\beta$ and the Lyman line $L\alpha$), carbon (C^+ and C^{++}), oxygen (O^+ and O^{++}) and magnesium (Mg^+) and absorption lines of carbon (C^{+3}), silicon (Si^{+3}), oxygen (O and O^{+5}), nitrogen (N^{+4}) and hydrogen (the Lyman line $L\beta$). Lines have been broadened by Doppler shifts of gas moving at velocities of a few hundred kilometers per second.

supernova remnant are able to accelerate electrons, protons and other subatomic particles to velocities close to the velocity of light. As these particles break away from the pulsar or the supernova remnant and interact with magnetic fields in the galaxy, they emit light and other forms of electromagnetic radiation. When they encounter atoms of the interstellar medium, they can generate gamma rays.

The spectrum of a primeval galaxy can now be obtained by adding together the spectra of the contributing phenomena. The greatest uncertainty in determining the spectrum comes in estimating the ratio of the number of massive stars that are shrouded by dust to the number that are not shrouded. We have assumed that the ratio is 1 : 1, which should suffice unless one situation strongly predominates. At red shifts of between 3 and 30 the calculated spectrum could be distinguished most easily by optical or near-infrared telescopes such as the space telescope. Other instruments that are still in the planning stage, such as orbiting far-infrared and X-ray observatories, should be able to detect distant primeval galaxies at other wavelengths.

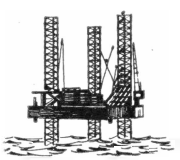
Are there any details in these spectra that would distinguish a primeval galaxy from other objects? In the visible and ultraviolet spectrum of a primeval galaxy one expects a continuous emission spectrum from stars, a line absorption spectrum from the atmosphere of stars and a line emission spectrum from the ionized gas that surrounds hot stars. The gas and stars in a primeval galaxy are moving rapidly and randomly, so that the lines will be much broader than those of slower-moving gas and stars that are not shifted from their normal position as much by the Doppler effect. The velocities associated with the line widths should never exceed the escape velocity of a primeval galaxy: a few hundred kilometers per second. An object with such features would be a strong candidate for being a primeval galaxy.

When the quasars were first discovered, it was thought they might be primeval galaxies chiefly because of their large red shifts. The spectrum of a quasar, however, is quite different from the calculated spectrum of a primeval galaxy. The continuous spectrum of 3C 273, one of the first identified quasars, slopes smoothly downward from the infrared to the ultraviolet, whereas a primeval galaxy should have one peak in the infrared region and one in the visible and ultraviolet region. Moreover, the continuous spectrum of a quasar is often partially polarized: the light waves oscillate in a preferred direction. Radiation from a primeval galaxy, on the other hand, is expected to exhibit little or no polarization. The spectrum of a quasar also has much greater line widths, which

You'd expect McDonnell Douglas to build the world's best fighter aircraft...

The F-15 Eagle. The peace keeper.
America's air ace. With extraordinary radar and electronics, more maneuverability, it is awesome enough to prevent attack, powerful enough to win. It's the best fighter aircraft in the world. And it's keeping watch over you.

but we also simulate the sea...



To help designers prepare ocean structures for a lifetime of hurricanes, rough seas, mud and air loads, and for the debilitation of fatigue. We use experience-based programs and one or more of our host of computers to help offshore technology companies build and work with greater confidence. OFF-SHORE—one of the hundreds of data capabilities of our fast-growing MCAUTO Division. To tell engineers all they need to know.

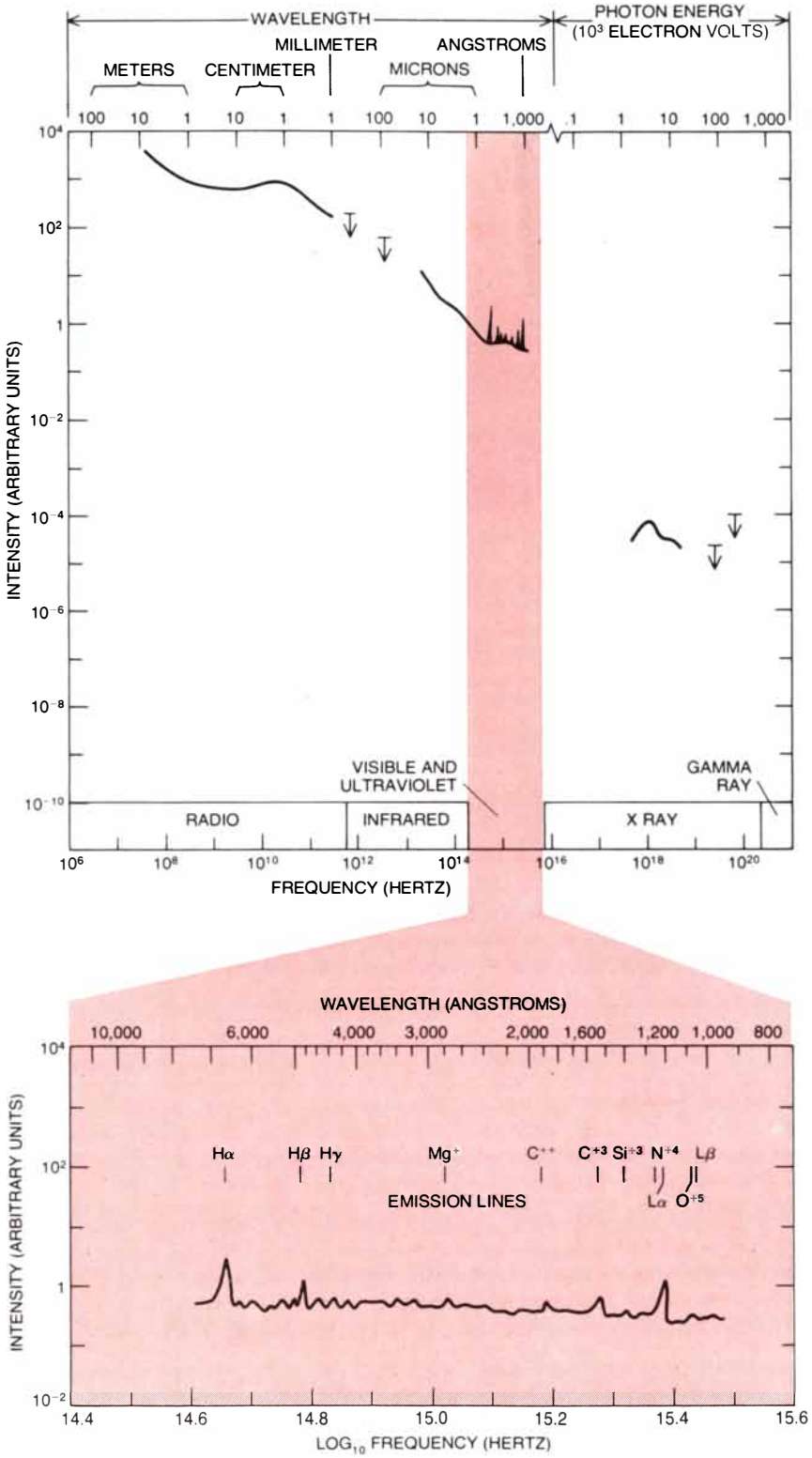
and make computers that talk.

Developed from our aircraft technology research, Mavis,™ is a computer-voice system we are making available to telephone companies so they can make more information available to you. With the Mavis system, you can dial a pre-determined number and Mavis computers can simultaneously tell you and thousands of others—sports scores, the latest stock quotes, ski-conditions, weather reports, just almost anything. For you, it's convenient; for phone companies, it's a bonanza.

At McDonnell Douglas, we've used our expertise in aircraft and aerospace technology to develop products that hopefully will make your life a little better, easier and even longer. To find out more about us and what we're doing, just write McDonnell Douglas, Box 14526, St. Louis MO 63178.

MCDONNELL DOUGLAS

We bring technology to life.



SPECTRUM OF A QUASAR (*top*) is very different from the expected spectrum of a primeval galaxy, so that the two objects could be distinguished even though they might look the same through a telescope. The quasar spectrum has neither the continuous emissions nor the line emissions that would be expected from the hot stars and gas in a primeval galaxy. The spectrum of a primeval galaxy peaks once in the infrared region and once in the ultraviolet region, whereas the spectrum of a quasar slopes smoothly downward. The lines in the visible and ultraviolet regions (*color*), which are shown in the expanded view at the bottom, are broader by a factor of about 100. This intrinsic spectrum is that of the quasar 3C 273, one of the first quasars to be discovered. Measured spectrum is shifted to the left by a factor of 1 plus the red shift .158.

suggests that such an object is a more energetic phenomenon than a primeval galaxy. The gas in quasars seems to be moving at speeds on the order of 10,000 kilometers per second. If clouds in a primeval galaxy traveled at such speeds, they would have escaped from the gravitational field of the galaxy long ago unless they were involved in a recent violent explosion or were bound to a large compact mass, such as a supermassive star or a black hole.

Quasars can also be distinguished from galaxies because their light can vary in brightness in the course of only a few months or years. Light from BL Lacertae objects, quasarlike entities with spectral peculiarities of their own, vary in intensity in only a few hours. Such variations are impossible for a collection of 100 billion stars, which cannot vary in unison. Even exploding supernovas, as we indicated above, can scarcely alter the brightness of a primeval galaxy, and they certainly cannot cause any variation in a matter of hours. Variable quasars must be extremely compact. They are probably smaller than the distance light can travel in the time it takes for the intensity of their light to change.

If quasars are not primeval galaxies, what are they? To answer this question astronomers have developed models of the region that is thought to emit the lines in the quasar spectrum. In a typical model the quasar is enveloped in a low-density gas that extends a few hundred light-years and has spectral lines with a width corresponding to a velocity of 1,000 kilometers per second. Within this region is a compact nest of dense, rapidly moving clouds of gas about a light-year in extent. The nest of clouds is the source of the broader lines in the spectrum of the quasar. At the center of the quasar is the "engine," at most a light-day in size, that emits the continuous spectrum and ionizes the denser clouds.

The exact nature of the quasar engine is the subject of current work. Three proposed models have been offered so far. In the first model the engine is a compact cluster of stars, the energy coming from a multitude of stellar collisions, supernova explosions and pulsars that are squeezed into a region with a radius perhaps only 10 times larger than the radius of the solar system. In the second model the engine is a supermassive star or pulsar, the energy coming from the gravitational contraction or rotation of a single magnetic star with a mass as great as a billion solar masses. In the third model, which is similar to the current model of binary X-ray sources such as Cygnus X-1, the engine is a supermassive black hole of a billion solar masses. The energy comes not from the black hole itself but from gas and decomposed stars that collide as

Is your work at the cutting edge of imagination? Then Rolex wants to help you change the world.

The five Rolex Awards for Enterprise, in the sum of 50,000 Swiss francs each, are conferred to provide financial help for projects which seek to break new ground in their particular sphere, and which capture the spirit of enterprise shown by Rolex and Rolex owners.

The kind of enterprise that takes a man like Julian Nott to world record-breaking altitudes in hot-air balloons, Sir Edmund Hillary to the top of Everest. Or Dr. Bill Lasley, a previous Awards for Enterprise winner, into new methods for saving and breeding rare species of birds.

These people epitomize the spirit of enterprise. The 1981 Rolex Awards for Enterprise seek others like them whose proposals for projects fall into any of three general categories:

Applied Sciences and Invention. Projects involving a step forward in research, experimentation, or application in the general field of science and technology.

Exploration and Discovery. Projects involving exploration or discovery will be concerned primarily with venturesome undertakings or expeditions, and should seek to inspire our imagination, or expand our knowledge of the world in which we live.

The Environment. Projects in this category will be concerned primarily with our environment, and should seek to protect and preserve, or to improve, the world around us.

In addition to the 50,000 Swiss francs, each of the five winners will receive a specially inscribed Rolex Chronometer.

For an Official Application Form with detailed rules and conditions write to:

The Secretariat, The Rolex Awards for Enterprise, P. O. Box 178, Dept. 316, 1211 Geneva 26, Switzerland.

Send your name and address and state the general nature of the project you are considering.

Entries must then be submitted in English on the Official Application Form and reach the Secretariat before April 30th, 1980.

The Rolex Awards for Enterprise selection process is exacting and arduous. Winners will be notified in December 1980, and their names will be published in the international editions of *Time* and *Newsweek* magazines in May 1981.

The Rolex Oyster Perpetual Day-Date Chronometer

ROLEX

they fall into it. In the collisions the gas and stars release about a tenth of their rest-mass energy (their mass at rest multiplied by the square of the speed of light). A loss of only 10 solar masses per year to the black hole would provide enough energy to power the brightest quasar and to strongly outshine a galaxy of a trillion stars.

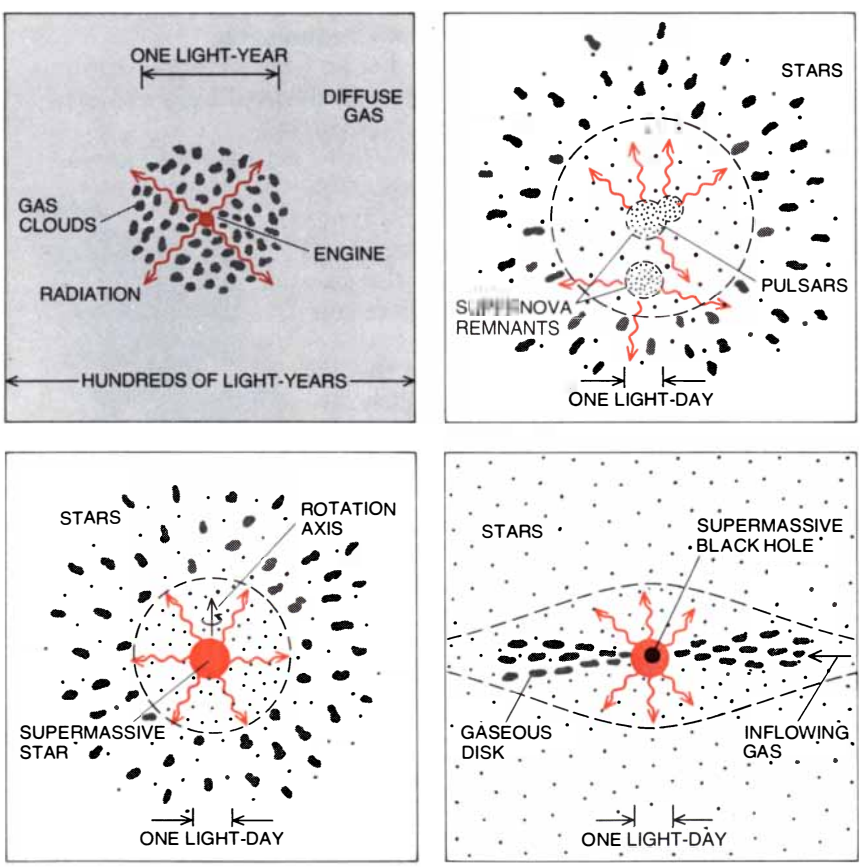
What is the origin of the quasars? The models of the quasar engine suggest that they exist at the nuclei of otherwise normal galaxies, where compact star clusters, supermassive stars and large black holes are most likely to form. The BL Lacertae objects, the extended radio sources that resemble extended radio quasars and the Seyfert nuclei, all of which are less luminous than quasars but have virtually the same spectrum, have been found in galaxies. It is reasonable to assume that quasars are simply scale-up versions of these phenomena, so bright and distant that the surrounding galaxy cannot be seen.

How do quasars form? This question remains largely unanswered. The abundance of quasars at red shifts just below the ones we expect primeval galaxies to have suggests that a quasar forms in a galactic nucleus soon after the galaxy forms. In fact, quasar formation may be a phase through which most galaxies pass when they are born. In the model of a collapsing galaxy outlined above activity need not stop when star formation stops. There may be events that cause the young galaxy to develop at its center the kind of supermassive object called for by the models of the quasar engine. (Processes that have been suggested are the collision and coalescence of stars packed into the galactic nucleus or the merging and collapse of gas clouds left over from star formation. Another possibility is that a large black hole is built up slowly from a small one that swallows up neighboring stars.) Then, by some poorly understood mechanism, the supermassive object briefly shines much brighter than the galaxy ever did. Finally the quasar dies, revealing the surrounding galaxy. At present this model is little more than speculation. Over the next few years astronomers hope to be able to uncover more evidence about how quasar formation and galaxy formation are linked.

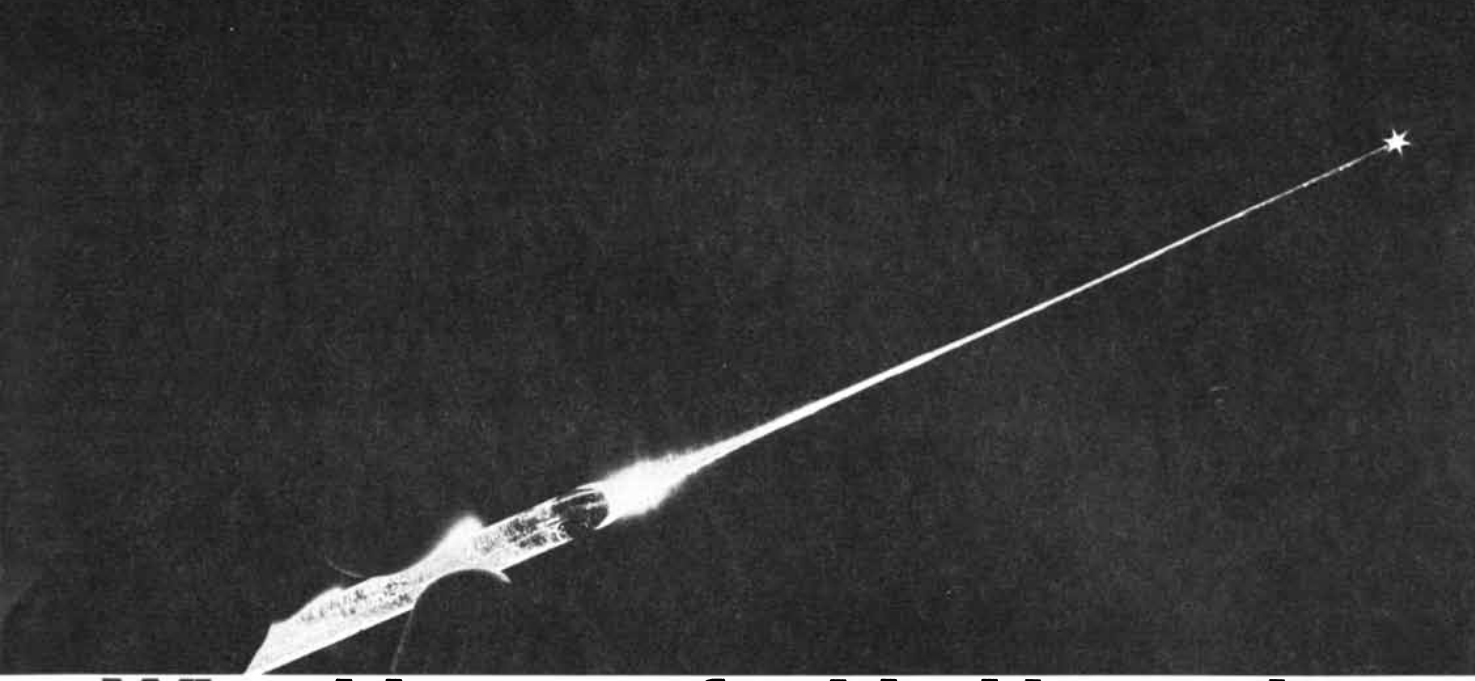
All the evidence suggests that quasars are not primeval galaxies. Nevertheless, some primeval galaxies may have been mistaken for quasars. Before combing the skies for new primeval-galaxy candidates it will first be necessary to examine the lists of quasars already catalogued for those that have emission lines with widths corresponding to velocities of only a few hundred kilometers per second. Such an examination could be a herculean project, because hundreds of quasars have been catalogued. Moreover, the spectra of many of these quasars have not been recorded in sufficient detail. The examination would be manageable, however, if the quasars with the highest red shifts were studied first. The next step will be to obtain new candidates with the objective-prism technique. This technique records rough spectra of many objects in the sky at the same time, allowing estimates of the red shift, line widths and other characteristics to be made quickly. It is already useful in finding quasars. An objective prism employed with the space telescope will perhaps be the best bet for detecting a primeval galaxy. Through the telescope very compact objects will remain points, whereas primeval galaxies should show some detail. That may eliminate the need to make detailed spectra to distinguish between primeval galaxies and quasars.

The discovery of a primeval galaxy would constitute a major triumph for modern astronomy. Current theories of cosmology, galaxy formation and stellar evolution would have new evidence that argued strongly in their favor. In a field as complex as astronomy, however, theoretical predictions are not always accurate; it is quite possible that unforeseeable phenomena could complicate the search for primeval galaxies. For example, it could turn out that not all galaxies formed in a violent collapse of an expanse of gas and dust with a mass of about 10^{11} solar masses. Perhaps some galaxies formed from the merging of much smaller star clusters that were themselves created from the numerous low-mass perturbations (ranging from 10^6 to 10^{11} solar masses) existing in the early universe. A primeval galaxy formed by merging would presumably be fainter than one formed by the collapsing process we have outlined, since the formation of stars would be slower.

The search could be complicated further if dust was so abundant in young galaxies that it shrouded most of the massive stars. At worst the entire galaxy could be hidden from view. It will be several years before instruments are introduced in space that could even detect the infrared radiation emitted by such dust, and it will be many years before the red shift of a dust-covered galaxy can be measured reliably. Most astronomers nonetheless believe primeval galaxies will someday be found. It will then be possible to study the formation of galaxies directly.



"ENGINE" OF A QUASAR is probably in the center of a nest of dense clouds embedded in diffuse gas (*top left*). Three theoretical models of the engine have been proposed. In the first model (*top right*) the engine is a compact cluster of stars. In the second model (*bottom left*) it is a supermassive star or a pulsar. In the third model (*bottom right*) it is a supermassive black hole.



Why this one-of-a-kind invention didn't end up as the only one of its kind.

Every new invention needs another new invention—the one that can mass-produce it at an affordable cost.

For example, Bell Labs invented a process for making the glass rods from which hair-thin fibers used in lightwave communications can be drawn. The fibers have far greater capacity than conventional copper wires, so they'll help keep costs down. In fact, they've been carrying voice, data, and video signals under city streets for about two years in a Bell System demonstration.

But standard lightwave systems will require *miles* of the fiber, produced at low cost and to specifications nothing short of microscopic.

That's where Western Electric's Engineering Research Center comes in.

A Unique Center

The Center is devoted exclusively to manufacturing research.

Here, a highly trained team of scientists and engineers probe fundamental questions about materials and processes. They provide Western Electric factories with pre-tested, proven ways to manufacture products based on the latest technology coming out of the laboratory.

For example, while Bell Labs scientists were inventing new glass fibers, Western Electric engineers and scientists were tackling the manufacturing problems involved.

The fibers had to be drawn from molten glass at high speeds, with less than a 1% deviation in diameter.

But how do you control a "thread" of glass being spun at rates up to 15 feet per second?

Scientists and engineers at the Center discovered that laser light beamed onto the fiber cast a characteristic pattern. By correlating the pattern to the fiber's diameter, they were able to build a monitoring system into the fiber drawing machinery. It measures the fiber 1000 times per second, automatically adjusting production to keep the diameter constant.

The system works so well that in all the miles of fiber produced by Western Electric, the diameter varies by no more than 30-millionths of an inch.

The Key to the Future

In the Bell System, technology is the key to keeping costs down. It is the key to constantly improving your phone service.

And Western Electric's Engineering Research Center is an essential link between the ideas of the laboratory and the realities of the factory.

So your Bell Telephone Company can make the best one-of-a-kind inventions a part of your phone service.

Keeping your communications system the best in the world.

Western Electric

The Ecology of the African Dung Beetle

Scarab beetles play a key ecological role in removing the dung left by herds of mammals. Success in the vigorous competition among such beetles depends in part on their body temperature

by Bernd Heinrich and George A. Bartholomew

It might cross the mind of anyone who contemplates the great herds of animals in the grasslands and savannas of East Africa that the animals produce large quantities of dung. Yet one does not see much dung on the ground. What happens to it? Most of it is rapidly removed and buried by legions of scarab beetles, for which it is a vital resource. The beetles do a good deal more than clean up material that would otherwise accumulate on the ground, choking out plants and so probably limiting the populations of animals the land could support; their activity also fertilizes and aerates the soil, retards the spread of parasites and disease organisms and reduces the number of bothersome flies that breed in dung. The story of the beetles is illustrative of the intricate relations in an ecosystem. It also reveals some remarkable physiological and behavioral adaptations of the beetles.

The dung beetles, of which Africa alone has more than 2,000 species, have evolved into many sizes and shapes. The largest weigh more than 20 grams and are seven or eight times larger than the smallest birds, bats and shrews. The smallest are hardly a thousandth as large, weighing only a few milligrams. In appropriate circumstances the beetles are so abundant that dung, even though the supply is constantly being replenished, is a scarce resource. Indeed, its availability can limit the beetles in both reproduction and growth. In tropical Africa many species compete strongly for dung, and a remarkable variety of patterns of utilization have evolved.

We went to study scarab beetles in Tsavo National Park in Kenya, which has the largest elephant population in the world. A group of four or five elephants can process a metric ton of food per day, much of which is returned to the ground as dung. In the rainy season at Tsavo Park the removal of elephant dung by beetles is amazingly rapid. During the day relatively few beetles are attracted to fresh elephant dung, but after sundown the flying beetles arrive in great humming clouds. Small samples of elephant dung that we collected during the day and put out as bait at night attracted swarms of beetles. More than 3,800 of them came to a half-liter sample left exposed for only 15 minutes. Only a few beetles appeared during the early part of the period, but then the rush was on.

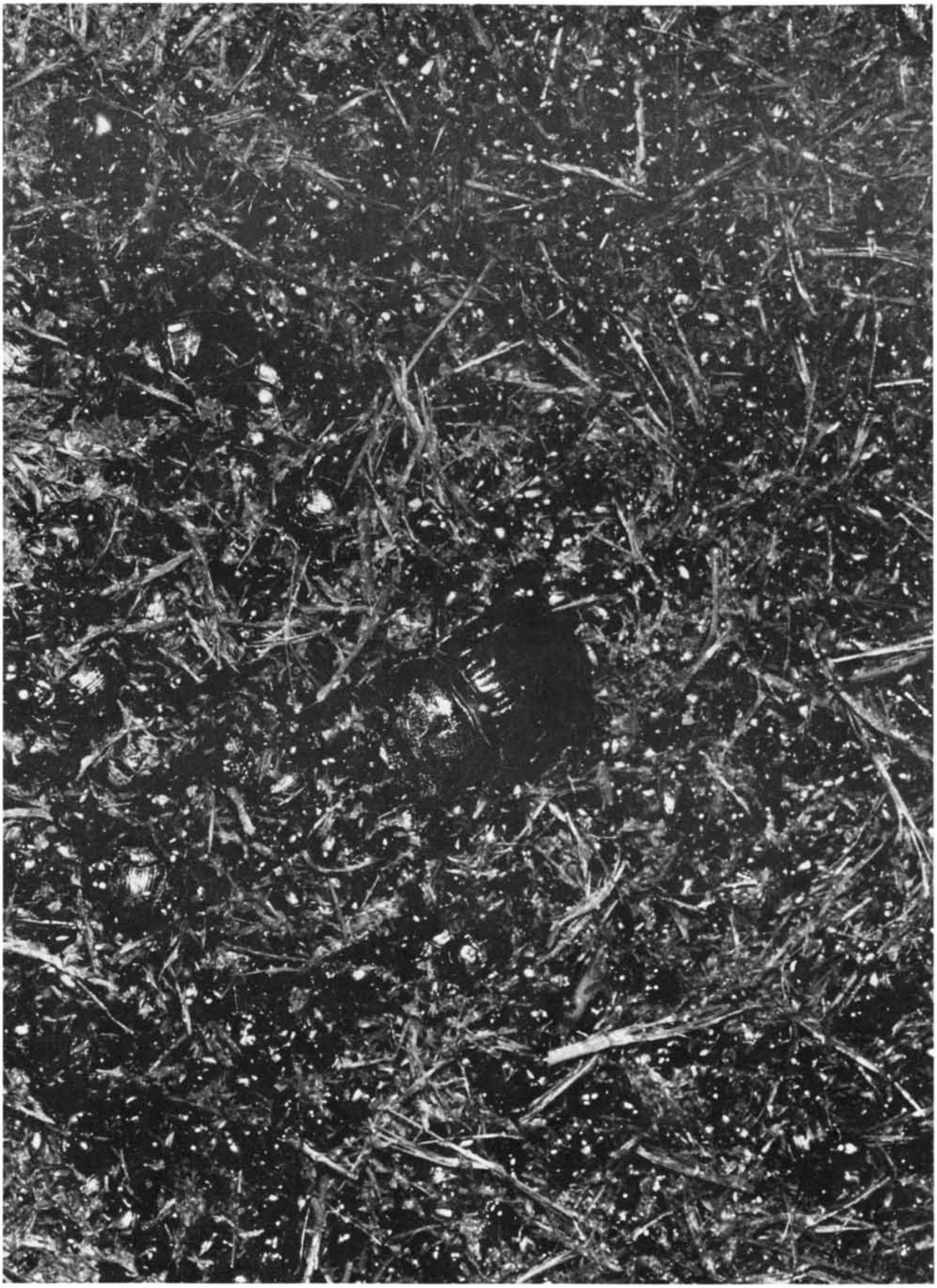
Another sample of about 30 liters of elephant dung brought an awe-inspiring onslaught. Clouds of small beetles swept in, landed on the dung and immediately tunneled into it. Within half an hour the pile of dung was transformed into a spreading, heaving mat consisting of a fluidized layer of beetles and moist dung covered by a thin sheet of fibrous material: the coarse, undigested remains of plants. The beetles transformed the firm, malodorous, football-size boluses of dung into a spreading mat two or three centimeters thick and as much as two meters in diameter.

It does not take an observer long to see that competition for elephant dung within and among the many species attracted to it is intense. What are the consequences? How have dung beetles accommodated to it? We shall deal with these questions by describing some of the patterns exhibited by scarab beetles in harvesting and utilizing dung and then by examining some of the behavioral and physiological responses that operate within these patterns.

A few species of scarabs specialize in the dung of particular kinds of mammals, but many species will accept any dung they find. Three major patterns of dung utilization can be discerned. Some beetles, which are termed endocoprids, burrow into dung and remain there living and feeding until the dung is exhausted or its structure breaks down. Other species tunnel into the earth below or next to a pile of dung and haul the dung into the burrow. Still others, such as the Egyptian scarab *Scarabaeus sacer*, cut out pieces of dung, shape them into balls and roll them from one to 15 meters or even more before burying them.

The beetles employing each of the three techniques have morphological adaptations that complement their behavior. The endocoprids are generally small, a characteristic suited to feeding inside dung that contains much fibrous material, as elephant and rhinoceros dung does. The digging beetles are often large and robust, with bodies that function as earth-moving machines. The most striking examples of this adaptation are the huge beetles of the species *Heliocopris dilloni*, which feed only on elephant dung. Some of these beetles weigh as much as 25 grams. They have powerful legs, with shovel-like front tibias and backward-directed spines on the tibias of their two hind pairs of legs that enable them to get good traction while they push headfirst through the soil. The leading and upper surfaces of the head resemble the blade of a bulldozer. The dung-rolling beetles, in contrast, have long, thin, outwardly bowed hind legs that are suited for running and that they can wrap around a mass of dung, thereby clinging to it as they form it into a ball. Their front tibias have rakelike extensions that are utilized to pat bits of dung onto the ball as it is being formed. Dung rollers have evolved into many sizes. Some weigh only a few milligrams and roll pea-size balls. Others, such as the species *Scarabaeus laevistriatus* and *Kheper platynotus*, weigh as much as 10 grams and cut from elephant dung spheres the size of tennis balls.

Burying dung has the advantage of getting it out of the reach of competitors. It is often imperative that burying beetles work fast. The species that bury dung below the pile (without any rolling) usually form sexual pairs and work as a team. They take dung into the burrow in loose bundles, ultimately forming some of it into round or pear-shaped brood balls. One egg is laid in each ball, which provides all the food and water a



SCARAB BEETLES in East Africa work at removing a pile of elephant dung during the night. The bright spots in the photograph are reflections from the backs of beetles, of which more than 200 can be counted. Visible beetles, taken together, weigh more than the dung.

single larva needs to complete its development. Only the females make brood balls. The females also do most of the work associated with nesting.

Generally the female makes the initial burrow, pushing the soil she loosens up to the surface and leaving it there. The male may then push it away from the entrance. As soon as the female stops bringing up soil the male seizes a piece of dung with his front legs and backs down the burrow. When the tip of his abdomen strikes the female's head, he drops his load and returns to the surface for more. The female takes the piece of dung and carries it the rest of the way down the tunnel. In this pattern of behavior the male works on or near the surface and so is exposed to predators, whereas the female, which contributes the greater share of the reproductive effort, works in the relative safety of a burrow that may extend a meter or more below the surface.

Both males and females of the ball-rolling species make dung balls. They eat some of the balls, but others (made by the male) are key elements in the mating sequence. The male and the female first encounter each other at the dung dropping, which has attracted them both. A female may join forces with a male that is making a ball or has already made one. The details vary among species, but the general pattern of reproductive behavior is the same.

Kheper platynotus, a large beetle that is active during the day, provides an example. The female usually joins the male at the dung pile while he is making a ball. When he starts to roll the completed ball away, the female clings to it and rides along while the male laboriously pushes it to the burial spot. The male buries the ball by excavating the dirt under it so that it sinks vertically into the earth. The female, still aboard, sinks with it. When the ball has been buried, the pair feed on it, mate and return to the surface.

In some species the female does not ride on the ball. For example, the female of the species *Gymnopleurus laevicollis* stands on the ground on her hind legs, putting her front legs on the forward side of the ball. She therefore appears to be pulling the ball while the male, on the other side, stands on his front legs and pushes with his hind legs in the usual scarab fashion. The female of the species *Scarabaeus sacer* trails along two or three centimeters behind the male as he rolls the ball to the burial site.

Among all scarabs the male does the work of burying the ball and the female nibbles on it as they mate. Because a ball serves the male in a sense as an offering to the female that enables him to mate, it is called a nuptial ball. (Nuptial feeding is found in other insects and in a number of vertebrates.) Plainly success in competing for food is tied closely to reproductive success.

BALL IS SHAPED from a bolus of elephant dung by a ball-rolling dung beetle of the species *Kheper platynotus,* which is active during the day. (Most East African dung beetles work at night.) Two finished balls, with a beetle climbing on one of them, are visible at the left. *K. platynotus* is the species portrayed on the cover of this issue. The male pushes the finished ball, with the female riding on it, over a distance of several meters until a suitable burial place is found. The East African region's lateritic soil gets its red color from iron oxides.

RESIDUAL MAT was all that was left of a bolus of elephant dung when hordes of beetles had worked on it for a single night after it was dropped. Most of the beetles that were there during the night had left by the time the photograph was made, but the ones remaining under the mat numbered in the thousands. Most of them would be gone within another day. One of the authors (Bartholomew) is at the left.

Mumm's the word.

For 150 years, people who know how to live have been celebrating life with Mumm premium French Champagne.

Although the male dung beetle is in no danger of being eaten by the female after mating, which is what happens among some insects such as the mantids, he does incur considerable risk in making a nuptial ball. Dung piles teeming with beetles attract birds (hornbills, spur fowl, guinea fowl) and mammals (the mongoose) that prey on the beetles. A male making a dung ball for a female to eat in relative safety underground exposes himself to risk while reducing the risk to a potential mate.

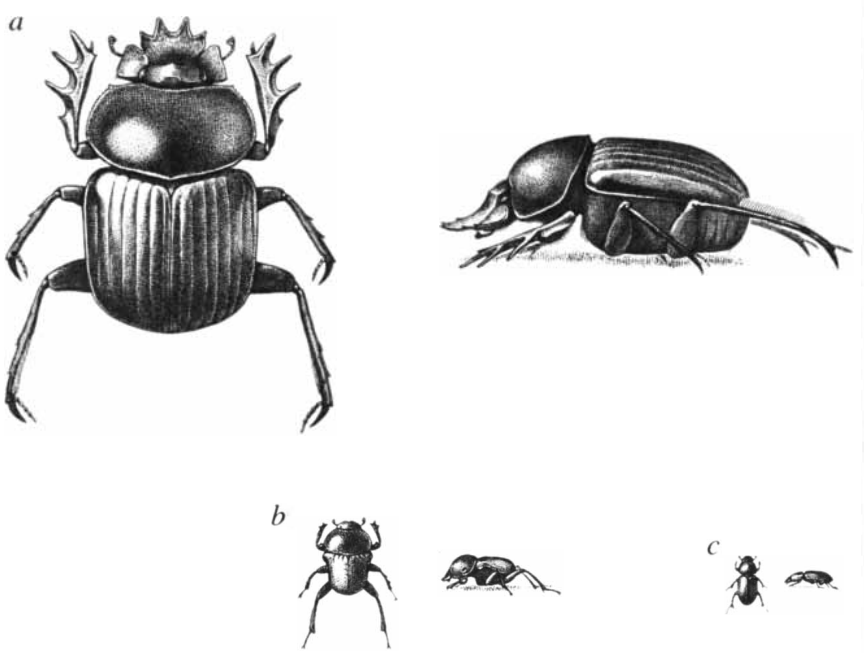
Among the East African scarabs the selective pressure to speed up the process of finding dung and removing it from the dung pile is intense. The pressure arises not only because the dung attracts predators on beetles but also because both the physical environment and the competition among the beetles make it imperative for a beetle to get in and out fast. One therefore wonders what behavioral and physiological mechanisms have evolved that speed up the acquisition of dung.

Dung beetles cannot ordinarily employ the obvious strategy of following herds of mammals because the mammals are active both day and night, whereas the beetles (depending on the species) are either diurnal or nocturnal. Moreover, beetles that have buried a ball stay with it for at least a day. By the end of that time the mammals that deposited the dung are likely to be far away.

It is apparent that the beetles find fresh dung by its odor: they always approach from downwind. We learned that enormous numbers of beetles could find dung bait that had been hidden, even when it was in an area without game animals.

It is not known whether beetles cruise continuously in search of dung or wait to take off until they catch the scent. It is known, however, that the control of body temperature plays an important role in the speed, the energy cost and the potential reward of finding fresh dung. The body temperature of a scarab beetle at rest is within one degree Celsius of the ambient temperature. It is known that some large insects during flight have a thoracic temperature as much as 35 degrees C. above the temperature of the air. In any one kind of insect the power of the wing beats and the speed of flight are closely correlated with the temperature of the muscles. If the temperature of the flight muscles in the thorax is lower than about 34 degrees, many large insects are unable to generate enough lift and thrust to remain airborne. What is the situation in dung beetles of various sizes?

We measured the thoracic temperature of beetles in flight by capturing them at night as they arrived at dung or at lights and immediately inserting into their thoracic muscles a fine hypodermic needle bearing a tiny thermocouple. The thoracic temperature of the large beetles was from 39 to 45 degrees, which is from two to eight degrees higher than the temperature of most mammals and as high as that of flying birds. Some of the beetles must have been flying with



THREE SPECIES of East African dung beetle are shown in top and side views. The scale is the same. The larger species (not the largest of African dung beetles) is *Scarabaeus laevistriatus*, a ball-rolling species active at dusk and during the night. At the left below it is a smaller nocturnal ball roller, *Gymnopleurus laevicollis*. The third beetle is an endocoprid: it burrows into dung and feeds on it until the resources are exhausted. Endocoprids are ordinarily small.

muscle temperatures only a degree or two below the level that is damaging and possibly lethal.

These high thoracic temperatures are necessary for flight in the larger dung beetles. They cannot get airborne with low muscle temperatures. The smaller scarabs, however, can fly well with a body temperature as low as 27 degrees, which is only a couple of degrees above the nighttime ambient temperature in Tsavo Park while we were working there.

What accounts for this difference? The answer involves at least in part the relation between size and the rate of heat flux. From our extensive measurements of thoracic temperature as a function of size we conclude that in beetles weighing up to two grams body temperature is a passive function of the production of heat (a necessary consequence of flight metabolism) and the rate of passive cooling. The larger the beetle, the smaller the relative amount of surface area for dissipating heat and the higher the body temperature. There is, however, an upper limit of tolerable temperature: about 45 degrees, which is in fact approached by beetles weighing only two or three grams.

Hence the larger beetles must actively dissipate some of the heat they generate in the intense metabolism associated with flight. They can do so only when the temperature gradient between their body and the air is large. Since they necessarily heat up, it can be inferred that they have evolved the biochemical machinery of the flight motor to function optimally at the temperature it develops during flight. The small beetles do not heat up, and so they have been under selective pressure to fly at lower muscle temperatures.

When enzymes and other components of the biochemical machinery have been committed to perform at maximal rates at one temperature, their function is usually impaired at another. This tradeoff may have significant implications in the strategies of competing for dung. The small beetles can take off immediately when they smell dung. The large beetles, however, must expend time and energy in warming up before they can fly. (We ascertained that a beetle of the species *Heliocopris dilloni* weighing 11.7 grams needed five minutes to warm up from 27 degrees to a flight temperature of 40 degrees.)

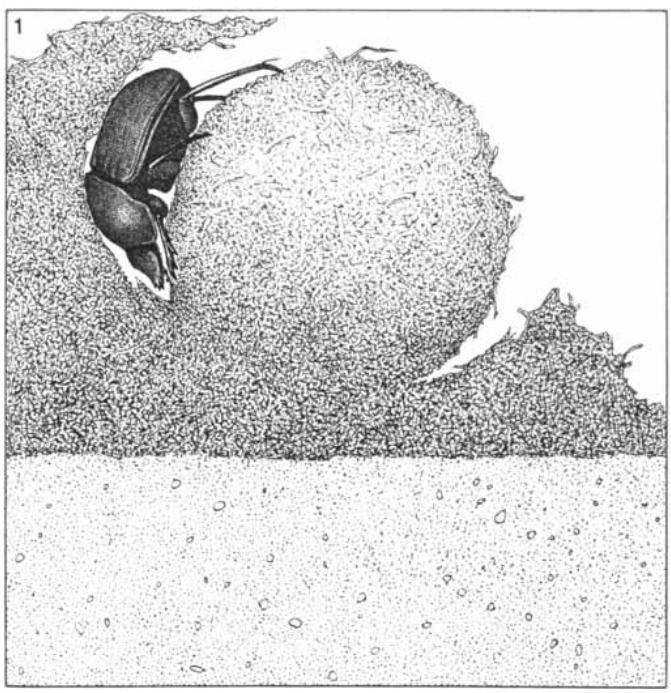
The temperature of the flight muscles has other implications. The flight muscles in the thorax also heat up other thoracic muscles that move the legs. Hence flight-muscle activity affects essentially all the activity of a beetle such as a ball roller: walking to a dung pile after landing some distance away from it, making a ball, rolling it away and defending it against competitors. Each of these activities is related to the rapidity of motion and power of movement of the legs. It is therefore possible that the flight temperatures, which are evolutionarily related to body size, set the rate for all other activities unless physiological mechanisms are activated to circumvent them.

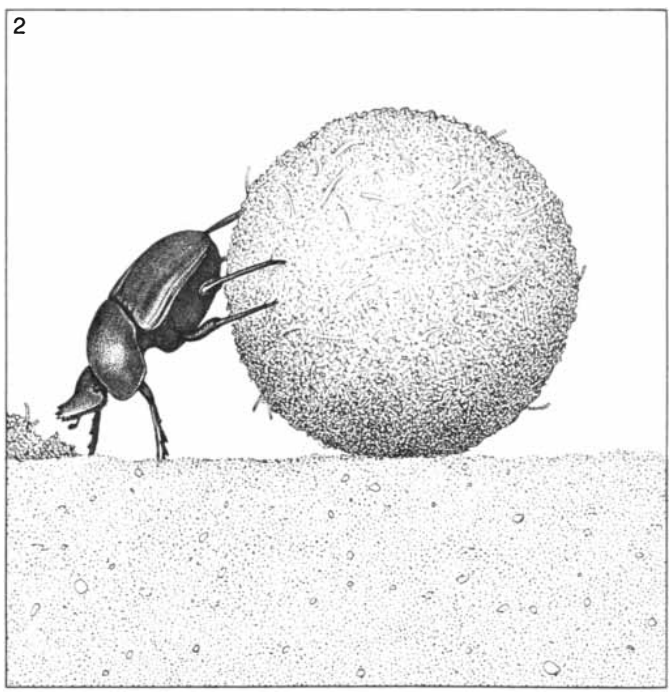
A physiological mechanism the large beetles have, which keeps them from being prisoners of the thermal environment, is shivering. Shivering involves contractions of the flight muscles. It enables the beetles not only to warm up for flight but also to keep warm after they land. Although shivering is essential for flight, it seems to be an option for other activities. The ball-rolling beetles can walk normally and do other things on the ground with a body temperature at about the level of the ambient temperature. Beetles with low body temperatures simply move slowly.

We found that the diurnal species *Scarabaeus catenatus,* which moves rather slowly anyway, had a thoracic temperature of 41 degrees in flight, 28.4 degrees while making balls in the shade, 32 degrees while rolling balls in the shade and 37 degrees while rolling balls in the sun. The female of the species *Kheper platynotus* that rides on the ball while the male pushes it consistently has a lower thoracic temperature than the male.

When a beetle is only walking, its thoracic temperature does not rise significantly. We therefore conclude that the animals sometimes shiver during certain activities on the ground in order to speed up their rate of work. One should remember, however, that the amount of competition among dung beetles active during the day is low, so that at least this one selective pressure for speeding up the rate of activity is reduced.

At night, when the endocoprids and the diggers make their great rush to dung, the ball rollers have only minutes in which to secure dung from fresh piles or from piles that have accumulated during the day. We concentrated our





DUNG-BALL ROUTINE of a female of the species *Kheper aegyptiorum* **is depicted. She first cuts the material for a ball out of a bolus of dung and then shapes it (*1*). When the ball is finished, she starts to roll it to a burial place (*2*), standing on her front legs and pushing**

attention on one ball-rolling species, *Scarabaeus laevistriatus,* which is active at night. These beetles behaved very differently from the ball-rolling species that are active during the day.

The *S. laevistriatus* beetles, which are large and long-legged, seldom walk in a leisurely way but rather move at an almost frenzied pace. Like other large East African dung beetles they are strong and swift fliers, reaching speeds as high as 30 kilometers per hour. They almost certainly find dung by smell. We do not know whether they search for it by flying crosswind, by tacking back and forth upwind or by waiting on the ground until the odor reaches them. In any event they always (in our experience) approach dung from downwind. They sometimes land on it directly, but more often they alight several meters downwind and scurry along the ground to the pile. Once they are on the dung they clamber over it rapidly, repeatedly manipulating it with their front legs and probing it with their head, apparently assaying it for the conditions of moisture and cohesiveness that make it suitable for being formed into balls.

Thereafter their behavior depends on the extent of activity by endocoprid beetles and the presence or absence of other *S. laevistriatus* beetles. If the dung has been extensively leavened by endocoprids, the *S. laevistriatus* beetles usually fly off within a minute or two, presumably to look for something more suitable. Otherwise they start to make a ball. They find a protuberance on the pile, round it up, cut it off and roll it away. Some of them burrow into the dung, make a ball there, push it to the surface and roll it away.

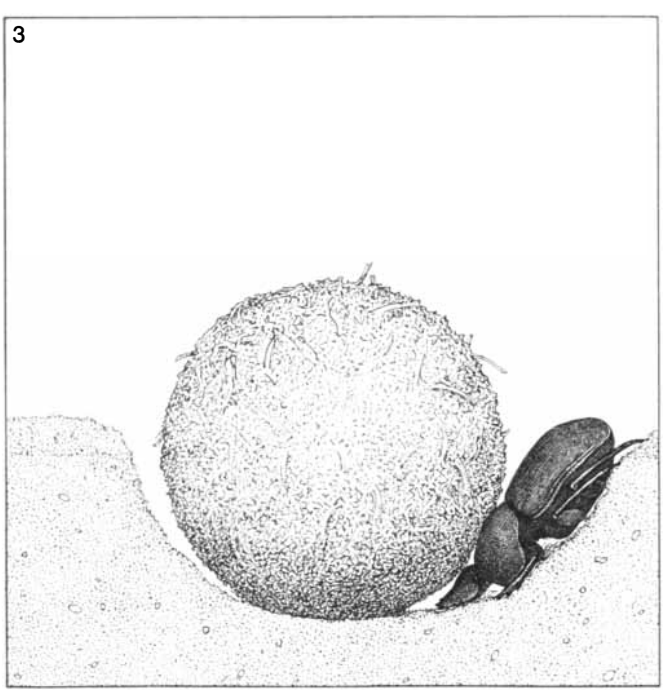
The time it took the beetles we watched to make a ball ranged from 1.1 to 53 minutes. The longer times were observed only when we put wire screens over dung piles to exclude endocoprids and other competitors. Endocoprids can render fresh dung unsuitable for ball making in 15 minutes.

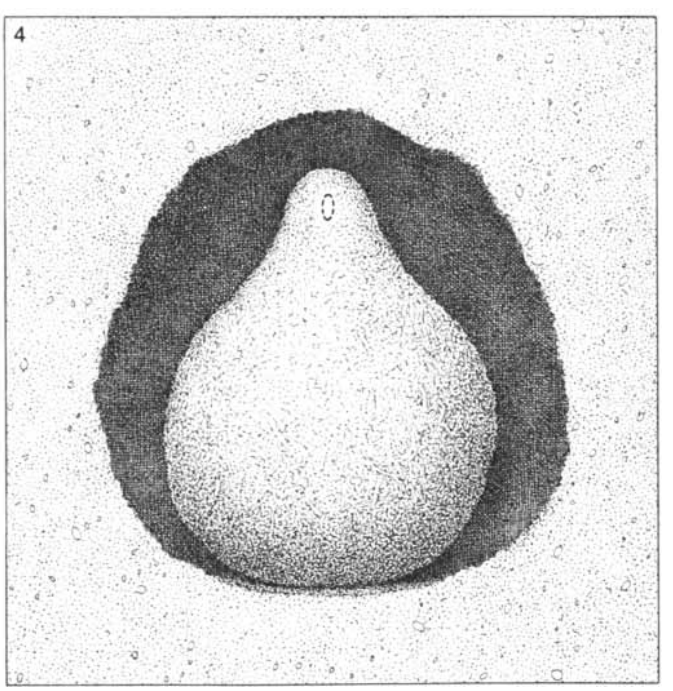
The ball-making beetle *S. laevistriatus* uses its front legs intensively to pat additional pieces of dung onto a ball and shape it. The rapidity of the patting movements is directly related to the thoracic temperature. The relatively cool diurnal beetles worked quite slowly, but the patting movements of the *S. laevistriatus* beetles, which were usually warm, were extremely rapid.

The relation between body temperature and speed of ball construction is affected by many variables, including the consistency of the dung, the size of the ball being made, the density of the ball, the presence or absence of other *S. laevistriatus* beetles attempting to steal the ball and the presence or absence of endocoprids that break up the ball as it is being formed. If a number of endocoprids invade a ball while it is under construction, the *S. laevistriatus* beetles abandon it. (One 33-gram ball that was abandoned before it had been finished had more than 50 small endocoprid beetles in it.) Notwithstanding such variables the hottest beetles were usually the first to leave with completed balls. An exception is the beetles working within the dung pile; they appear to sacrifice speed of getaway in order to make a dense, more cohesive ball.

Even after 20 minutes of ball construction in piles of dung that had cooled to the temperature of the air some beetles had thoracic temperatures near 40 degrees. On the basis of the rate at which dead animals cool we estimate that if these beetles had not generated heat, they would have cooled to within two or three degrees of the air or dung temperature in about 10 minutes. On the other hand, some beetles of the same species had low thoracic temperatures as they made balls. We do not know why some beetles shiver and stay warm and others cool off, but we suspect that (as with certain other animals) the difference may be related to the amount of stored energy available. The beetles that cool off probably have depleted their reserves of energy; the others have not. The energy cost of maintaining an elevated body temperature is high but the benefits are significant. A warm beetle is likelier than a cold one to leave with a ball of dung.

We also found that the success of a given *S. laevistriatus* beetle in making and burying a dung ball depends in large part on its fighting ability, which we found to be closely related to thoracic temperature. The best strategy for obtaining a ball quickly is to steal it. A new arrival at a dung pile often makes such an attempt. The ball roller maneuvers quickly to place itself between the attacker and the ball, meanwhile continuing to roll the ball away. If the attacker manages to climb on the ball anyway, a spirited contest develops. Both beetles



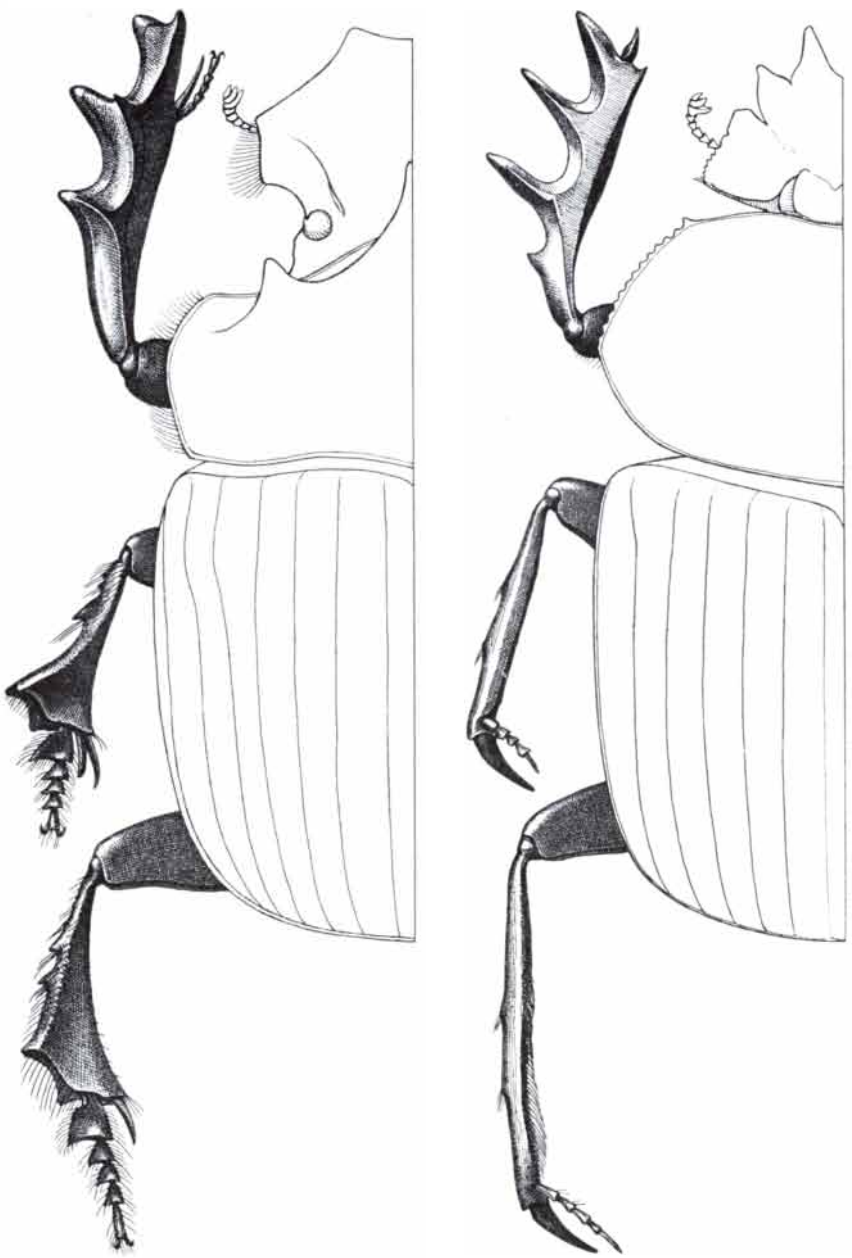


with her hind legs in the usual manner of African dung beetles. Once a suitable spot has been found she burrows (*3*) as much as half a meter into the ground to bury the ball. After the ball has been buried (*4*) the beetle lays a single egg in it. The ball serves as food for larva.

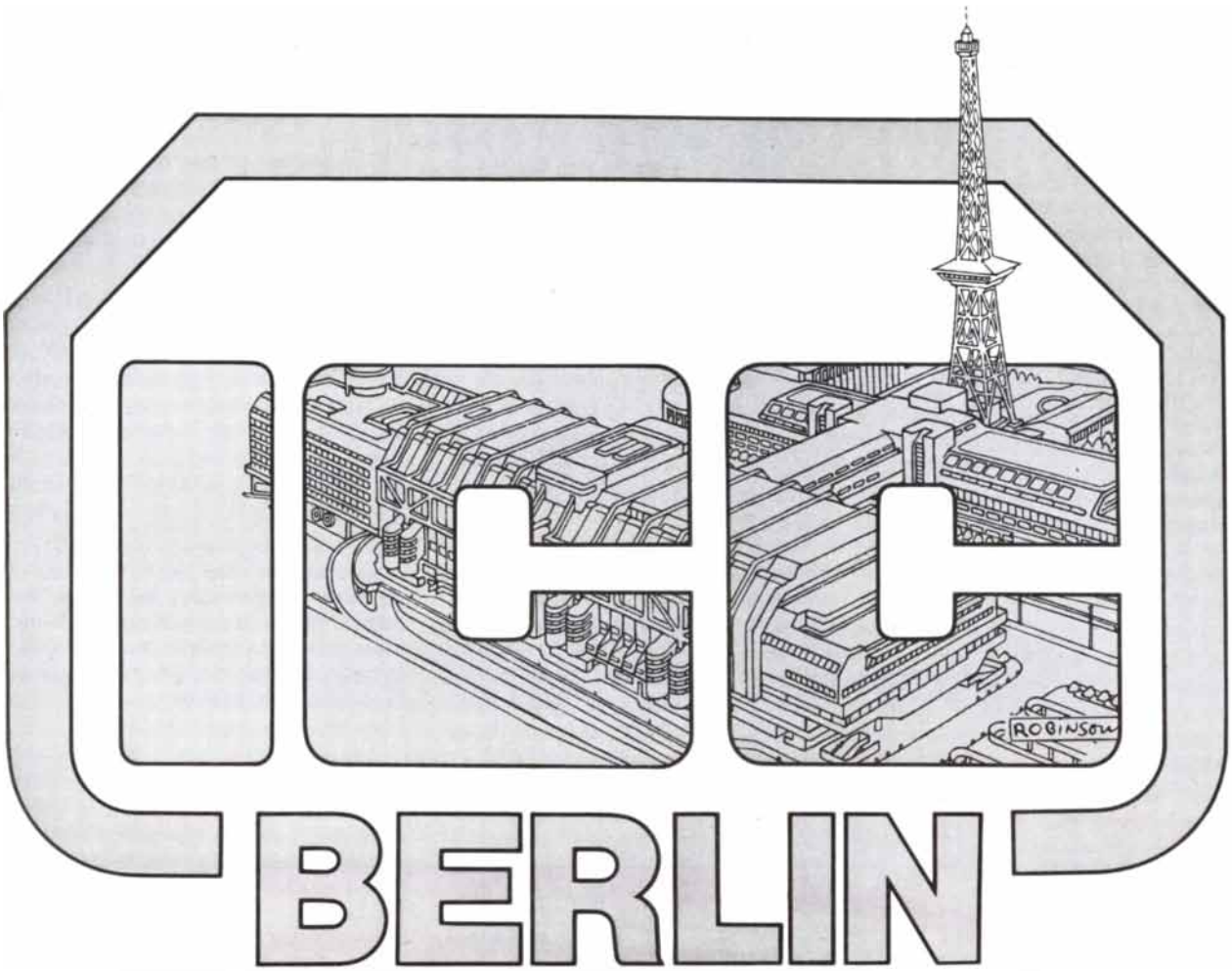
grasp the ball firmly with their middle and hind legs and try to dislodge each other by forward and upward flicking motions of their powerful front legs. Sometimes they grapple with their middle and hind legs while kicking or pushing with their front legs. The loser is often flicked 10 centimeters or more through the air, usually landing on its back. We never saw a contestant injured, however, in the hundreds of natural and staged fights we watched.

To assay the effects of temperature on this kind of aggression we staged contests with balls made by the beetles. The balls lasted for only a few minutes, however, because endocoprids entered and quickly destroyed them and because the vigor of the fighting soon tore the balls apart. We therefore made artificial balls of clay, squeezing fluid from elephant dung on them. Arriving beetles accepted them readily, attempted to roll them away and defended them vigorously. In our staged contests we observed as many successful takeovers as successful defenses. The winner was usually the beetle with the higher temperature. Indeed, high body temperature was more important than size in determining the winner of a contest.

A beetle has little to lose, except for



MORPHOLOGICAL ADAPTATIONS of African dung beetles reflect the different ways of dealing with dung. The large *Heliocopris dilloni*, which is a digger that feeds only on elephant dung, has shovel-like front tibias to aid digging and backward-pointing spines on the tibias of the hind legs to improve traction (*left*). The ball roller *Scarabaeus laevistriatus* (*right*) has long, thin hind legs that are suited for running and for clinging to dung while the beetle forms a ball.



More than just an exciting destination for your (inter)national meeting.

This is where things are happening. Through our contacts you can reach Science and Technology, Trade and Industry, Theater, Art and Business too. Give us a call or write. We'll send you the facts or brief you in person.

ICC BERLIN

International Congress Center Berlin

AMK Berlin
Company for Exhibitions, Fairs and Congresses, Ltd.

International Congress Center Berlin
Congress Hall Berlin
Exhibition Grounds Berlin
Deutschlandhalle/Ice Palace Berlin

Box 191740, Messedamm 22,
D-1000 Berlin 19

Our Representative: Dieter H. von Schramm
1735 Eye Street, N.W., Suite 718,
Washington, D.C. 20 006
Telephone: (202) 659-0440, Telex: ttm 64 134

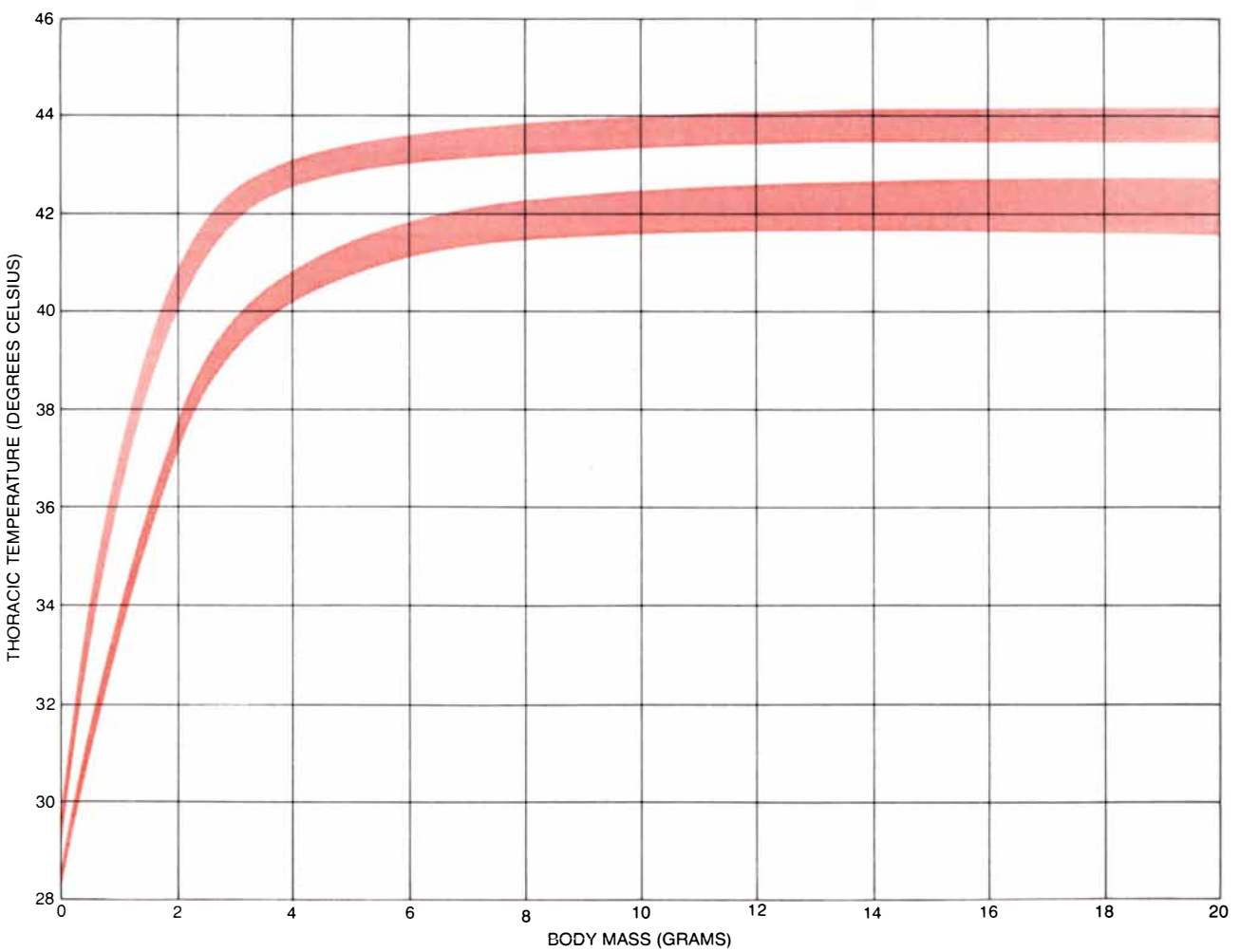
the investment of energy, from risking a fight. The fights are short (seldom lasting longer than 10 seconds) and result in few if any injuries. Under these conditions it is clearly advantageous for a newly arrived beetle, still hot from flying, to risk a fight. It follows that if a beetle has already invested the time and energy needed to make a ball, it should roll the ball away and bury it as quickly as possible.

The *S. laevistriatus* beetles rolled their dung balls with alacrity. We measured rates as high as 14 meters per minute on level ground, but only in beetles with a thoracic temperature of 40 degrees or more. The velocity of ball-rolling was a direct function of the thoracic temperature. Beetles with a thoracic temperature of 42 degrees rolled at an average rate of 11.4 meters per minute, whereas at 32 degrees the average rate was 4.8 meters per minute. In all the *S. laevistriatus* beetles that we observed the thoracic temperature was elevated while the beetles were rolling balls.

The metabolic rate of endothermic beetles during such activity is at least as high as that of a small mammal such as a shrew, which must eat a quantity of food at least equal to its own weight every day in order to sustain the high energy cost of endothermy. When food is in short supply, many small birds and mammals depart from endothermy, becoming temporarily ectothermic and allowing their body temperature to drop almost to the temperature of their surroundings. The nocturnal *S. laevistriatus* beetles are on an energy tightrope. The amount of food energy they can get increases with the extent of their endothermy, but so does their energy expenditure. They can maximize the net amount of energy they get from food only if the timing of their energy expenditure is precise. Their periods of endothermy must coincide with the times when the potential yield of energy is highest.

Since *S. laevistriatus* beetles are often active just before and at sunset, one wonders why they have not become diurnal. In that way they would escape the intense competition from endocoprids at night. We do not know the answer. We can only suggest that in the evolutionary past (and perhaps now) diurnal birds and other predators have taken such a heavy toll of beetles during the day that it is advantageous to the beetles to remain nocturnal.

The findings on the role of endothermy in dung beetles are helpful in understanding the significance of its evolution in other animals, including birds and mammals. Dung beetles achieve their highest rates of activity on the ground when their body temperature is at the level necessarily generated during flight; they have not evolved the capacity to achieve high rates of activity at low body temperature. In beetles as in other animals it appears that the biochemical machinery is adapted to the highest body temperature it reaches when it has to work at its maximum rate, which is the case with a large beetle when it is flying. Hence the maximum rates of activity that can be achieved at all lower temperatures depend on the passive effects of temperature on metabolic processes, which usually double in rate with each increase of 10 degrees C. in body temperature.



IMPORTANCE OF TEMPERATURE to dung beetles is indicated by these curves, which represent the average thoracic temperature in flight (*top*) and at takeoff (*bottom*) of several species of African dung beetle. At 44 degrees Celsius a beetle is near the lethal temperature.

If you want to know how to read our confusing economy, read these.

Will the economy pick up again in the months ahead? Or can we expect stagnation—or worse—from higher oil prices, a shortage of capital, slowing productivity?

To help answer questions like these, some of the finest economic thought leaders in America have participated in 23 lecture series at U.S. colleges and universities.

These programs, the ITT Key Issues Lecture Series, have included Dean Rusk, Isaac Asimov, Vernon E. Jordan, Jr., Juanita Kreps, Daniel Boorstin, former Congressman John E. Moss, Lee Loevinger, Raymond Vernon, among others.

Their talks have been brought together in a number of provocative, enlightening volumes.

Each has an insightful foreword by Harold S. Geneen, ITT Chairman.

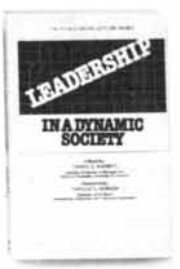
Here are the latest.

ITT

The best ideas are the ideas that help people.

An acute look at the role of technology, capital formation and other factors in shaping America's economic future. Jules Backman (ed.), Edwin Mansfield, Norma Pace, others.

☐ hardcover @$9.95
☐ paperback @$4.95



Leadership looms as our latest shortage—and without leadership, other problems facing us are magnified. The alternatives are discussed by six authorities, among them: Frank E. Kuzmits (ed.), John B. Miner, Earl G. Graves.

☐ hardcover @$9.95
☐ paperback @$4.95

Instead of reforming taxes, recent laws could cause still more problems. Questions raised by Donald H. Skadden (ed.), Murray L. Weidenbaum, others.

☐ hardcover @$9.95
☐ paperback @$4.95

How will dwindling resources, population trends, energy costs affect business? An interdisciplinary review. Howard F. Van Zandt (ed.), Edward T. Hall, Sir James Lindsay, James M. Voss, others.

☐ hardcover @$9.95
☐ paperback @$4.95

Like everything else, accounting procedures are undergoing significant changes these days. Experts in business, government and academic life provide a lucid, practical guide. They include Dhia D. AlHashim and James W. Robertson (eds.), James Don Edwards, former Congressman John E. Moss.

☐ hardcover @$12.50
☐ paperback @$6.95

The impact of inflation, government policy and international events on capital formation—and the impact that capital scarcity is likely to have. Among those discussing the issues: Alan Heslop (ed.), David I. Meiselman, Geoffrey Bell.

☐ hardcover @$9.95
☐ paperback @$4.95

Order these essential books today. Write to:

Bobbs-Merrill Educational Publishing
4300 West 62nd Street, Dept. 1079Q
Indianapolis, Indiana 46206

Please send me copies of the books I've indicated (for more than one copy of any title, simply indicate number in parentheses). I enclose check or money order for $_____ postpaid by return mail. *(Orders received on company letterhead can be invoiced after shipment.)*

Name_____
Company_____
Street_____
City_____ State_____ Zip_____

The Quantum Theory and Reality

The doctrine that the world is made up of objects whose existence is independent of human consciousness turns out to be in conflict with quantum mechanics and with facts established by experiment

by Bernard d'Espagnat

Any successful theory in the physical sciences is expected to make accurate predictions. Given some well-defined experiment, the theory should correctly specify the outcome or should at least assign the correct probabilities to all the possible outcomes. From this point of view quantum mechanics must be judged highly successful. As the fundamental modern theory of atoms, of molecules, of elementary particles, of electromagnetic radiation and of the solid state it supplies methods for calculating the results of experiments in all these realms.

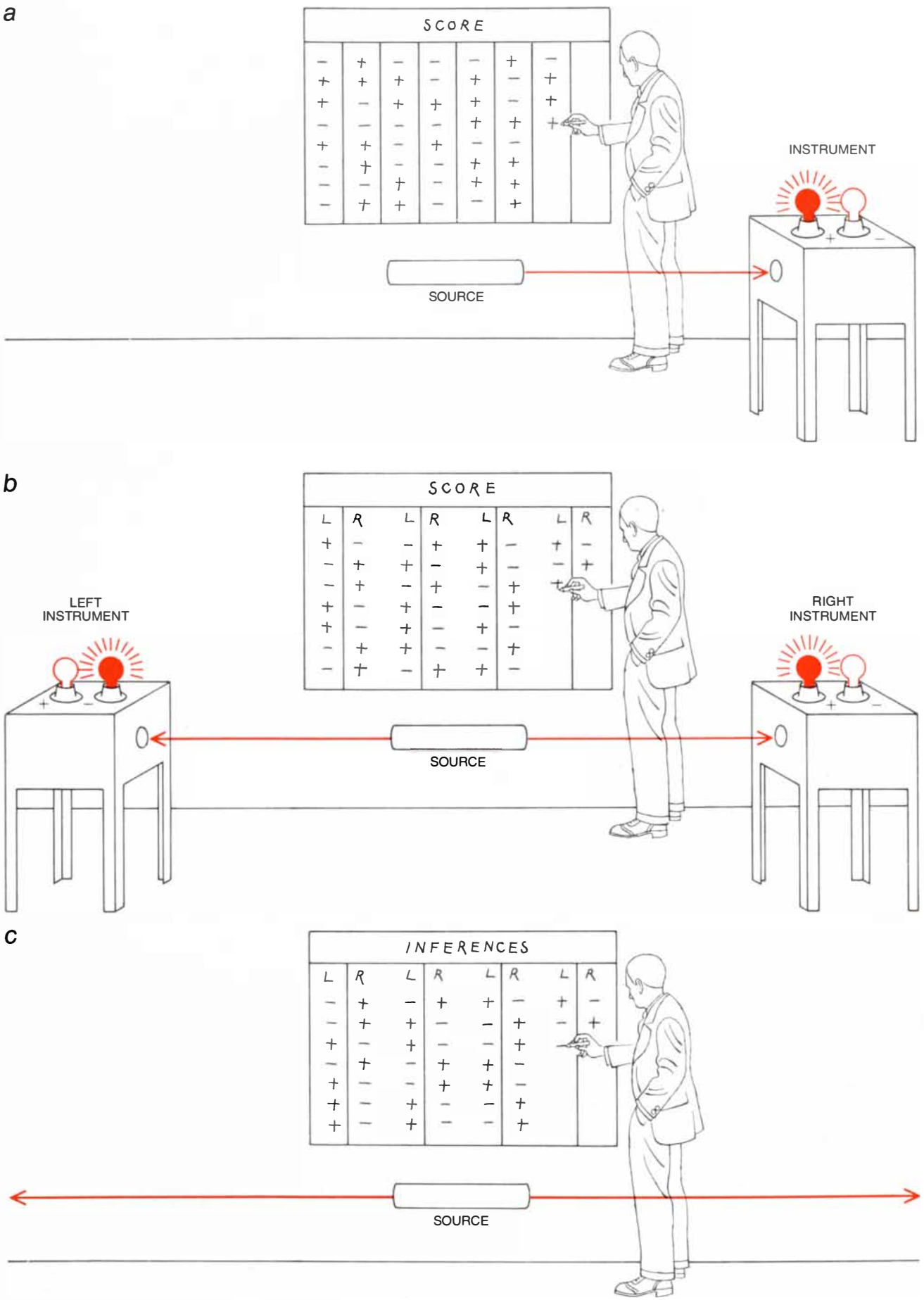
Apart from experimental confirmation, however, something more is generally demanded of a theory. It is expected not only to determine the results of an experiment but also to provide some understanding of the physical events that are presumed to underlie the observed results. In other words, the theory should not only give the position of a pointer on a dial but also explain why the pointer takes up that position. When one seeks information of this kind in the quantum theory, certain conceptual difficulties arise. For example, in quantum mechanics an elementary particle such as an electron is represented by the mathematical expression called a wave function, which often describes the electron as if it were smeared out over a large region of space.

This representation is not in conflict with experiment; on the contrary, the wave function yields an accurate estimate of the probability that the electron will be found in any given place. When the electron is actually detected, however, it is never smeared out but always has a definite position. Hence it is not entirely clear what physical interpretation should be given to the wave function or what picture of the electron one should keep in mind. Because of ambiguities such as this many physicists find it most sensible to regard quantum mechanics as merely a set of rules that prescribe the outcome of experiments. According to this view the quantum theory is concerned only with observable phenomena (the observed position of the pointer) and not with any underlying physical state (the real position of the electron).

It now turns out that even this renunciation is not entirely satisfactory. Even if quantum mechanics is considered to be no more than a set of rules, it is still in conflict with a view of the world many people would consider obvious or natural. This world view is based on three assumptions, or premises that must be accepted without proof. One is realism, the doctrine that regularities in observed phenomena are caused by some physical reality whose existence is independent of human observers. The second premise holds that inductive inference is a valid mode of reasoning and can be applied freely, so that legitimate conclusions can be drawn from consistent observations. The third premise is called Einstein separability or Einstein locality, and it states that no influence of any kind can propagate faster than the speed of light. The three premises, which are often assumed to have the status of well-established truths, or even self-evident truths, form the basis of what I shall call local realistic theories of nature. An argument derived from these premises leads to an explicit prediction for the results of a certain class of experiments in the physics of elementary particles. The rules of quantum mechanics can also be employed to calculate the results of these experiments. Significantly, the two predictions differ, and so either the local realistic theories or quantum mechanics must be wrong.

The experiments in question were first proposed as "thought experiments," intended for the imagination only. In the past few years, however, several versions of them have been carried out with real apparatus. Although not all the findings are consistent with one another, most of them support the predictions of quantum mechanics, and it now seems that unless some extraordinary coincidence has distorted the results the quantum-mechanical predictions will be confirmed. It follows that the local realistic theories are almost certainly in error. The three premises on which those theories are founded are essential to a common-sense interpretation of the world, and most people would give them up only with reluctance; nevertheless, it appears that at least one of them will have

CORRELATIONS BETWEEN DISTANT EVENTS can form the basis of conclusions about the structure of the world. Suppose a physicist sets up an experiment in which subatomic particles such as protons are fired one at a time into an instrument that can give only two possible readings, plus and minus (*a*). He finds that for some protons the reading is plus and for others it is minus, but he cannot tell whether the instrument measures some real property of the protons or merely records random fluctuations. The physicist then arranges two identical instruments with a source that emits two protons simultaneously (*b*). He observes a strict negative correlation: whenever one instrument reads plus, the other reads minus. On the basis of this correlation the physicist concludes that a real property of protons is responsible for the readings and that its value is determined before the protons leave the source. If the sample of particles measured meets certain statistical tests, he can go on to infer that every pair of protons emitted by the source consists of one proton with the property plus and one with the property minus, even if neither proton is submitted to a measurement (*c*). The conclusions are reasonable if three premises are accepted as valid: that at least some properties of the world have an existence independent of human observers, that inductive inference can be applied freely and that a measurement made with one instrument cannot influence the result of a measurement made with the other instrument. A more restrictive form of the last premise forbids such influences only if the two measurements are so nearly simultaneous that the influence would have to propagate faster than light. The premises can be identified as realism, the free use of induction and separability; the more restrictive version of the separability premise is called Einstain separability or Einstein locality. Any theory that incorporates them is called a local realistic theory.

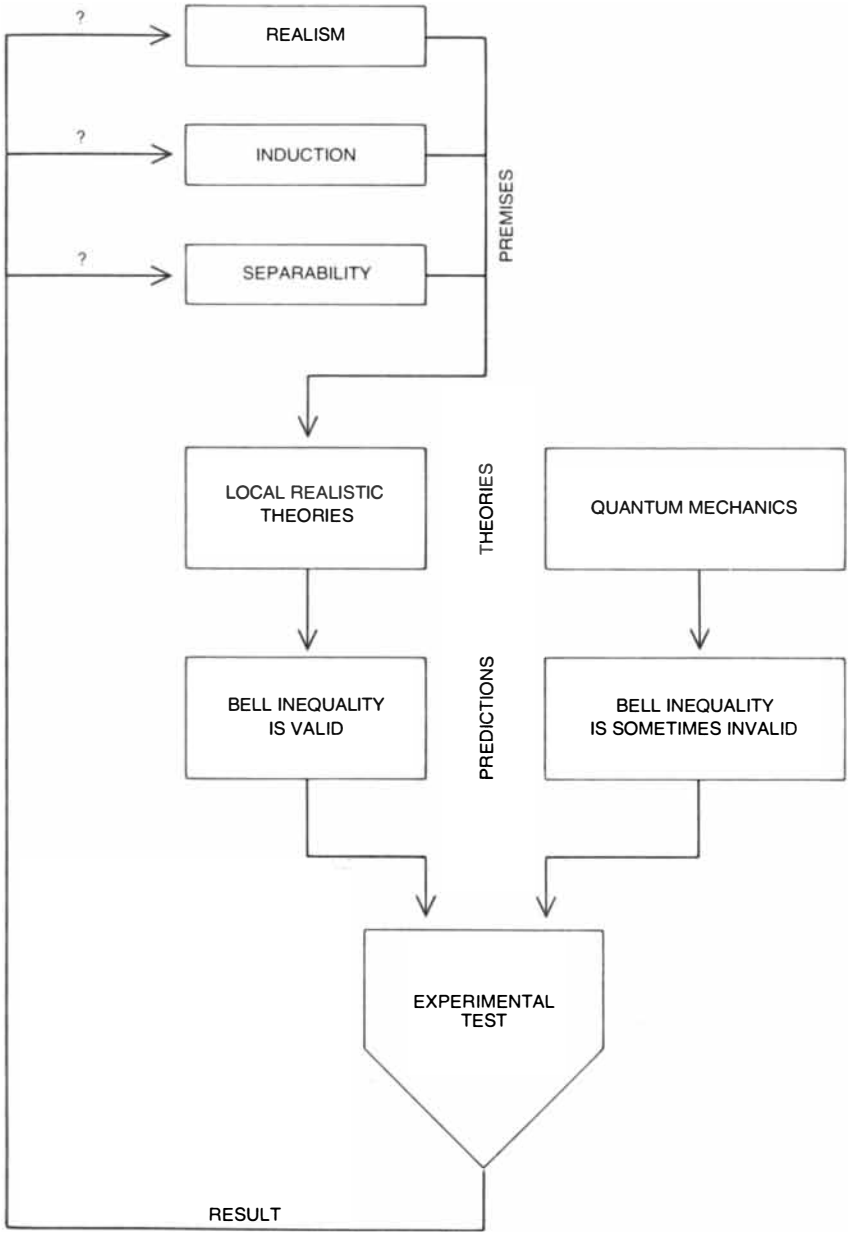


to be abandoned or modified or in some way constrained.

The experiments are concerned with correlations between distant events and with the causes of those correlations. For example, suppose two particles a few meters apart are found to have identical values of some property, such as electric charge. If this result is obtained once or a few times, it might be dismissed as coincidence, but if the correlation is detected consistently in many measurements, a more systematic explanation is called for. It would make no difference if the measured values were always opposite instead of the same; the correlation would then be a negative one, but its magnitude would be just as great, and it would be just as unlikely to arise by chance.

Whenever a consistent correlation between such events is said to be understood, or to have nothing mysterious about it, the explanation offered always cites some link of causality. Either one event causes the other or both events have a common cause. Until such a link has been discovered the mind cannot rest satisfied. Moreover, it cannot do so even if empirical rules for predicting future correlations are already known. A correlation between the tides and the motion of the moon was observed in antiquity, and rules were formulated for predicting future tides on the basis of past experience. The tides could not be said to be understood, however, until Newton introduced his theory of universal gravitation.

The need to explain observed correlations is so strong that a common cause is sometimes postulated even when there is no evidence for it beyond the correlation itself. Whether or not this procedure can always be justified is a central issue in the conflict between quantum mechanics and local realistic theories. The correlations in question are between observations of subatomic particles, where a quantum-mechanical description, with its attendant epistemological hazards, is indispensable. The predictions of local realistic theories, however, can be illustrated by considering how correlations between distant events are explained in a more familiar context, where quantum mechanics need not be introduced.



LOCAL REALISTIC THEORIES and quantum mechanics make conflicting predictions for certain experiments in which distant events are correlated. In particular, local realistic theories predict that a relation called the Bell inequality will be obeyed, whereas quantum mechanics predicts a violation of the inequality. There is strong experimental evidence that the inequality is violated in the way predicted by quantum mechanics. Local realistic theories therefore seem to be untenable, and at least one of the premises underlying those theories must be in error.

Imagine that a psychologist has devised a simple test, which a subject must either pass or fail, so that there can be no ambiguity in the results. The psychologist finds that some people pass and some fail, but he does not know what distinguishes the two groups other than their performance on the test itself. In other words, he cannot tell whether the test measures some real aptitude or attribute of the subjects or whether the results are haphazard.

It seems there is no general solution to this problem, but in a special case it might be solved. Suppose the test is administered not to a series of individuals but to a series of married couples and that a strong correlation is detected in their answers. The procedure might consist in separating the husbands from the wives before the test and then giving the test to each of them in isolation. When the results are analyzed, it is found again that part of the population has passed and part has failed, but in the case of each couple where the husband passed so did the wife; similarly, whenever the husband failed so did the wife.

If this correlation persists after many couples are tested, the psychologist is almost sure to conclude that the response of each subject is not determined randomly at the time of testing. On the contrary, the test must reveal some real property or attribute of the subjects. The property must already be present in the subjects before they are tested, and indeed before they are separated. Chance may have had some influence on the development of the property,

Technics

Nobody's perfect. But Technics quartz-locked, direct-drive Q-Series: the Q-2 semi-automatic and Q-3 fully automatic come incredibly close.

So close that many discos and FM stations choose Technics quartz-locked, direct-drive turntables over any other. It's no wonder, with speed accuracy of 0.002%, wow and flutter of only 0.025% WRMS and rumble of −78dB (DIN B). They're impressive specs.

What's just as impressive is Technics soft-touch in-line controls conveniently mounted on the front panel. You can operate every electronic function without ever lifting the dust cover.

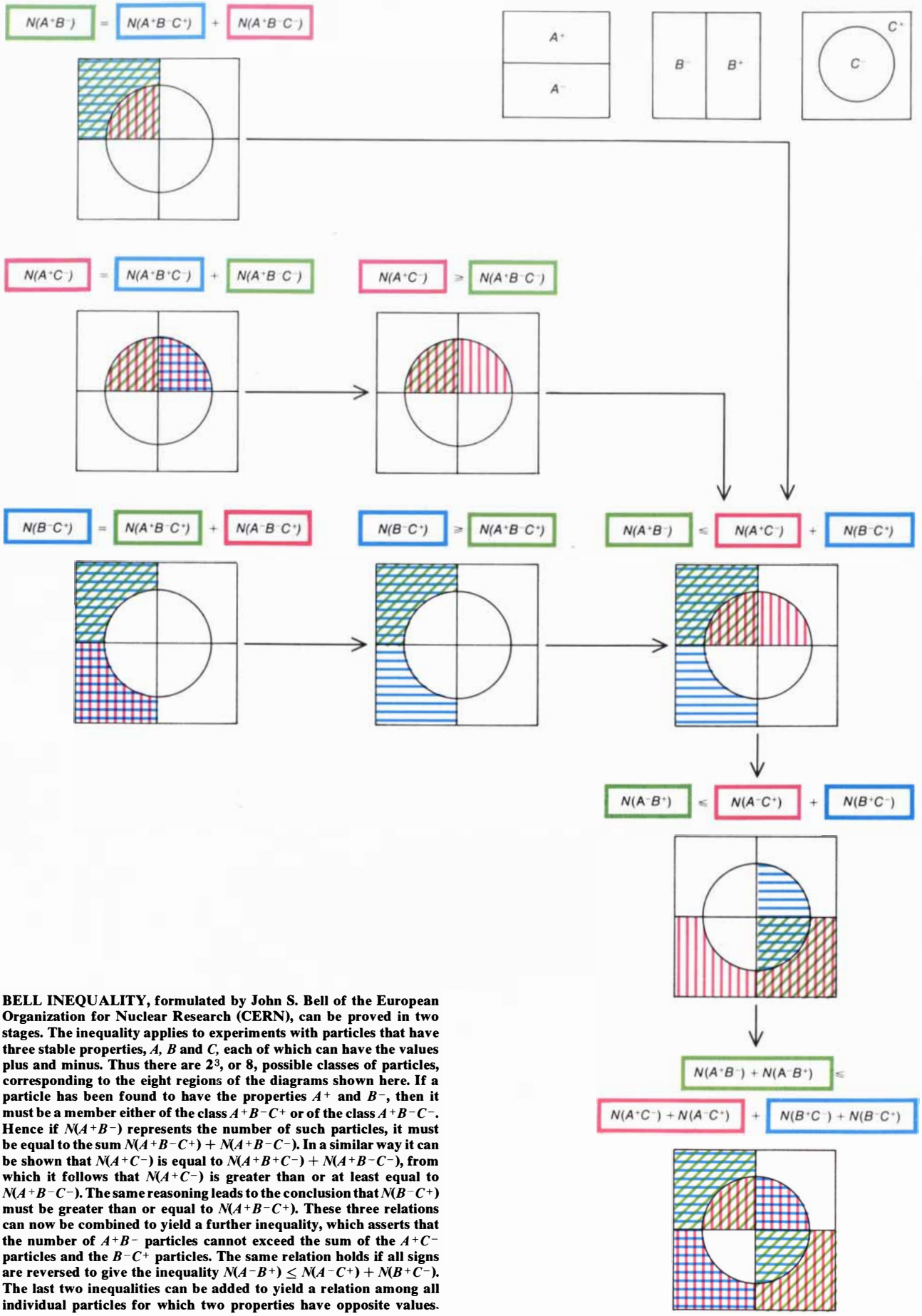
Or Technics statically balanced S-shaped tonearm. With only 7 mg friction on both the vertical and horizontal planes, it's more than sensitive, it's sensational. Even the computer-designed headshell contacts are gold-plated for maximum conductivity.

To help protect against acoustic feedback, Technics Q-Series turntables are all mounted in a precision aluminum diecast base with a unique non-resonant compound, TNRC. It's so effective it resists feedback at the highest music levels.

By this time you might think you have to be rich to afford Technics Q-Series. You don't. Both models are surprisingly reasonable.

Technics Q-Series. We can't say they're perfect. You will.

We can't say the speed accuracy of our new quartz-locked turntables is 100%. Just 99.998%.



BELL INEQUALITY, formulated by John S. Bell of the European Organization for Nuclear Research (CERN), can be proved in two stages. The inequality applies to experiments with particles that have three stable properties, A, B and C, each of which can have the values plus and minus. Thus there are 2^3, or 8, possible classes of particles, corresponding to the eight regions of the diagrams shown here. If a particle has been found to have the properties A^+ and B^-, then it must be a member either of the class $A^+B^-C^+$ or of the class $A^+B^-C^-$. Hence if $N(A^+B^-)$ represents the number of such particles, it must be equal to the sum $N(A^+B^-C^+) + N(A^+B^-C^-)$. In a similar way it can be shown that $N(A^+C^-)$ is equal to $N(A^+B^+C^-) + N(A^+B^-C^-)$, from which it follows that $N(A^+C^-)$ is greater than or at least equal to $N(A^+B^-C^-)$. The same reasoning leads to the conclusion that $N(B^-C^+)$ must be greater than or equal to $N(A^+B^-C^+)$. These three relations can now be combined to yield a further inequality, which asserts that the number of A^+B^- particles cannot exceed the sum of the A^+C^- particles and the B^-C^+ particles. The same relation holds if all signs are reversed to give the inequality $N(A^-B^+) \leq N(A^-C^+) + N(B^+C^-)$. The last two inequalities can be added to yield a relation among all individual particles for which two properties have opposite values.

since not all the couples possess it, but that influence must have been exerted at some time before the husbands and the wives were separated. It was only then, while the husbands and the wives were still united, that they could have acquired any traits that would induce them to respond consistently the same way. Thus the correlation is explained by attributing it to a common cause antecedent to the test.

One other explanation that must be excluded in deriving this conclusion is the possibility that husbands and wives could communicate with each other while they were taking the test. If some means of communication were available, there would be no need for any tested attribute to exist beforehand. Whichever spouse was given the test first could choose a response at random and send instructions to the other, thereby creating the observed correlation. In giving a psychological test it would not be hard to guard against subterfuge of this kind. In the extreme case the tests could be made so nearly simultaneous, or husbands and wives could be tested at sites so far apart, that a signal moving no faster than light could not arrive in time to be of any value.

Once having decided that the test measures some real property of individuals, the psychologist can take a further step and make an inductive inference. If the couples already tested constitute an unbiased sample of some population of couples, and if the sample meets certain statistical standards, the psychologist can infer that any couple taken from the same population will be made up of a husband and a wife who either both possess or both do not possess the property measured by the test. By the same principle he can conclude that in any large, unbiased sample of couples who have not yet been tested some of the couples will have the property and some will not. The confidence of these assertions approaches certainty as the size of the sample increases. Hence both the correlation within couples and the existence of differences between couples are inferred to exist even in the segment of the population that has not been submitted to any test.

These conclusions rest on the same three premises that form the basis of local realistic theories. Realism is a necessary assumption if one is to believe at least some tests measure stable properties that exist independently of the experimenter. It was necessary to assume the validity of inductive inference in order to extrapolate from the observed data to the segment of the population that had not yet been tested. Separability was incorporated in the assumption that husbands and wives being tested cannot communicate with each other. If the tests are given simultaneously, so that any signal passing between hus-

THE BMW 528i. A CAR THAT MEETS THE DEMANDS OF THE 80's WITHOUT VIOLATING THE CONCEPT OF A BMW.

Conventional automotive wisdom has it that any increase in a car's environmental control paraphernalia must inevitably result in a corresponding decrease in performance.

Indeed, many serious automotive writers have warned that pollution systems and fuel efficiency regulations have all but sounded the death knell for high-performance automobiles.

These gloomy predictions, however, do not take into consideration the determination of the engineers at the Bavarian Motor Works to build extraordinary automobiles.

Against all odds, the BMW 528i not only meets the demands of society, it also provides the kind of exhilarating driving experience that most people have all but given up for lost in today's luxury automobiles.

AMIDST INCREASING MEDIOCRITY, BMW GETS EVEN BETTER.

Press the accelerator and the 528i's fuel-injected, six-cylinder, overhead cam engine responds in a manner that can only be described as exhilarating.

And yet, the 528i (with standard transmission) delivers an impressive 17 EPA estimated mpg, 26 estimated highway mileage and, based on these figures, an estimated mpg range of 279 miles and a highway range of 426 miles.

(Naturally our fuel efficiency figures are for comparison purposes only. Your actual mileage and range may vary, depending on speed, weather and trip length. Your actual highway mileage and highway range will most likely be lower.)

Its four-speed manual transmission (automatic is available) runs smoothly and precisely up through the gears.

Its suspension system—independent on all four wheels—provides the driver with an uncanny feel of the road.

And, while the 528i provides as long a list of luxury items as one could sanely require—AM/FM stereo cassette, full-power accessories, air conditioning, etc.—its luxury is purposefully engineered to help prevent driver fatigue.

Vital controls are within easy reach; the tachometer, speedometer and ancillary instruments are well-marked and easy to read.

Its front seats are designed to hold their occupants firmly in place, and are so thoroughly adjustable that it is all but impossible not to find a comfortable seating position.

If you'd care to judge the extraordinary performance of the 528i for yourself, phone your nearest BMW dealer and he will arrange a thorough test drive at your convenience.

THE ULTIMATE DRIVING MACHINE.

Bavarian Motor Works, Munich, Germany.

bands and wives would have to propagate faster than the speed of light, the assumption is equivalent to Einstein separability.

At first the conclusions drawn from this hypothetical experiment in psychology seem to follow quite obviously from the data. An epistemologist might nonetheless maintain that the conclusions are uncertain. In particular an epistemologist trained in the foundations of quantum mechanics might argue that there is no logical necessity for accepting the three premises of the psychologist's argument; hence neither would it be necessary to conclude that a correlation existed between the husbands and wives before they were tested, or that differences existed between the couples before any tests were given. The psychologist is likely to find these objections laughable, an expression of misplaced doubt or of a very unscientific adherence to paradox. In the literature of quantum mechanics, however, there are numerous arguments similar or equivalent in form to this one, all purporting to show that correlations and differences need not exist until they are measured.

A singular feature of quantum mechanics is that its predictions generally give only the probability of an event, not a deterministic statement that the event will happen or that it will not. The wave function employed to describe the motion of an elementary particle is often interpreted probabilistically: the probability of finding the particle at any given point is proportional to the square of the wave function at that point. As I mentioned above, a wave function can sometimes be spread out over a large region, which implies that the probability can also be broadly distributed. Of course, when a measurement is actually made at some chosen point, the particle must either be detected or not be detected; the wave function is then said to collapse. Suppose the particle is detected. The question of epistemological interest is then: Did the particle have that definite position all along, even before the measurement was made?

The conclusions of the psychologist, if they could be transferred to this context, would imply that the position of the particle was well defined from the start, just as the attribute discovered in some members of the population was deduced to have existed before any tests were given. According to this argument the position of the particle was never indeterminate but was merely unknown to the experimenter.

Most authorities on the quantum theory would disagree. One exception among physicists was Einstein, who throughout his life remained dissatisfied with the probabilistic nature of the interpretations generally given to quantum mechanics. He based his most incisive criticism of those interpretations on an argument that was somewhat similar to the one I have attributed to the psychologist. In 1935 Einstein published a paper with two young colleagues, Boris Podolsky and Nathan Rosen, in which he stated his objections explicitly. He did not maintain that the quantum theory is wrong; on the contrary, he assumed that at least some of its predictions must be correct. What he proposed was that the quantum-mechanical description of nature is incomplete or approximate. The motion of a particle must be described in terms of probabilities, he argued, only because some of the parameters that determine the motion have not yet been specified. If the values of these hypothetical "hidden parameters" were known, a fully deterministic trajectory could be defined.

A number of counterarguments to Einstein's proposal have been formulated. For now I shall mention only one of them, which is based on the criterion of utility. It is immaterial, the argument states, whether or not hidden parameters exist, or whether differences between married couples exist in the absence of a test. Even if they do exist, they should not be incorporated into any theory devised to explain the observations, and so they can be said to have no scientific existence. The exclusion of the hidden parameters is justified by the conjunction of three facts. First, the mathematical formalism of the theory is simpler if any hidden parameters are ignored. Second, this simple formalism predicts results that are confirmed by experiment. Third, adding the hidden parameters to the theory would give rise to no supplementary predictions that could be verified. Thus the assertion that hidden parameters exist is beyond the reach of experiment and is a proposition not of physics but of metaphysics.

This defense of the conventional interpretation of quantum mechanics dismisses any hidden parameters as being superfluous and ultimately, perhaps, meaningless. Recent theoretical developments have shown that their actual status is quite different. The hypothesis that hidden parameters exist does in fact lead to experimental predictions differing from those of quantum mechanics. Hidden-parameter theories, and local realistic theories in general, place a limit on the extent to which certain distant events can be correlated; quantum mechanics, in contradistinction, predicts that under some circumstances the limit will be exceeded. Hence it should be possible, at least in principle, to devise an experimental test that will discriminate between the two theories.

Suppose a physicist has devised a test that can be carried out on subatomic particles such as protons. After many trials he finds that some protons pass the test and others fail, but he does not know whether he is measuring some real property of the protons or merely observing random fluctuations in his apparatus. He therefore tries applying the test not to individual protons but to pairs of them. The protons that make up each pair are initially in close proximity, having been brought together by some well-defined procedure that is the same for all the pairs. The protons are then allowed to separate, and when they have moved some macroscopic distance apart, they are tested, simultaneously for some pairs and with an interval between the tests for the remaining pairs. The physicist discovers a strict negative correlation: whenever one proton in a pair passes the test, the other proton invariably fails.

The situation of the physicist has obvious similarities to that of the psychologist giving a test to married couples, and the same reasoning might be applied to the results of the physical experiment. If realism, the free use of induction and Einstein separability are all accepted as premises, then the physicist is justified in concluding that his test does measure some real property of protons. For the correlation to be explained the property must exist before the protons in each pair are separated, and it must have some definite value from then until the measurement is made. Furthermore, if additional pairs of protons are prepared by the same method, the physicist knows that in each case one proton will have the property and one will not, even if neither proton is actually tested.

Is there any real test that can be carried out on subatomic particles with results like these? There is. It is a measurement of any one component, defined along some arbitrary axis, of the spin of a particle. The spin attributed to a subatomic particle is analogous only in some respects to the spin angular momentum of a macroscopic body such as the earth. For the purposes of this discussion, however, there is no need to introduce the details of how spin is treated in quantum mechanics. It will suffice to note that the spin of a particle is represented by a vector, or arrow, that can be imagined as being attached to the particle. A projection of this vector onto any axis in three-dimensional space is the component of the spin along that axis. A well-established but nonetheless surprising property of protons (and many other particles) is that no matter what axis is chosen for a measurement of a spin component the result can take on only one of two values, which I shall designate plus and minus. (A measurement of a component of the earth's spin would give very different results; depending on the direction of the component, it could have any value from zero up to the total angular momentum of the earth.)

A strict negative correlation between spin components is observed when any two protons are brought together in

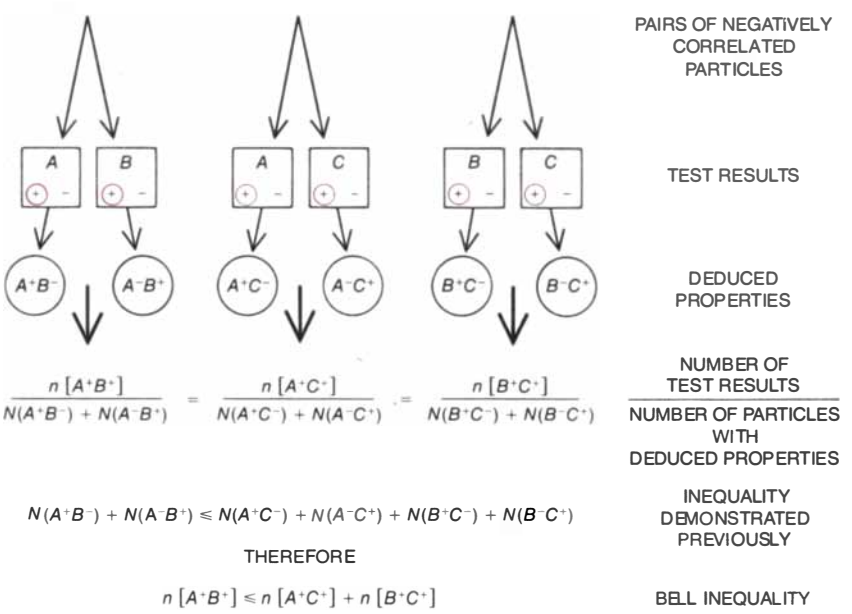
the quantum-mechanical configuration called the singlet state. In other words, if two protons in the singlet state are allowed to separate and the same component of spin is subsequently measured on both particles, it will always be plus for one proton and minus for the other. There is no known means of predicting which particle will have the plus component and which the minus component, but the negative correlation is well established. It makes no difference what component of the spin the experimenter chooses to measure, provided the same component is measured for both particles. It also makes no difference how far the protons travel before the measurement is made, as long as there are no perturbing influences, such as other particles or radiation, along their paths.

In this simple measurement there is no conflict between the predictions of quantum mechanics and those of local realistic theories. A conflict can arise, however, when the experiment is made somewhat more complicated.

The vector that represents the spin of a particle is defined by components along three axes in space, which need not necessarily be at right angles to one another. For a vector associated with a macroscopic object in everyday life, one would assume as a matter of course, and with good reason, that all three components have definite values at all times; the value of a component might be unknown, but it cannot be undefined. When this assumption is applied to the spin vector of a particle, however, it becomes highly suspect, and indeed in the conventional interpretation of quantum mechanics it is dismissed as an instance of a hidden-parameter theory. The problem is that no experiment can be devised, even in principle, that would provide information about the simultaneous values of all three components. A single instrument can measure only one spin component, and in doing so it generally alters the values of the components. Hence in order to learn the values of three components three measurements would have to be made in succession. By the time the particle emerged from the third instrument it would no longer have the same spin components it had when it entered the first instrument.

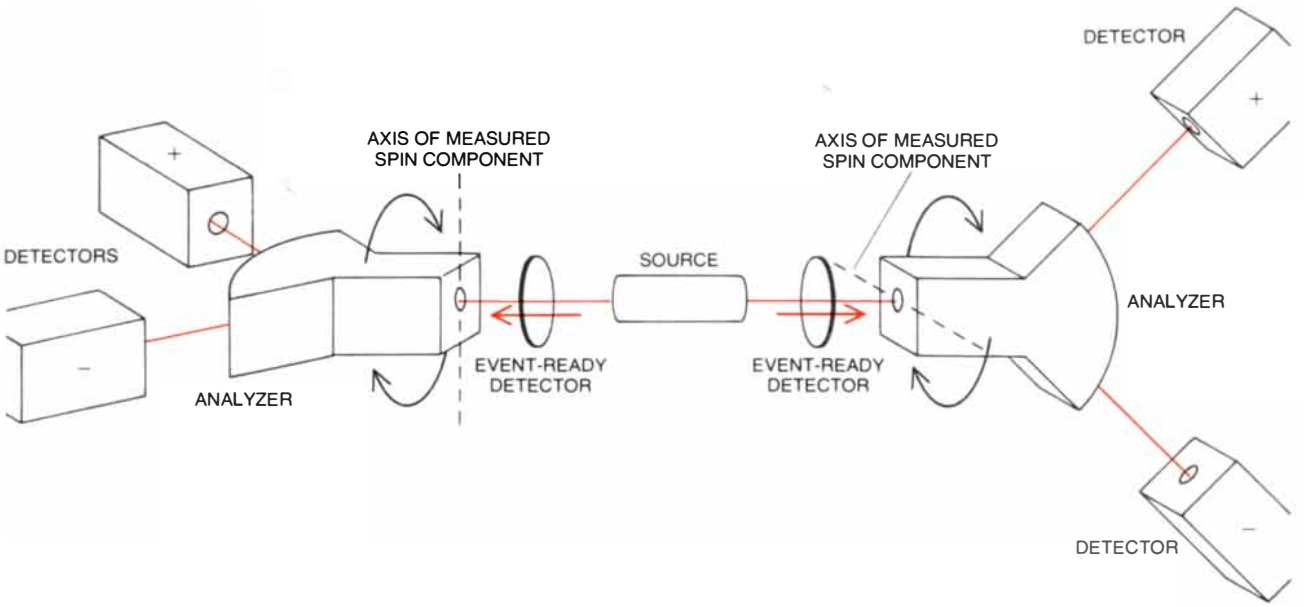
Although no instrument can measure more than one spin component at a time, a device can be built that is capable of being adjusted to measure the spin component along any one of three arbitrarily chosen axes. I shall designate these axes *A, B* and *C* and note the results of experiments as follows. If the spin component along axis *A* is found to be plus, it is labeled A^+; if the component along axis *B* is minus, it is given as B^-, and so on. The physicist can now prepare a large batch of protons in the singlet state. He finds that if he measures component *A* for both protons in each pair, some protons are A^+ and others are A^-, but whenever one member of a pair is A^+, the other member is always A^-. If he decides instead to measure component *B,* he observes the same negative correlation: whenever one proton is B^+, its singlet partner is B^-. Similarly, a C^+ proton is invariably accompanied by a C^- one. These results hold no matter how the axes *A, B* and *C* are oriented.

It is important to emphasize that in these experiments no proton is submitted to a measurement of more than one spin component. Nevertheless, if the physicist accepts the three premises of local realistic theories, he can draw conclusions from these findings about the values of all three components, following an argument much like that of the hypothetical psychologist. Considering a fresh batch of proton pairs in the singlet state on which no spin measurement has yet been made (and perhaps on which no such measurement will ever be made), he can infer that in every pair one proton has the property A^+ and the other has the property A^-. Similarly, he can conclude that in every pair one proton has the property B^+ and one B^- and one has the property C^+ and one C^-.



SECOND STAGE OF THE PROOF extrapolates from the case of single particles for which two properties are known to that of pairs of particles, each particle of which is tested for one property. The pairs are created in such a way that there is always a strict negative correlation for any property considered separately, that is, if one particle in a pair has the property A^+, the other must have the property A^-. Because of this correlation, if one particle in a pair is found to be A^+ and the other is found to be B^+, it is possible to deduce both properties of both particles. The doubly positive test result can arise only if one particle has the two properties A^+B^- and the other has the properties A^-B^+. Hence the number of such doubly positive test results, which can be designated $n[A^+B^+]$, must be proportional to the total number of particles with the properties A^+B^- and A^-B^+. Similar proportionalities can be derived for the number of doubly positive results observed when pairs of particles are tested for properties *A* and *C* and for properties *B* and *C;* these are the quantities $n[A^+C^+]$ and $n[B^+C^+]$. The constant of proportionality depends only on the number of pairs submitted to each set of tests and on the total number of pairs, and so the constant is the same in all three cases. It follows that the three ratios of the number of doubly positive test results to the number of individual particles that can give rise to those results must also be equal. A relation has already been demonstrated between the numbers of individual particles with the indicated properties; it is the inequality proved in the illustration on page 162. If that inequality is to hold, there must be a similar inequality between the numbers of doubly positive test results. This is the Bell inequality. The proof is valid only if the three premises of local realistic theories are assumed to be valid.

These conclusions require a subtle but important extension of the meaning assigned to a notation such as A^+. Whereas previously A^+ was merely one possible outcome of a measurement made on a particle, it is converted by this argument into an attribute of the particle itself. To be explicit, if some unmeasured proton has the property that a measurement along the axis *A* would give the definite result A^+, then that proton is said to have the property A^+. In other words, the physicist has been led to the conclusion that both protons in each pair have definite spin components at all times. The components may be unknown, since the physicist cannot say which proton in a pair has the property A^+ and which has the property A^- until a measurement along axis *A* has been made, but he can argue from the premises of local realistic theories that the values are quite definite even in the absence of any measurements. This view is contrary to the conventional in-



THOUGHT EXPERIMENT would test the Bell inequality by measuring the components of the spin of protons or other elementary particles. A spin component is a projection along some axis of the proton's intrinsic angular momentum; each component can have only two possible values, which can be designated plus and minus. The experiment, which assumes the availability of perfect instruments, would have a source where pairs of protons are brought together in a quantum-mechanical configuration called the singlet state. The pairs would then be broken up, and the protons would fly apart in opposite directions. "Event-ready" detectors would issue a signal whenever a suitable pair of protons had been emitted. Each proton would then enter an analyzer, where it would be deflected to one of two detectors depending on the value of its spin component along the axis defined by the analyzer. If the analyzers were set to measure the spin components along the same axis, a strict negative correlation would be observed. If one analyzer were rotated, so that they measured different components, local realistic theories predict that the correlation observed would be no greater than that allowed by the Bell inequality regardless of what the angle between the analyzers was. Quantum mechanics predicts a violation of the Bell inequality for some angles.

terpretation of quantum mechanics, but it is not contradicted by any fact that has yet been introduced.

The strict negative correlation for protons in the singlet state is expected only when the same spin component is measured on both protons. What happens when the instruments are set to measure different components? To be precise, consider the following experiment. Pairs of protons are brought together in the singlet state by the same method employed in the earlier experiments and are allowed to separate under exactly the same conditions. Each proton is then tested for just one spin component, A, B or C, but which one of the components is measured in each case is determined entirely at random. Sometimes by coincidence the same component will be measured on both protons in a pair; those results are discarded, since they provide no new information. The remaining pairs must then be made up of either one proton tested along axis A and one tested along axis B, or one tested along axis A and one along axis C, or one along axis B and one along axis C. For the sake of brevity I shall refer to the pairs in each of these three populations as AB, AC and BC. A pair that on testing yields the results A^+ for one proton and B^+ for the other can be labeled an A^+B^+ pair. The number of such pairs observed can be represented by the notation $n[A^+B^+]$. Can any relation among these quantities be expected?

In 1964 John S. Bell of the European Organization for Nuclear Research (CERN) discovered such a relation. For any large sample of singlet proton pairs Bell showed that the tenets of local realistic theories impose a limit on the extent of correlation that can be expected when different spin components are measured. The limit is expressed in the form of an inequality, which is now called the Bell inequality. Given the experimental conditions described above, it states that the number of A^+B^+ pairs cannot exceed the sum of the number of A^+C^+ pairs and the number of B^+C^+ pairs. The inequality can be expressed in symbols as

$$n[A^+B^+] \leq n[A^+C^+] + n[B^+C^+].$$

Many similar inequalities could be constructed with the various symbols transposed or with the signs reversed. Because the directions along which the spin components are defined were chosen arbitrarily, all such formulations are interchangeable, and I shall discuss only this one.

The Bell inequality can be proved, within the context of local realistic theories, through a straightforward argument in the mathematical theory of sets. It is convenient to begin with an assumption contrary to fact: that some means exist for independently measuring two components of the spin of a single particle. Suppose this impossible instrument has revealed that a particular proton has the spin components A^+ and B^-. The third component, C, has not been measured, but it can have only one of two values, plus or minus; hence the measured proton must be a member of one of two sets of protons, either the set with spin components $A^+B^-C^+$ or the set with components $A^+B^-C^-$. There are no other possibilities.

If many protons with the spin components A^+B^- are detected, one can write an equation about their number:

$$N(A^+B^-) = N(A^+B^-C^+) + N(A^+B^-C^-).$$

In order to avoid confusion the symbol $N(A^+B^-)$ has been employed to represent the number of individual protons with the two spin components A^+ and B^-; the symbol $n[A^+B^-]$ gives the number of proton pairs in which one particle has the component A^+ and the other has the component B^-. The equation states the obvious fact that when a set of particles is divided into two subsets, the total number of particles in the original set must be equal to the sum of the numbers in the subsets.

The protons found to have the spin components A^+C^- can be analyzed exactly the same way. Every such proton must be a member either of the set $A^+B^+C^-$ or of the set $A^+B^-C^-$, and the total number $N(A^+C^-)$ must be equal to the sum $N(A^+B^+C^-) + N(A^+B^-C^-)$.

You can wait years for new developments to appear in standard texts, spend hundreds of hours extracting them from journals, or you can subscribe to SCIENTIFIC AMERICAN *Medicine*.

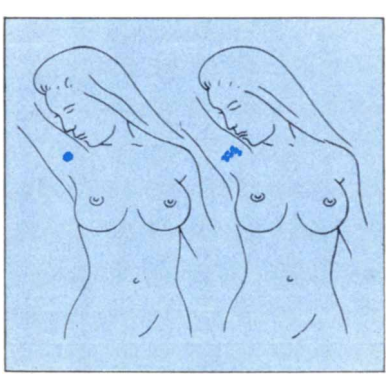
SCIENTIFIC AMERICAN *Medicine* is the clinician's answer to the diffuseness of the periodicals and the prepublication obsolescence of the textbooks.

Because its authors update SCIENTIFIC AMERICAN *Medicine* every month, it is always current. Because the new information that affects patient management appears in a single source, there can be no loss through diffusion.

This 2,000-page, innovative union of publishing and electronic technology is the work of leading scholar-practitioners from Harvard and Stanford. The editors are Edward Rubenstein, M.D., F.A.C.P., and Daniel D. Federman, M.D., F.A.C.P.

Each month as authors update their contributions, revisions are entered on the magnetic tape on which the text and index are stored. The tape drives high-speed phototypesetting equipment so that subscribers receive about eight new chapters and a new index every four weeks; a

—Adjuvant chemotherapy for breast cancer utilizing CMF enhances survival among premenopausal women with axillary lymph node involvement.



—Posttransfusion hepatitis is usually caused by "non-A, non-B" viruses.

—Lidocaine fails to control ventricular arrhythmias in as many as 20 percent of acute M.I. patients. Procainamide, given by a bolus and continuous infusion method, is effective in most of these cases.

—Poorly ventilated recreational vehicles are a new, significant carbon monoxide poisoning hazard.

—Cerebral embolism, producing both transient ischemic attacks and permanent strokes, is a complication of mitral valve prolapse.

—Spectinomycin is effective for disseminated gonococcal disease caused by penicillinase-producing organisms.

—Adenine arabinoside has been licensed for treatment of herpes simplex encephalitis.

bulletin highlights the new developments.

New material replaces old material in the living text, so that the information is there — up-to-date, at your fingertips.

A CME program of eight patient management problems offered over a 12-month period is available at no extra cost. Authorities from first-line institutions create the problems under the direction of Stanford's Office of Postgraduate Medical Education and the Educational Testing Service.

As an organization accredited for continuing medical education, the Stanford University School of Medicine designates this continuing medical education activity as meeting the criteria for 32 credit hours in Category 1 for Educational Materials for the Physician's Recognition Award of the American Medical Association provided it has been completed according to instructions.

This program is approved by the American Academy of Family Physicians for 32 Elective Hours of CME credit.

Trial Offer

Please send us the card and you will receive over a period of 60 days at no cost, the two-volume text and two monthly updates. You may also take a CME test for credit. At the end of 60 days, if you decide to continue the subscription, we will bill you for $150 for the full 12 months; otherwise just return the volumes to us.

We urge you to subscribe now. We anticipate a price increase in January 1980 to $218 for the first year and $165 for the annual renewal as of January 1981. If you subscribe in 1979 for $150, you will be able to renew next year at $150, a saving of $83 over two years.

So please mail the card today and let us take the hassle out of keeping up.

SCIENTIFIC AMERICAN MEDICINE
415 Madison Avenue, New York, N.Y. 10017

Please enroll me as a subscriber to SCIENTIFIC AMERICAN *Medicine*. On receipt of this coupon you will send me the advanced two-volume text described in your announcement and update it regularly by sending me new monthly subsections. I understand that the annual subscription of $150 for SCIENTIFIC AMERICAN *Medicine* is tax-deductible. If I am not entirely satisfied, I may cancel at any time during the first 60 days, returning all materials for a complete refund.

- ☐ Please enter my subscription for SCIENTIFIC AMERICAN *Medicine*
- ☐ I shall also enroll in the CME Program
- ☐ I enclose a check made out to SCIENTIFIC AMERICAN *Medicine* for $150
- ☐ Please bill me

Please add sales tax for California, Illinois, Massachusetts, Michigan, Ohio and New York

Name _____

MD Specialty _____

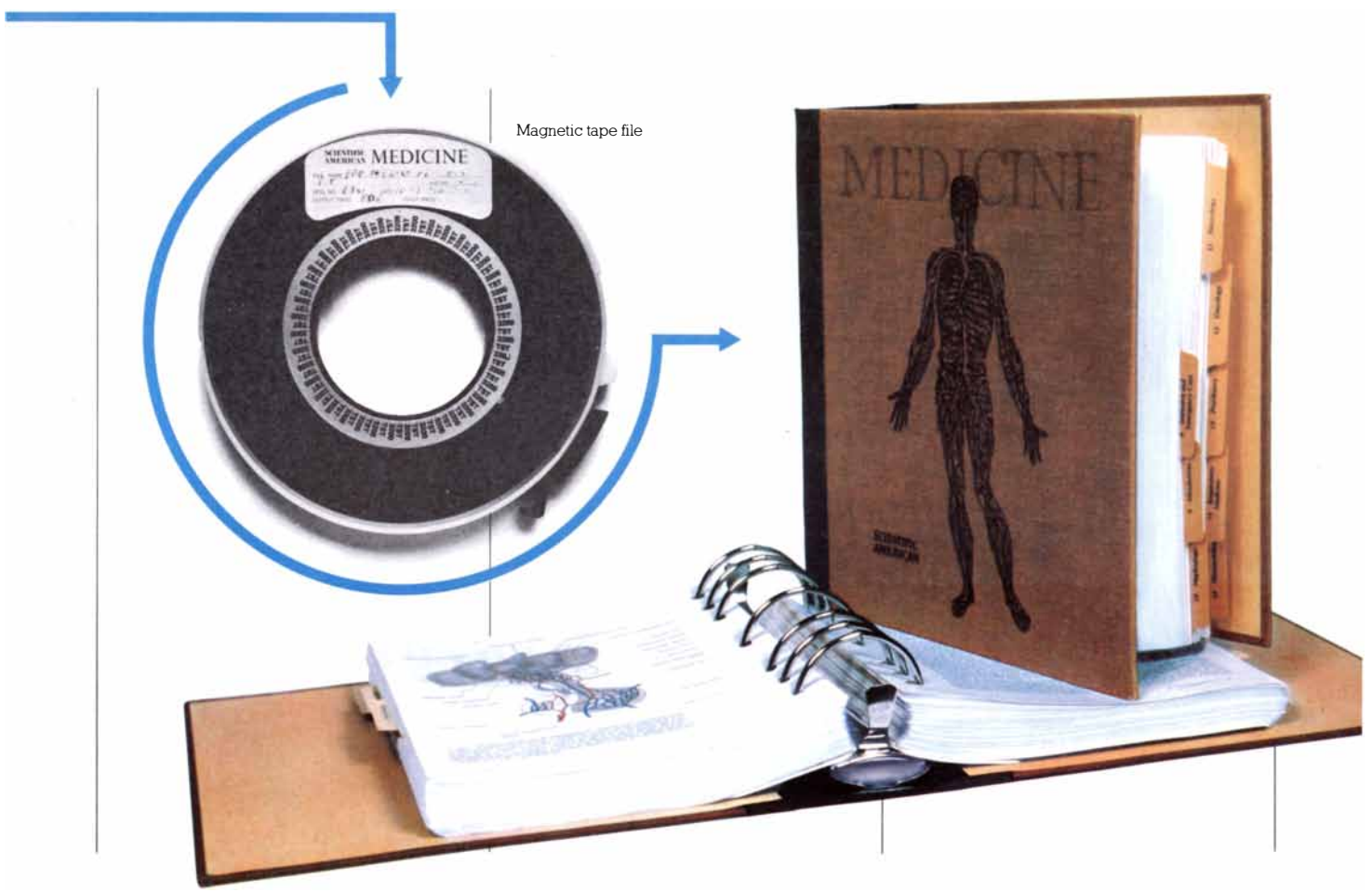
Address _____

City _____

State _____ Zip _____

Signature _____

Subscribers outside of the U.S. and the possessions will be charged extra for shipping and handling; updates will be sent by surface routes unless air mail delivery is requested. 1E



Working with one of the most modern fleets in the world is a source of pride for us.

Iberia's fleet —made up by Boeing 747's and 727's, Douglas DC-10's, DC-8's and DC-9's, Fokker F-27's and, soon, the Airbus- is one of the most modern commercial fleets of our times.

We are 5.440 engineers, technicians, specialists and other professionals working together to keep our fleet in top shape. With jobs that range from the simplest daily inspection, through the different periodic ones, to what we call the Major Overhaul which implies the dismantling and total renewal of the plane.

And, since each of those specialists knows his job inside out, we are proud to say that not even the smallest detail ever escapes us.

Keeping one of the most modern, better maintained fleets in the world is yet another way of showing our passengers that we care.

This is Iberia today.
But we want to be better.

IBERIA
INTERNATIONAL AIRLINES OF SPAIN

MORE THAN 50 YEARS MAKING FRIENDS.

A further step can now be taken. If the number of protons $N(A^+C^-)$ is equal to $N(A^+B^+C^-) + N(A^+B^-C^-)$, then it must be greater than or at least equal to $N(A^+B^-C^-)$. (The two sets will be equal if the B components of all the particles' spins happen to be minus, so that the subset $(A^+B^+C^-)$ is empty; otherwise $N(A^+C^-)$ will be larger. In other words, a part of the whole cannot be greater than the whole.) The same reasoning can be applied once again to prove that the number of protons with spin components B^-C^+ must be equal to the sum $N(A^+B^-C^+) + N(A^-B^-C^+)$ and hence that $N(B^-C^+)$ must be greater than or equal to $N(A^+B^-C^+)$.

Consider again the first equation derived above:

$$N(A^+B^-) = N(A^+B^-C^+) + N(A^+B^-C^-).$$

It has just been demonstrated that $N(B^-C^+)$ is greater than or at least equal to $N(A^+B^-C^+)$, which is the first term on the right side of the equation. It has also been shown that $N(A^+C^-)$ is greater than or equal to $N(A^+B^-C^-)$, which is the second term on the right side of the equation. It is therefore permissible to make the appropriate substitutions in the equation, changing the equals sign to one signifying "less than or equal to." The result is the inequality

$$N(A^+B^-) \leq N(A^+C^-) + N(B^-C^+).$$

Although this inequality is hereby formally derived, it cannot be tested directly by experiment because no instrument can independently measure two spin components of a single proton. The experiments under consideration, however, are carried out not on individual protons but on correlated pairs of them, and there is no need to make such impossible measurements. Suppose one proton in a pair is submitted to a measurement of its spin component along the A axis and is found to have the value A^+. No other measurements are carried out on this particle, but its singlet partner is tested for the component along the B axis and the result is found to be B^+. The latter measurement, which might be made at a distant site after the protons have been moving apart for some time, conveys additional information about the state of the first proton. To be explicit, the existence of a strict negative correlation implies that the first proton, which is already known by direct measurement to have the spin component A^+, must also have the component B^-.

By this means the observation of a pair of protons one of which has the spin component A^+ and the other the component B^+ can be employed as a signal indicating the existence of a single proton with the components A^+B^-. Furthermore, it can be demonstrated by a statistical argument that $n[A^+B^+]$, the number of such doubly positive pairs, must be proportional to $N(A^+B^-)$, the number of individual protons with the spin components A^+B^-. In the same way $n[A^+C^+]$ must be proportional to $N(A^+C^-)$ and $n[B^+C^+]$ must be proportional to $N(B^-C^+)$. The constant of proportionality in all three cases is the same. For single protons each of which is subjected to an imaginary double measurement an inequality has already been proved, showing that $N(A^+B^-)$ can be no greater than the sum of two terms: $N(A^+C^-) + N(B^-C^+)$. It is now possible to replace each of these unmeasurable quantities by the corresponding numbers of doubly positive proton pairs. The resulting expression is

$$n[A^+B^+] \leq n[A^+C^+] + n[B^+C^+].$$

This is the Bell inequality.

Of course the inequality is proved by this argument only if the three premises of local realistic theories are considered valid. Indeed, it is here that the premises have their most important application and ultimately their most questionable one. If the premises are granted, at least for the sake of argument, it should be clear that the Bell inequality must be satisfied. Moreover, the orientation of the axes A, B and C has nowhere been specified, so that the inequality should be valid regardless of what axes are chosen. The only possible violation of the inequality would result from a statistical fluke, where many particles with the spin components A^+ and B^+ happened to appear through random coincidence. The probability of such a coincidence approaches zero as the number of particles tested increases.

The Bell inequality constitutes an explicit prediction of the outcome of an experiment. The rules of quantum mechanics can be employed to predict the results of the same experiment. I shall not give the details of how the prediction is derived from the mathematical formalism of the quantum theory; it can be stated, however, that the procedure is completely explicit and is objective in the sense that anyone applying the rules correctly will get the same result. Surprisingly, the predictions of quantum mechanics differ from those of the local realistic theories. In particular, quantum mechanics predicts that for some choices of the axes A, B and C the Bell inequality is violated, so that there are more A^+B^+ pairs of protons than there are A^+C^+ and B^+C^+ pairs combined. Thus local realistic theories and quantum mechanics are in direct conflict.

The conflict raises two questions. First, what are the experimental facts of the situation? Is the Bell inequality satisfied or is it violated? Whatever the outcome of an experimental test there must be a flaw of some kind either in the rules of quantum mechanics or in local realistic theories. The second question there-

fore is: What premise underlying the refuted theory is at fault?

The thought experiment proposed in 1935 by Einstein, Podolsky and Rosen called for measurements of the position and momentum of particles. The experiment on spin components of protons was first discussed in 1952 by David Bohm of Birkbeck College in London, but still in the context of a thought experiment. It was not until 1969, after Bell had introduced his inequality, that real experiments exploring these questions were contemplated. The feasibility of such experiments was discussed by John F. Clauser of the University of California at Berkeley, R. A. Holt of the University of Western Ontario and Michael A. Horne and Abner Shimony of Boston University. They found that for a practical experiment the Bell inequality would have to be generalized somewhat, but a meaningful test of the alternative theories would still be possible.

The technical difficulty of the experiments should not pass unmentioned. In a thought experiment both protons of every pair always reach the instruments and the instruments themselves always yield an unambiguous measurement of the spin component along the chosen axis. Real apparatus cannot reproduce these results. The detectors are never perfectly efficient: many protons are simply not registered at all. Because of the imperfections of the instruments the number of protons counted in each category cannot be interpreted directly; instead an allowance must be made for the inefficiency of the detectors, which adds to the uncertainty of the results.

Of seven experiments reported since 1971, six have not concerned measurements of the spin components of protons but have instead measured the polarization of photons: the quanta of electromagnetic radiation. Polarization is the property of a photon that corresponds to the spin of a material particle. In one series of experiments atoms of a particular element and isotope were raised to an excited state by the absorption of laser light and then allowed to return to their original energy level in two steps. At each step a photon with a characteristic energy or wavelength was emitted. The photons moved off in opposite directions, and they had opposite polarizations. In other words, if the polarization of both photons was measured along any single direction, a strict negative correlation was observed.

The differences between ideal instruments and real ones are quite plain in these experiments. There is no single device that can intercept a photon and report directly on its polarization. Instead two devices are necessary, a filter and a detector. The filter is designed to allow the passage of those photons that have the selected polarization and to stop or deflect all others; the detector counts the number of photons that pass through the filter. Neither of these components is perfect, so that the failure to register a photon does not necessarily mean that it had the wrong polarization.

Experiments have also been done on the polarization of gamma rays, which are high-energy photons. The gamma rays were created by the mutual annihilation of electrons and their antiparticles, positrons. Such an annihilation gives rise to two gamma rays, which are emitted in opposite directions and have opposite polarization. The experiments are therefore formally equivalent to the atomic ones, but the apparatus required is quite different. In general detectors are more efficient for high-energy photons, but polarization filters are more efficient for low-energy ones.

One experiment has measured the correlations of spin components of protons and therefore closely resembles the original thought experiment. The pairs of protons are created by injecting protons of comparatively low energy into a target made up partly of hydrogen atoms. The nucleus of a hydrogen atom consists of a single proton. When an incident proton strikes a hydrogen nucle-

EXPERIMENT DATE	PARTICLES STUDIED	RESULTS
Stuart J. Freedman and John F. Clauser, University of California at Berkeley 1972	Low-energy photons emitted during transitions in calcium atoms.	In agreement with quantum mechanics.
R. A. Holt and F. M. Pipkin, Harvard University 1973	Low-energy photons emitted during transitions in atoms of mercury 198.	In agreement with Bell inequality.
John F. Clauser, University of California at Berkeley 1976	Low-energy photons emitted during transitions in atoms of mercury 202.	In agreement with quantum mechanics.
Edward S. Fry and Randall C. Thompson, Texas A. & M. University 1976	Low-energy photons emitted during transitions in atoms of mercury 200.	In agreement with quantum mechanics.
G. Faraci, S. Gutkowski, S. Notarrigo and A. R. Pennisi, University of Catania 1974	High-energy photons (gamma rays) from annihilation of electrons and positrons.	In agreement with Bell inequality.
L. Kasday, J. Ullman and C. S. Wu, Columbia University 1975	High-energy photons (gamma rays) from annihilation of electrons and positrons.	In agreement with quantum mechanics.
M. Lamehi-Rachti and W. Mittig, Saclay Nuclear Research Center 1976	Pairs of protons in the singlet state.	In agreement with quantum mechanics.

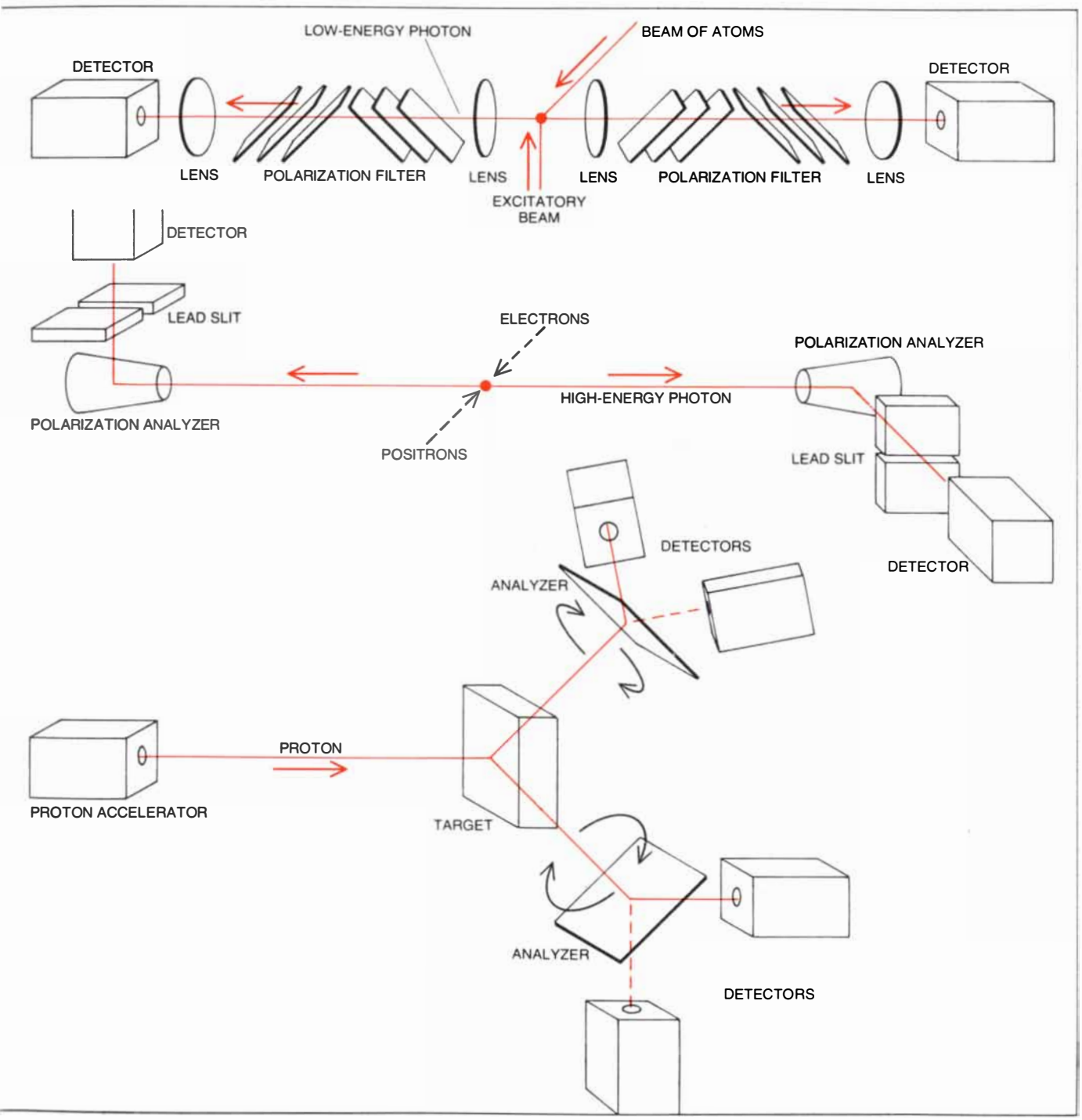
REAL TESTS OF THE BELL INEQUALITY have been carried out by seven groups of investigators. Only one of the experiments measured the spin components of protons; the others studied the polarization of photons, or quanta of electromagnetic radiation. In four experiments pairs of low-energy photons with opposite polarization were emitted by atoms that had been raised to an excited state. Pairs of oppositely polarized gamma rays, or high-energy photons, were created in two other experiments by the mutual annihilation of electrons and their antiparticles, positrons. In the remaining experiment protons from a particle accelerator struck a target made up partly of hydrogen; the accelerated protons and the hydrogen nuclei formed pairs in the singlet state. Five of the experiments gave results in violation of the Bell inequality and in agreement with quantum mechanics. That the Bell inequality is violated is now generally accepted. The cause of the discrepancy in the results of the other two experiments is uncertain.

us, the two protons interact briefly and enter the singlet state. Both then leave the target, sharing the momentum of the incident proton, but if they are undisturbed, they remain in the singlet state. Preliminary measurements of the same spin component on both protons give opposite results.

The instruments for an experiment with proton pairs again consist of filters and detectors. In the one experiment that has been completed the filter was a carbon foil, which scattered each proton into one of two detectors depending on the value of the measured component.

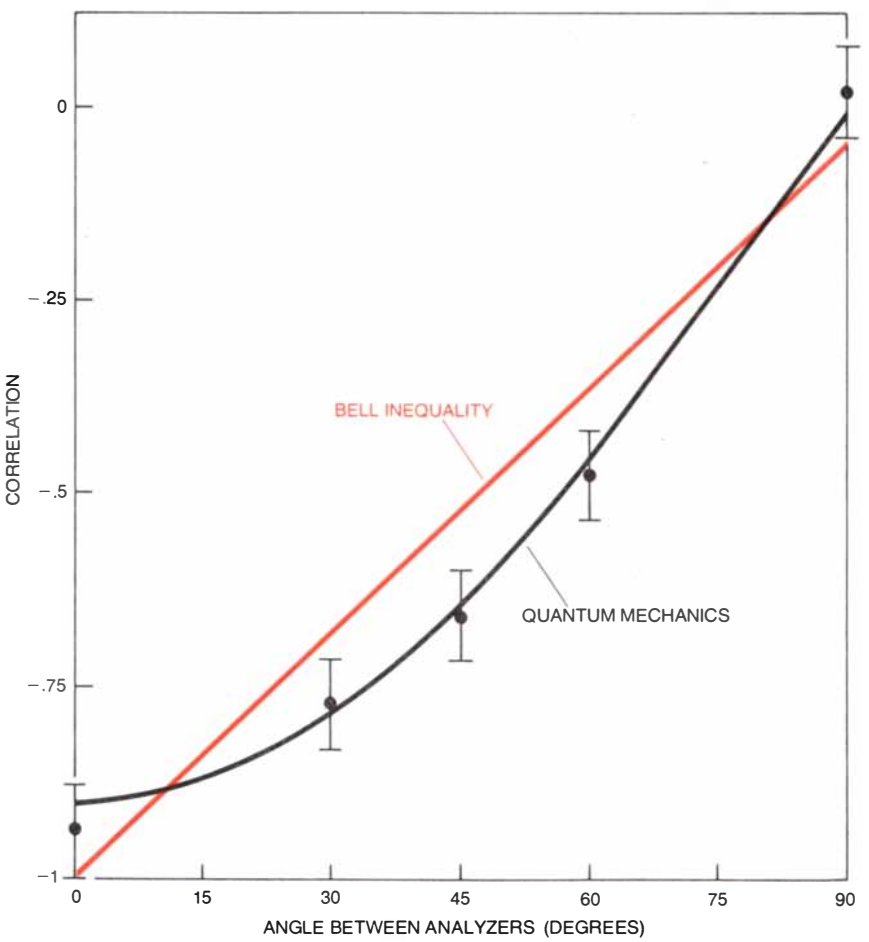
Regardless of what particles are being studied, the experiment consists of three series of double measurements. Three axes, *A*, *B* and *C*, are selected; in general the angles between them are set to the values where the maximum discrepancy between quantum mechanics and local realistic theories is expected. One filter is then set to admit particles with the polarization or spin component A^+ and the other is set to pass particles with the component B^+. After a large enough sample of particles has been recorded in this configuration the filters are rotated to measure the components along axes *A* and *C* and further data are recorded. Finally the filters are reoriented again to axes *B* and *C*. The coincidences recorded in each configuration are counted and corrections are made for the inefficiency of the apparatus. It is then a matter of simple addition to compare the results with the Bell inequality.

Of the seven completed experiments five endorse the predictions of quantum mechanics, that is, they indicate a violation of the Bell inequality for some choices of the axes *A*, *B* and *C*. The other two give correlations no greater than those allowed by the Bell inequality and



therefore support local realistic theories. The score is thus five to two in favor of quantum mechanics. Actually the support for quantum mechanics is much stronger than this ratio would seem to imply. One reason for attributing greater credibility to the five experiments that violate the Bell inequality is that they represent a larger sample of data and are therefore statistically more significant. Some of those experiments were done after the two anomalous results were reported and included refinements in the instrumentation designed explicitly to avoid any biases that might account for the two discrepant results. Clauser and Shimony have pointed out that there is also an epistemological justification for disregarding the two experiments that are in disagreement with the majority. Quantum mechanics predicts a larger correlation between events and local realistic theories predict a smaller one. A great variety of systematic flaws in the design of an experiment could destroy the evidence of a real correlation, yielding results within the limit set by the Bell inequality. On the other hand, it is hard to imagine an experimental error that could create a false correlation in five independent experiments. What is more, the results of those experiments not only violate the Bell inequality but also violate it precisely as quantum mechanics predicts. For the results of the five experiments to be produced by random coincidence would require an extraordinary statistical fluke that is not credible given the number of particles that have now been detected.

Further tests of the Bell inequality are under consideration, and at least one additional experiment is already in preparation. Most physicists concerned with these problems, however, have substantial confidence, based on the five consistent results, that the issue has already been decided. For some choices of the axes *A*, *B* and *C* the Bell inequality is violated in nature, and local realistic theories are therefore false.



RESULTS OF AN EXPERIMENTAL TEST of the Bell inequality show that it is clearly violated. The experiment is the one that employed pairs of protons in the singlet state, which was carried out by M. Lamehi-Rachti and W. Mittig of the Saclay Nuclear Research Center in France. The negative correlation between the values of different spin components is given as a function of the angle between the settings of the two analyzers. A correlation of −1 would indicate that the components invariably had opposite values. The Bell inequality states that the correlation at any angle must be on or above the colored line. The observed correlations at 30, 45 and 60 degrees are below the line. The results not only violate the Bell inequality but also are in good agreement with the predictions of quantum mechanics, which fact adds to their credibility. The violation of the Bell inequality implies that at least one of the three premises of local realistic theories must be false; Einstein separability is considered the most plausible candidate.

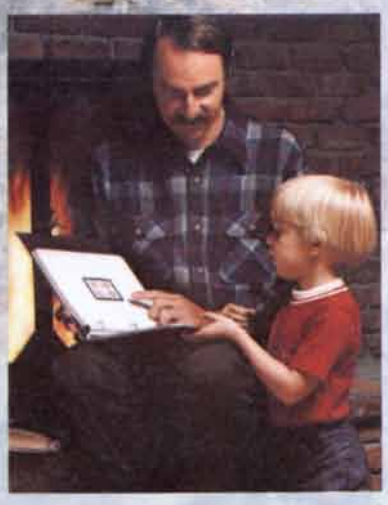
If it can be considered as having been demonstrated that local realistic theories are in error, which of the three premises underlying those theories is to blame? A first step in answering this question should be to make sure no additional assumptions were made in formulating the experimental test.

As it happens, at least one subsidiary assumption was needed. Because of the limitations of practical instruments, it was necessary to generalize the Bell inequality slightly, and that generalization must be assumed to be valid; it cannot be proved. It seems most unlikely, however, that this circumstance could alter the phenomena in such a way that the results of the experiments not only would violate the Bell inequality but also would be consistent with the predictions of quantum mechanics. In any case it is possible more refined experiments will test the inequality without the generalization. Because the subsidiary assumption is susceptible to an experimental test it seems less fundamental than the other three, and so it will not be considered further here.

Another area that might be scrutinized for unacknowledged assumptions is the proof of the Bell inequality. Indeed, it seems the proof does depend on the assumed validity of ordinary, two-valued logic, where a proposition must be either true or false and a spin component must be either plus or minus. Some interpretations of quantum mechanics have introduced the idea of a many-valued logic, but those proposals have nothing to do with the reasoning applied in this proof. Indeed, in the context of the proof it is difficult even to conceive of an alternative to two-valued logic. Unless such a system is formulated it seems best to pass over the problem.

The entire series of experiments founded on the ideas of Einstein, Podolsky and Rosen is sometimes regarded as merely a test of hidden-parameter theories. The experiments do indeed test those theories, but it should be emphasized that the existence of hidden parameters is not an additional premise of local realistic theories. On the contrary, the existence of parameters specifying the deterministic properties of a particle was derived from the three original assumptions. Remember that the psychologist did not assume that his invented test measured any real attribute of the tested subjects; instead he deduced the existence of such an attribute after observing a strict correlation. In the same way the existence of hidden parameters was derived from the negative correlation detected when a single spin compo-

George Wendt, rapids-shooting philatelist.



George Wendt, white water expedition guide, has found something exciting to do on dry land. He collects stamps.

George finds that collecting U.S. Commemoratives lets him explore the thrilling panorama of America. He discovers new places and things, meets fascinating people and relives exciting events in our history.

You can start the adventure of stamp collecting with U.S. Commemoratives. It's easy and affordable. Every few weeks, there are new, beautifully designed issues at your local Post Office. You'll also find an informative, colorful guidebook called *Stamps & Stories* to help you get your feet wet. And the new 1980 Olympic Stamps are an exciting place to begin.

"Collecting U.S. Commemoratives is a special kind of thrill," says George. "And it's one adventure I can share with my family."

Olympic Summer Events (Available now.)

U.S. Postal Service®

Start the adventure of stamp collecting with U.S. Commemoratives.

nent was measured on pairs of protons in the singlet state.

It is probably not possible to prove rigorously that no other supplementary assumptions enter into the argument supporting the local realistic theories. The chain of reasoning is simple enough, however, that if other assumptions are implicit in it, they should be easily recognized. None has yet been pointed out. It therefore seems that attention must be focused on the three premises of realism, the free use of induction and Einstein separability.

Of the three premises realism is the most fundamental. Realism can be stated formally as the belief that a mere description of data is not all that should be required of a theory. Even an empirical rule for predicting the patterns of future measurements is not enough. The mind demands something more: not necessarily determinism—there is nothing intrinsically irrational about randomness—but at least objective explanations of observed regularities, or in other words causes. Underlying this demand is the intuitive notion that the world outside the self is real and has at least some properties that exist independently of human consciousness.

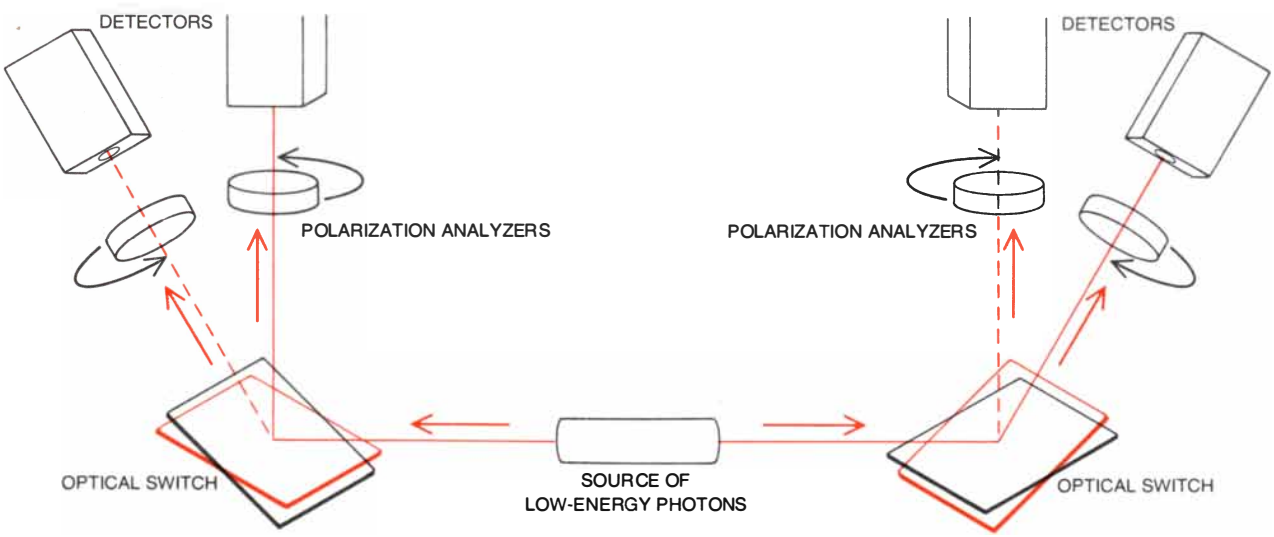
A number of philosophers, who can collectively be called positivists, have rejected the realistic viewpoint. The positivists do not assert that the world external to the mind does not exist; they merely dismiss as meaningless any statement about an external reality that does not refer directly to sensory impressions. In the 20th century some radical positivists have had an appreciable, if indirect, influence on the thinking of theoretical physicists.

The sense of paradox induced by the finding that the Bell inequality is violated can certainly be alleviated by adopting a positivist attitude, and such a course of action was first proposed long ago. When all the consequences of abandoning realism are considered, however, it is too great a renunciation to have much appeal. In the context of this experiment positivism asserts that it would be meaningless to attribute anything resembling a definite spin component to a particle before the component is measured; that the only quantity with any verifiable reality is the observation itself, the sensory impression; and that the psychologist's demand for an objective explanation of the remarkable correlation he observes should ultimately be rejected. If this refusal to seek underlying causes of observed regularities is applied consistently, it trivializes the entire scientific enterprise. Science is reduced to a set of recipes for predicting future observations from a knowledge of past ones. Any notion of science as "the study of nature" is impossible; nature is a phantom. One can imagine a physics grounded on positivist principles that would predict all possible correlations of events and still leave the world totally incomprehensible. Given the extreme consequences of abolishing realism, one is inclined to cling to this first premise.

Realism enters the argument supporting local realistic theories at another point: it is the justification for postulating the free use of induction. It is induction that enabled the physicist to extrapolate from a series of observed negative correlations to the conclusion that any two protons in the singlet state have opposite values of any single spin component, even if none of the components is measured. The extrapolation was an essential step in the proof of the Bell inequality, but it is clearly insupportable if the concept of unmeasured properties has no meaning.

This use of induction might be regarded by some as a weak link in the chain of argument. Shortly after the paper by Einstein, Podolsky and Rosen appeared, Niels Bohr published a reply in which he defended the completeness of the quantum-mechanical description of nature; the basis of his criticism was that Einstein's use of induction was unwarranted. Bohr's reply is a central document in what has come to be known as the Copenhagen interpretation of quantum mechanics. His reasoning amounts to an argument that a particle and an instrument adjusted to make a specific measurement on it constitute in some respects a single system, which would be altered in an essential way if the setting of the instrument were changed. For this reason it is not allowable to make any inferences about the state of a particle without specifying at the same time the settings of the instruments that will interact with the particle.

Bohr's views have been widely influential, and in a sense rightly so; after all, the recent work under discussion here has shown that in these matters he was



EINSTEIN SEPARABILITY will be tested rigorously in an experiment now being prepared by Alain Aspect of the Optics Institute of the University of Paris. Earlier experiments tested only the less restrictive separability principle: the settings of the analyzers were determined well in advance, so that some influence of one measurement could be communicated (by an unknown mechanism) to the other measurement at a speed well below the velocity of light. This possible explanation of the observed correlation is extremely unlikely, but it would be excluded entirely if the settings of the analyzers were changed so quickly that a signal moving no faster than light could not pass from one detector to the other in time to influence the result of the second measurement. In Aspect's experiment, which will measure the polarization of low-energy photons, this condition will be met. Two sets of analyzers and detectors will be provided for each photon, and the analyzers will measure different components. A fast optical switch will determine which analyzer the photon enters only when it is too late for the decision to influence the other measurement (assuming that the hypothetical influence propagates no faster than light). The switch is shown as a moving mirror; actually the switching will be accomplished by ultrasonic waves on the surface of a crystal.

closer to the truth than Einstein was. Nevertheless, when Bohr's ideas are considered in their essence, they are subject to objections much like those that were raised against a retreat to positivism. Because realism provides the ultimate rationale for the free use of induction, it can be argued that Bohr was not a realist, or at least not a consistent one. Any explanation of the distant-correlation experiments that relies on Bohr's reply to Einstein, Podolsky and Rosen may turn out to be inconsistent with even a moderate version of realism.

If realism and the free use of induction are to be retained, the violation of the Bell inequality can be explained only by giving up the assumption of Einstein separability. In the psychological experiment separability was understood to imply that the husbands and the wives, once they were separated, could not communicate with each other. In the physics experiment the separability assumption expressed the intuitively reasonable idea that the spin components of one proton have no influence over those of the other proton, provided the two particles are far apart. The more restrictive assumption of Einstein separability forbids such an influence only if it would have to propagate with a speed greater than the speed of light. As I have shown, this assumption must now be regarded as highly questionable.

Before considering the consequences of this conclusion it should be pointed out that none of the experiments completed so far has rigorously tested the assumption of Einstein separability. In those experiments the settings of the instruments were determined well in advance (on the time scale of particle physics). Therefore the setting of one instrument could conceivably affect events observed at the other instrument, or it could modify hidden parameters at the source of the proton pairs; in either case there would be no need for the influence to travel faster than light. An experiment with instruments whose setting can be changed rapidly could exclude this possibility. The decision to measure a certain spin component with one detector would not be made until it was too late for any influence of that decision to reach the other instrument or the source, even at the speed of light, in time to alter the outcome of the second measurement. Such an experiment is now being done by Alain Aspect of the Optics Institute of the University of Paris.

Quite apart from the question of how fast a hypothetical influence could travel from one instrument to another, the influence itself seems extremely implausible. It would be required to alter the distant observations in precisely the manner needed to produce the observed violation of the Bell inequality. Hence it seems best to search for some other ex-

Now available by mail . . .
The perfect gift for **anyone** — including yourself!

The Everything Set

Completely portable (AC/DC)
. . . amazingly complete!
Even receives CB!

- **Watch TV — VHF and UHF**
- **Enjoy AM and FM radio**
- **Play tape cassettes or record them via the built-in mike or right off the air (radio or TV)**
- **Eavesdrop on CB — all channels**
- **Get weather reports — instantly**
- **Listen in on public service bands — police, fire, marine, etc.**
- **Hear pilots talk to air controllers**
- **Even use it as a P.A. system**
- **Play tape cassettes or record them via the built-in mike or right off the air (radio or TV)**

And note the tape cassette at the bottom of the photo. We put it there to help indicate how incredibly compact The Everything Set really is — only 15½"W x 10"D x 5½"H.

Never before has so much been offered in so compact a unit. A 100% transistorized/IC chassis — plus superb electronic design — makes it possible, without sacrificing features or performance.

IT'S SMALL, BUT BIG

The set is small enough to fit, without intruding, on a corner of your desk...on your kitchen table...or on your night stand. It weighs just 12½ lbs.

Yet the black-and-white TV picture is big enough to see without squinting, 3"when measured diagonally, big enough — and sharp enough — for you to read the fine type on commercials and the credits on movies.

The sound is also big — thanks to a 5" heavy-duty speaker. And there are all the controls you need for perfect pictures — up-front knobs for channel selection (VHF 2-13 and UHF 14-83) and contrast...plus controls for brightness, vertical hold and horizontal hold.

The collapsible antenna extends to nearly 3 feet, and boasts a deluxe swivel/tilt mount to help you get the best reception. There's also provision for an external antenna.

AM...FM...PSB...CB

The Everything Set brings you four diverse bands for a smorgasbord of entertainment, information and eavesdropping.

There's the AM band (525-1605 KHz) and the FM band (88-108 MHz). They share a circular, easy-to-read tuning dial that lights up at the touch of a button, as does the separate dial for VHF (112-174 MHz) and CB (channels 1-40). A separate, built-in meter indicates maximum station reception.

On VHF, the aviation band extends from 118-136 MHz; you can hear pilots talking to airport controllers. Tune a bit higher, and listen in on the public service band: police, fire, marine, etc. Weather forecasts are broadcast continuously at the high end of the band.

Switch to CB, and eavesdrop on any of the 40 channels. (You can receive, but not transmit.) A squelch control reduces or eliminates unwanted noise.

A TAPE RECORDER, TOO

Want to record right off the air — radio music, TV audio, sports broadcasts? Just tune in the program you want, then press the Play and Record buttons of the built-in tape recorder. Want to make a live recording? Use the built-in mike...or plug in an external mike (not included).

The Everything set accepts standard tape cassettes, and has just about every feature you'd want. There's a tape index counter (with reset button), so you can find any section of a tape quickly and easily. A variable sound monitor lets you control the speaker volume without affecting the recording level. (Useful when recording late at night.) And other features include fast forward, fast rewind, pause, cue, review, auto-stop, stop/eject, and automatic level control.

There's even a control to fade recordings in and out, to combine off-the-air recordings with input from an external mike, or to use The Everything Set as a portable public address system (indoors or outdoors).

FOR SLEEPYHEADS

A built-in timer works with the auto-stop to turn off the radio, TV or tape player after 15, 30 or 45 minutes, depending on whether you load the tape player with a C-30, C-60 or C-90 cassette. So you can doze off confident the unit will shut itself off.

PLUS, COMPLETE PORTABILITY

The Everything Set operates on any of four types of power, so you can use it anywhere — in your home or office...in your car, van, truck or boat...at a picnic...on a camping trip...wherever. Or send it off to college.

Most economical is AC (120V) power, using the 6-ft. cord provided. Also provided is an 8-ft. car cord that plugs into a cigarette lighter socket for 12V operation.

For complete portability, slip 6 alkaline D cells (not included) into the built-in battery compartment. Better yet, get 6 rechargeable D cells and a low-cost recharger. (G.E. makes one.) Or order the optional, rechargeable 6V Ni-Cad battery pack from us.

90-DAY LIMITED WARRANTY

The Everything Set is designed and manufactured by the Sampo Corporation of America, a pioneer in electronics since 1936. Their limited warranty covers the entire set for 90 days and the picture tube for 1 year. They have service centers throughout the U.S.

IT'S THE PERFECT GIFT

Just in time for the holiday season, we're making The Everything Set available by mail...for only **$249.95,** plus $6 for shipping/handling. This includes the AC cord, the 12V car cord, a collapsible sun shield for the TV screen, and a blank tape cassette.

If you order two sets at a time (keep one, give one) our price is even less — only **$229.95** (plus $6 shipping) for each.

The rechargeable 6V battery pack (for 6 hours of continuous TV operation) is yours for only **$49.95** plus $3 shipping/handling.

(New Jersey residents: add 5% sales tax).

MONEY BACK GUARANTEE

Try The Everything Set for 15 days. If not delighted, return it (insured) for a prompt refund.

DON'T MISS OUT — ORDER NOW

Inventories may be tight for the holiday season, so order now. Send us your check or money order...or charge it to your VISA, Master Charge, American Express or Carte Blanche credit card. Please include your name, address, signature, type of credit card, account number and expiration date.

For fastest service on credit card orders
Phone us Toll-free: **800-257-7850.**
In N.J. call **1-800-322-8650.**
Call any time
24 hours a day — 7 days a week.

Or mail your order to:

mercury
INTERNATIONAL SALES GROUP

Dept. P,Lakewood Plaza
Lakewood, New Jersey 08701

planation and to assume, pending the results of Aspect's experiment, that if ordinary separability is violated, Einstein separability will be violated too.

I have discussed a pair of protons as if they were independent entities that come together in the target and then move apart again. They can also be regarded as the elements of a single physical system that is created during the first interaction and becomes progressively more extended in space until it is disrupted by the first measurement. With respect to separability these descriptions are equivalent. In each case a violation of Einstein separability requires instantaneous action at a distance, either between independent systems or within a single extended system.

Must the principle of the finite propagation of signals therefore be abandoned? To that question no rash answer should be given. The principle was introduced as a premise of the theory of relativity, which cannot be made consistent without it. Moreover, signals that outrace light give rise to bizarre paradoxes of causality in which observers in some frames of reference find that one event is "caused" by another that has not yet happened. It turns out, however, that the instantaneous influences that seem to be at work in the distant-correlation experiments do not require such a drastic revision of accepted ideas. It seems quite certain these influences could not be employed to transmit any "useful" information, such as orders or instructions. No event that causes another event can be linked to it through this mechanism; the instantaneous influences can pass only between events that are related by a common cause. Hence the concept of a signal could be redefined in such a way that only those means of communication that transmit useful information would be called signals. The principle of the finite velocity of signals would then be preserved.

Even this solution impairs scientific realism to some extent. The basic law that signals cannot travel faster than light is demoted from a property of external reality to a feature of mere communicable human experience. Although this represents a step toward philosophical positivism, the concept of an independent or external reality can still be retained as a possible explanation of observed regularities in experiments. It is necessary, however, that the violation of Einstein separability be included as a property, albeit a well-hidden and counterintuitive property, of that independent reality. It should be noted in passing that Bohr's refutation of Einstein's argument for hidden parameters introduces an implicit violation of separability. It is founded on a strange indivisibility of the system of particles and the instruments of observation.

The argument that proceeds from an

observed correlation to the Bell inequality to the violation of Einstein separability is not particularly complicated, but it is indirect. Could the same result have been obtained in some more straightforward way? As it happens, it could not have been demonstrated without the Bell inequality, but it could have been suspected, and in fact it was. The suspicion arose from the fact that the wave function for a system of two or more particles is generally a nonlocal entity, which is considered to collapse suddenly or even instantaneously when a measurement is made. If the wave function is regarded as a kind of bizarre real jelly, the instantaneous collapse obviously violates Einstein separability. This naive argument was never taken very seriously, however, because the conventional interpretation of quantum mechanics does not identify the wave function of a system with whatever is meant by the reality of the system. Bohr, for example, considered the wave function a mere tool for doing calculations. Besides, the wave function for a system of several particles describes them only in an approximation that ignores the theory of relativity, and so its structure hardly seems a reliable argument against Einstein separability. For these reasons it was possible until a few years ago to believe in an independent, external reality and simultaneously to regard Einstein separability as a completely general law bearing on that reality.

One conceivable response to the distant-correlation experiments is that their outcome is inconsequential. The experiments themselves might represent a rare and therefore interesting test of quantum-mechanical phenomena observed at long range, but the results are merely what was expected. They show that the theory is in agreement with experiment and so provide no new information. Such a reaction would be highly superficial. It is indeed true that the experiments, now that they have been completed, have turned out to have little to do with quantum mechanics. That does not make them trivial; rather, it indicates that their real bearing is elsewhere. A discovery that discredits a basic assumption about the structure of the world, an assumption long held and seldom questioned, is anything but trivial. It is a welcome illumination.

Most particles or aggregates of particles that are ordinarily regarded as separate objects have interacted at some time in the past with other objects. The violation of separability seems to imply that in some sense all these objects constitute an indivisible whole. Perhaps in such a world the concept of an independently existing reality can retain some meaning, but it will be an altered meaning and one remote from everyday experience.

A Neolithic and Iron Age Site on a Hilltop in Southern England

The Celts encountered by the Romans when they invaded Britain had built enormous fortifications on hilltops. On Crickley Hill such a fort is superposed on enigmatic works preceding the Celts by 2,000 years

by P. W. Dixon

When the Romans invaded Britain in A.D. 43, they encountered a flourishing Celtic society. The most characteristic and impressive monuments left by the Celts are their imposing hilltop fortifications, with an extent of as much as several hundred acres. Most of the fortifications are now marked only by the ruined and grass-grown remains of defensive banks and ditches. The earliest of these works are currently thought to date back to late in the second millennium B.C.; the latest are thought to date to the end of the first millennium B.C. Whatever the function of the hill forts was in the communities from which they sprang, their construction represents such an immense expenditure of skill, labor and raw materials that an analysis of them is essential to any understanding of the structure and the cultural and economic development of society in late prehistoric Europe. The hill forts clearly imply that in addition to the peaceful pursuits of the time—agriculture, religious observance, manufacture and trade—there was a substantial element of migratory and military activity.

The remains of more than 1,000 hill forts are known in southern England. Nondestructive field work—contour surveying, collecting the chance finds thrown up by agriculture or erosion, measuring anomalies in the soil with geophysical instruments—have yielded information about the position, size and shape of many fortifications, but the surface evidence of grass-covered mounds is insufficient to chart the development of the more complex sites. Archaeologists have therefore devoted much labor to the excavation of hundreds of hill forts, with varying degrees of success.

The study of hill-fort settlements has suffered from an understandable tendency for excavators to concentrate on the upstanding ramparts, particularly the visible entrances; there a modest amount of work is likely to yield a substantial amount of information about how the structure evolved. A similar tendency has been to focus attention on the largest and strongest sites; these are likely to have been the most significant, at least in the politics of the time. Even the largest excavation, however, cannot deal adequately with an area of 30 or 40 acres, and the unrelieved grass or scrubland interior of many of the hill forts in southern Britain suggests to the prospecting archaeologist an unprofitable digging terrain. As a result much is known about the defenses of hill forts and little about the settlements they enclose. What was needed was a hill fort of modest size, one whose total excavation by hand might be achieved in the lifetime of a single director and whose remains had not been damaged by farming or earlier excavations. In 1968 I found such a site in the Gloucestershire Cotswolds, overlooking the Severn valley four miles south of Cheltenham and six miles east of Gloucester. It is named Crickley Hill.

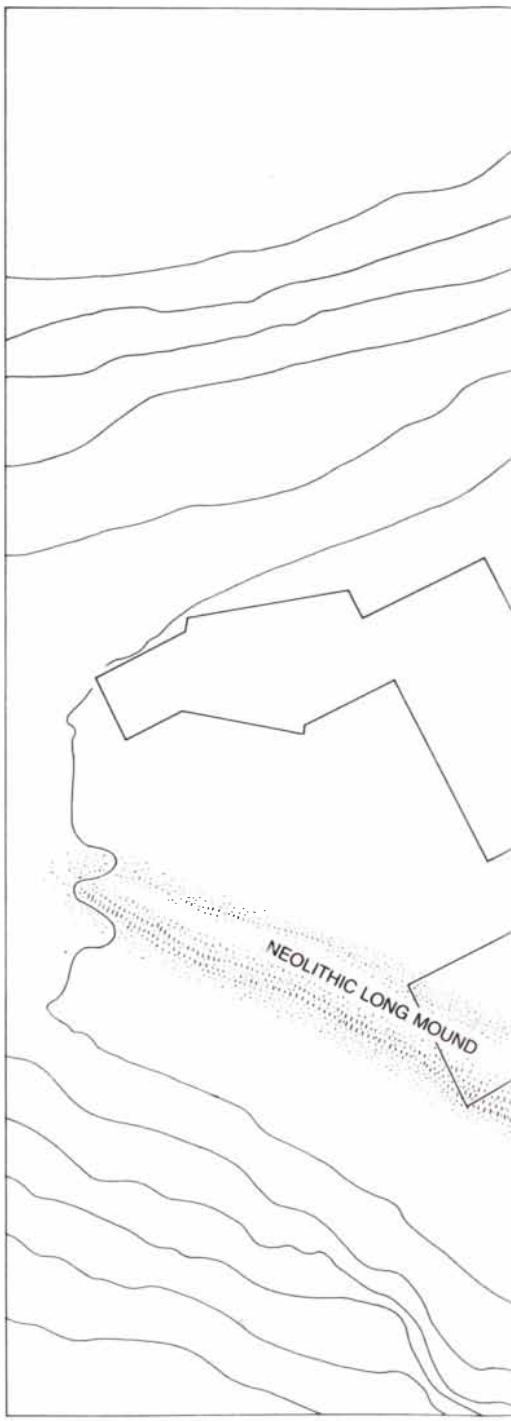
The steep western slope of the Cotswold limestone is divided by small valleys into flat-topped, spurlike hills. Many of the hilltops were fortified by banks and ditches. The outermost rampart at Crickley Hill, 300 meters long, cuts off a triangular area of about seven acres. Here in the summer of each year since 1969 volunteer diggers and supervisors working under my direction have excavated nearly half of the hilltop. Our efforts have unearthed signs of prehistoric activity of unexpected complexity and duration. We have discovered that Crickley Hill was the site not only of two successive Celtic Iron Age hill forts but also of two pre-Celtic Neolithic enclosures.

The top of Crickley Hill is not completely flat; at the center of the fort stands a small knoll that aerial photographs showed was enclosed by an eroded bank and ditch. In this area we found five successive phases of Neolithic occupation. The remains of the earliest and smallest structure were on the highest part of the knoll. There a series of small pits formed a roughly oval ring about eight meters long. The spoil from the pits had been piled into a mound in the center; the gaps between the pits had been cleared of grass and topsoil, leaving the bedrock exposed. Why this was done is not known, but the finished work must have resembled a tiny burial barrow. Later most of the mound was removed and the pits were filled with clean material; our own excavations emptied them of nearly 10 tons of stones and earth (but uncovered only one small piece of bone). This structure had been buried under the bank of the first Neolithic enclosure, a double line of ditches cutting across the triangular spur of the flat-topped hill.

In southern England about 50 Neolithic enclosures are known from excavations or are suspected from aerial photographs. Called "causewayed camps" because their encircling ditches are crossed by numerous causeways, they generally consist of two or more concentric rings. When they were first discovered early in this century, they

AERIAL VIEW of the central region of the Iron Age settlement on Crickley Hill shows the evidence of two distinct phases of occupation. The ring of excavated postholes near the center held the supports of the largest round house of the second hill fort. The postholes of the rectangular houses of the earlier hill fort are also visible. Parallel piles of turf at the top of excavated area mark the entrance of both forts. At the lower right some of the volunteer diggers rest.

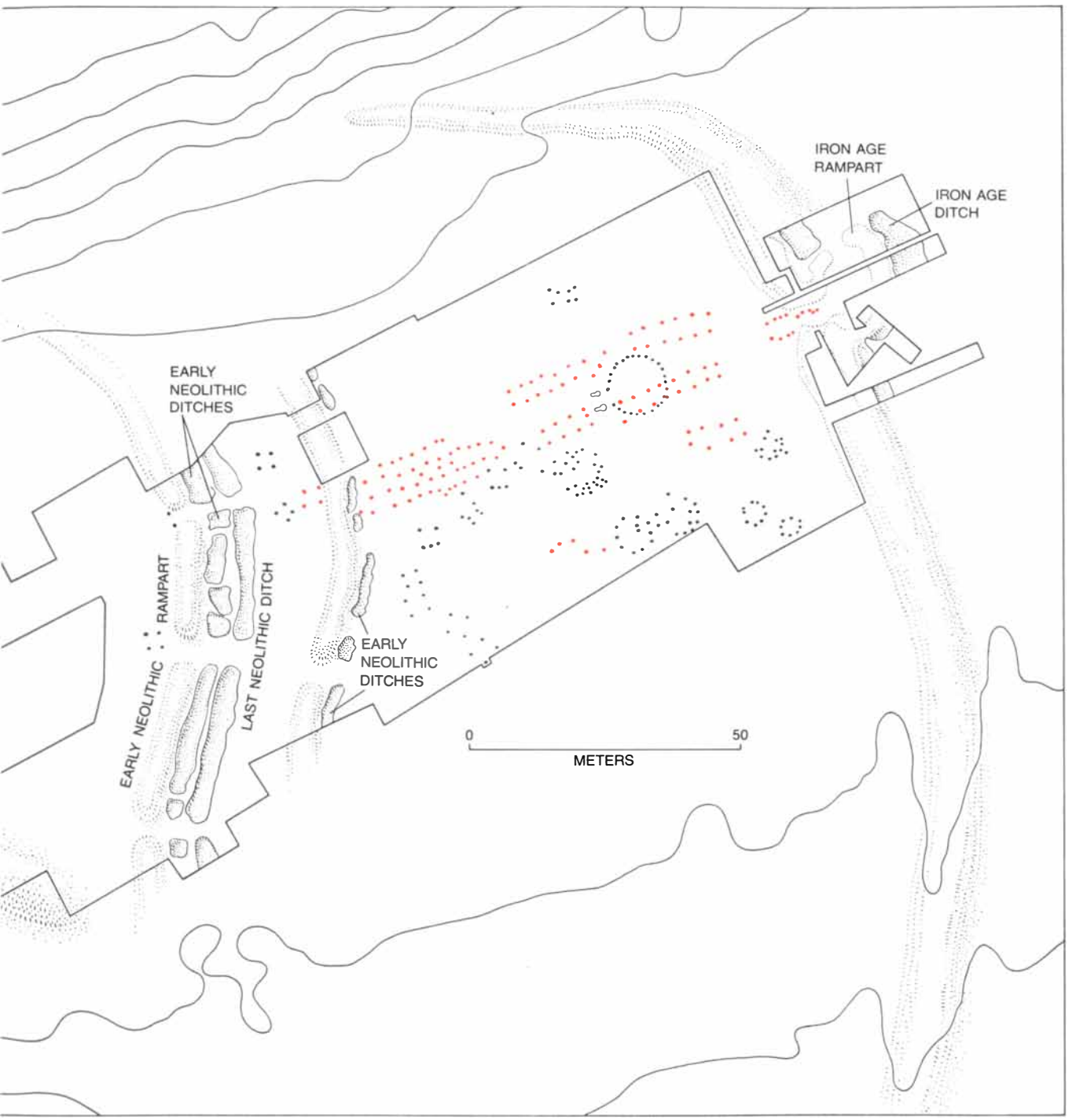
were thought to be the defenses of settlements and were also known as Neolithic towns. Excavations in the 1930's at Windmill Hill in Wiltshire raised doubts, because no remains of buildings were found inside the enclosure. Although the weathering of the chalk bed-



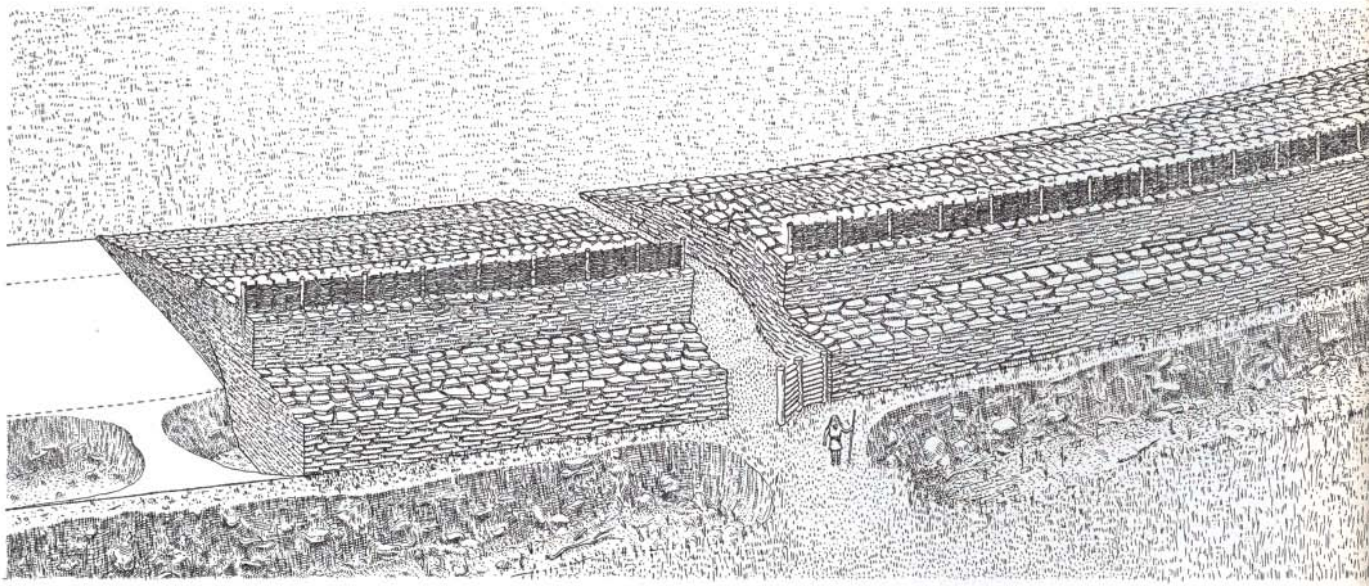
THREE OCCUPATION PHASES are represented in this plan of Crickley Hill. At the left, near the highest point of the hill, is the last Neolithic enclosure, built in about 2500 B.C. It had two long and narrow entrance pas-

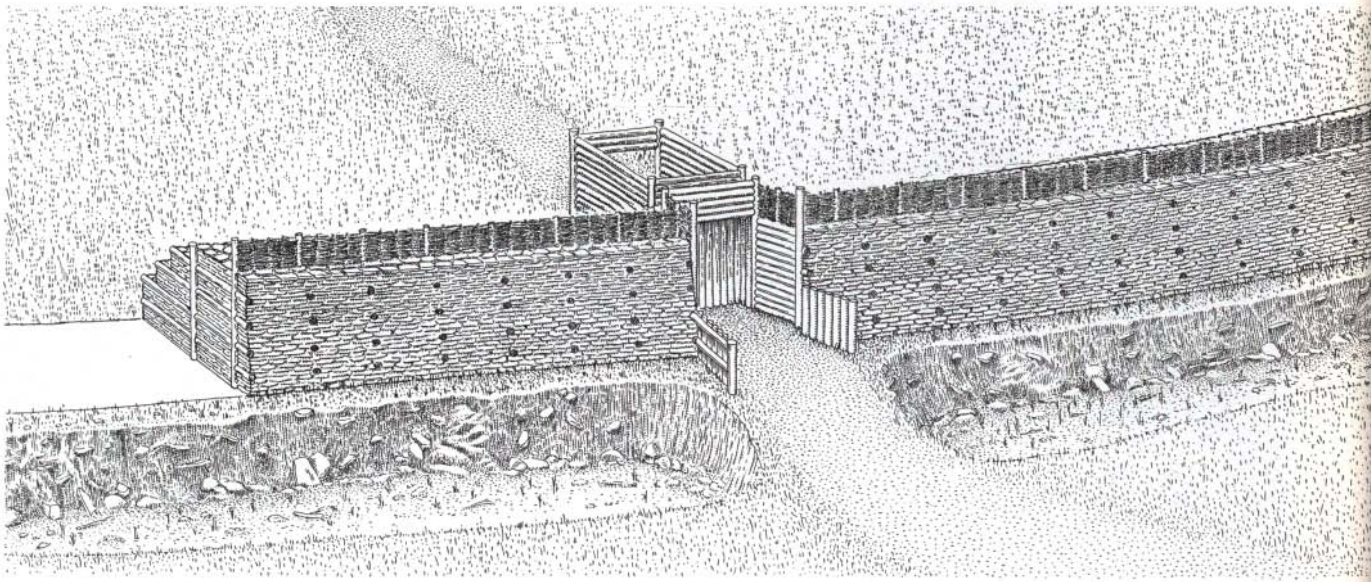
rock might have removed all traces of the buildings, most archaeologists came to accept the view that the enclosures were not defensive for that reason and two others: (1) the enclosures seemed to have numerous entrances (offering relatively easy access to attackers) and (2) the Neolithic builders often ignored the best defensive site in favor of one that was not as strong.

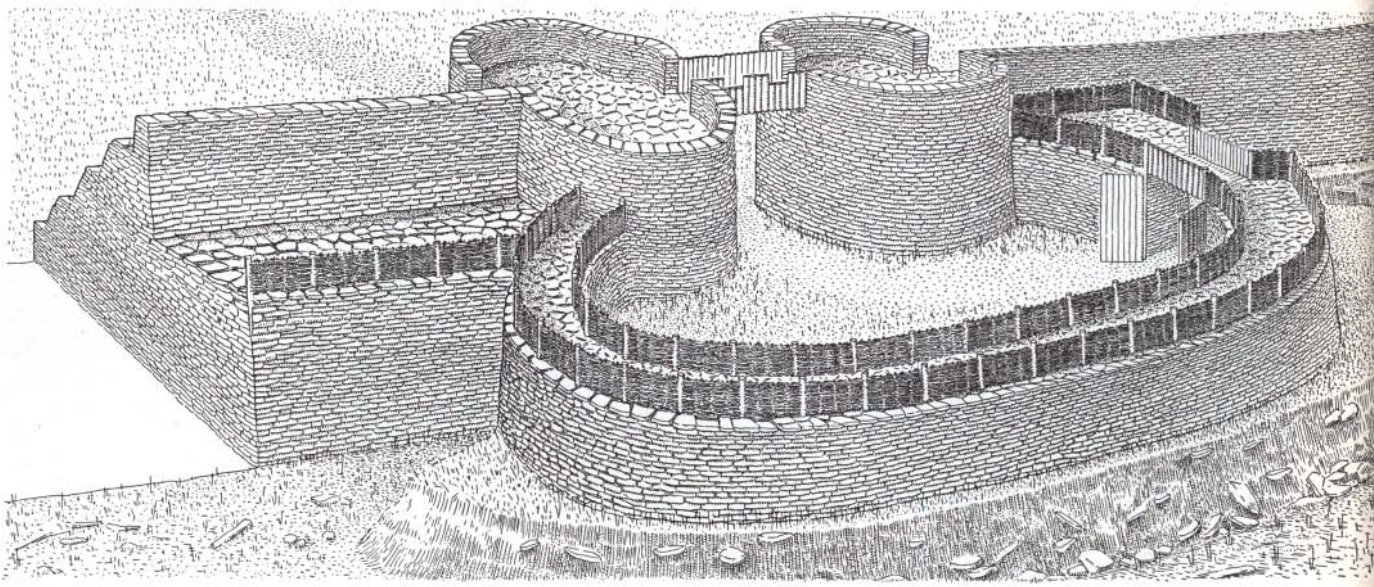
Accordingly it became the general view that the "camps" were probably livestock enclosures, perhaps serving for roundups at the end of summer. Examination of animal bones excavated at some of the sites has shown a preponderance of the bones of young cattle but not those of calves one would expect to find from a program of winter slaughtering. Some ceremonial killing is now considered likely. Burials of articulated

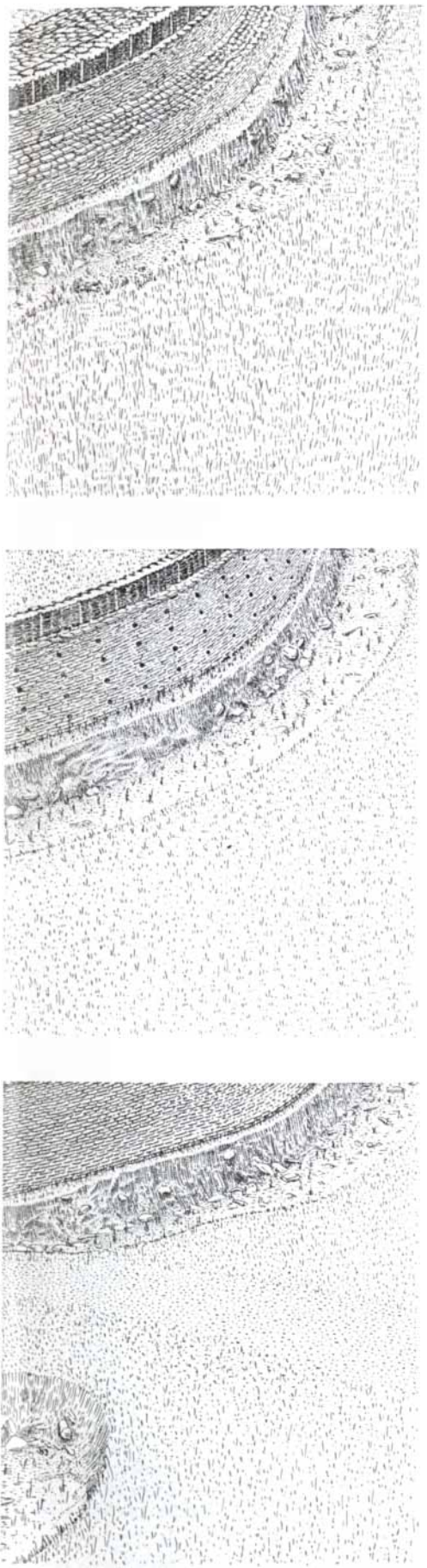


sageways. The enclosure was soon destroyed, and Crickley Hill was deserted until new settlers built the first of the two hill forts in about the sixth century B.C. The new fortification was 300 meters long and cut off an area of seven acres. Most of the settlement lay to the right of the Neolithic "causewayed camp." Color designates features belonging to the first hill fort, such as the postholes of its distinctive rectangular houses; black designates structures of the second hill fort, built only a few years after the first. The large round house near the entrance partly overlapped the earlier fort's postholes. In spite of the substantial new defenses the second settlement was also destroyed.









animal skulls and feet in distinct positions in the ditches indicate that the enclosures were the spiritual centers of the tribal communities that built them. Recently at Hambledon Hill in Dorsetshire excavators discovered human skulls set in identical positions at regular intervals on the bottom of the encircling ditch, which suggests to the excavator that the enclosure was a great necropolis for neighboring settlements. As more work is done the problem becomes more complex, and current opinion is inclined to regard the causewayed camps as serving a variety of practical and ritual functions, some of them being multipurpose and others specialized. Indeed, the various causewayed camps may have in common only the interruption of their ditches by causeways.

The significance of the causeways has in the past perhaps been exaggerated because the ditches have supplied the main body of evidence; erosion and plow damage have in most cases removed the accompanying banks and levels of human occupation. At Crickley Hill the banks are relatively well preserved and give a more detailed picture of how the enclosure was built. The two lines of ditches are crossed by 18 causeways, but 14 lacked corresponding gaps in the bank behind them and never gave access to the enclosure. The number of entrances is still large for a defensive fortification: four in 140 meters, with the original number likely to have been at least six in 260 meters. Postholes in the entrance passageways show, however, that all the entrances were closed by secure gates. The central gate of the inner ring is evidence of careful planning; after the ditch there had been dug about 30 centimeters deep work on it was stopped and a length of it was filled in to make an entrance causeway topped with a road surface. Dirt and debris from subsequent occupation covered the road and accumulated in the bottom of the ditch. The fact that the opening of the central gate matches one in the outer ditch indicates that outer and inner ditches were part of the same plan and that the plan was changed, by design or inadvertence, during the construction of the enclosure.

Excavations at the top of the knoll at the center of the Crickley Hill enclosure have uncovered hundreds of filled prehistoric postholes, gullies and small pits. Broad distinctions can be made between Iron Age and Neolithic features, but it is much harder to distinguish the various phases of Neolithic occupation. According to studies of the pottery associated with the Neolithic levels, all the Neolithic phases shared a culture so similar that separation of the cultural materials into periods would be unreliable. No carbon-14 dates for these early phases are yet available, but the pottery suggests a span for the Neolithic enclosures of between about 3500 and 2500 B.C. So far we can be sure only that there was substantial Neolithic occupation of the hilltop and that the inhabitants made tools out of flint, which since it is not found in the immediate area must have been imported from a distance of at least 30 miles. Analysis of thin sections of the pottery and the microfauna of the pits and postholes may in due course help to distinguish the various phases.

After a period of use the enclosure was apparently abandoned, but not before the banks were thrown back into the ditches and the inflammable material in the rubble was thoroughly burned. The sequence suggests a ritual, since one would expect an abandoned structure to be left to decay on its own. Our sections cut through the filled-in ditches show that in the next phase of occupation all the ditches were partially emptied, presumably in order to re-create the banks along them. In the same phase dry-stone walls about a meter high were built along the outer edge of the ditches. At the southern entrance to the old enclosure the walls turned inward to flank a new entry road, slightly offset from the preceding one, giving the impression of restoration and continuity. This enclosure too was later dismantled and its ditches were refilled, but in some areas the ditches seem to have been partially emptied once again before the entire complex was abandoned.

This odd succession of events preceded a still odder development in the next (and last) Neolithic phase. A much larger ditch was dug immediately outside and parallel to the main inner ditch of the earlier causewayed ditches. Unlike the earlier ditches, the new one was interrupted by only two entrance causeways. In this respect it resembled the defensive ditch of a hill fort. Its bank also had a military aspect: it was faced with a vertical stone wall and was surmounted by at least one fence. Its two long and narrow entrance passageways, the smaller one no more than a meter wide, were closed by strongly supported gates. The construction was not, however, entirely rational: the builders chose to dig the new ditch in solid lime-

FORTIFIED ENTRANCES enclosed at least three prehistoric settlements on Crickley Hill. Builders of the last Neolithic causewayed camp covered its banks with stone facing and interrupted the ditch with only two entrance causeways, one of which appears in the reconstruction at the top. The ramparts of the first Iron Age hill fort are reconstructed in the middle. They consisted of dry-stone walls held together by interlacing timbers and had a single entrance. Defenses of the second fort, reconstructed at the bottom, were more complex than preceding fort's.

stone rather than emptying and enlarging the earlier ditches. Moreover, the bank of the new ditch was lower than it could have been, because it was raised on the subsiding surface of the filled-in old ditch.

Did the new ditch deliberately avoid the old one? Its plan seems to have been an overt act of piety. Near the southern entrance the causewayed ditches bend irregularly outward for a short distance. The diggers of the new ditch, initially excavating a trench about half a meter deep, failed to notice the deviation and broke into the side of one of the earlier ditches, which they then refilled. They completed their ditch to its full depth of two meters on a new line that had a pronounced wrinkle. The builders' superstition is reflected not only by the layout of the enclosure but also by small pits at the ends of the ditch, which are sealed by large, flat stone slabs and contain the bones of butchered cattle.

In this final Neolithic phase, if not in the earlier ones, the bank and ditch did enclose a settlement. From the smaller entrance a cobbled road flanked by a fence led to the interior. Beside the road we found the postholes of at least one rectangular house, and the surrounding area was strewn with the debris of human occupation. The total area available for habitation was less than two acres, much of which remains to be excavated. The variations in the concentration of finds so far suggest an irregular pattern of settlement with two dense areas, one around the central knoll and the other just inside the banks. Between the two areas is an industrial zone for the manufacture of flint tools. We do not know the size of the population or even whether the site was occupied seasonally or permanently. The careful elaboration of the ramparts does, however, give the site an air of permanence. The postholes of the rectangular house were recut, indicating that the original house was rebuilt, probably twice, and that it was occupied for at least a generation or two.

The end of the final Neolithic settlement was sudden and violent. In the ditch and its rampart and in both entrances to the enclosure we found scores of leaf-shaped flint arrowheads, some of them scorched from the fire that destroyed the gates and the house and left a red streak along the back of the rampart where a wood fence had stood. These unmistakable signs of siege confirm evidence from Carn Brea in Cornwall, where great numbers of arrowheads have been recovered. Such arrowheads have been regarded as relics of hunting, not of warfare, because archaeologists have long believed early Neolithic society in Britain was fragmented and peaceful. The picture that is now emerging is one of a sedentary and stratified society capable of achieving substantial communal works, a society that both built fortified settlements and attacked them. Recently excavations at Hambledon Hill in Dorsetshire uncovered in the ditch of a small Neolithic enclosure the skeleton of a man with a leaf arrowhead in his chest.

Deserted after the burning, the Crickley Hill enclosure was left to fall into ruin. The stratigraphy of the deposits behind the small southern entrance indicates the next event. After the rear wall of the rampart had collapsed but before there had been time for much humus to collect in the area immediately behind it a great linear mound, four meters wide and nearly 100 meters long, was built in the area. The mound consisted almost entirely of topsoil that had been laboriously scraped from the sides of a natural gully. When the soil was in place, the builders laid limestone slabs intermittently to form a curb on each side of the mound. The eastern end of the structure, which overlapped the back of the old bank, was semicircular, its margin defined by a crescent of cobbling. In the middle of the crescent we found a large posthole, apparently the socket for a marker or totem.

LINES OF POSTHOLES indicate the size of a rectangular house in the first Iron Age fort. Side walls stood two meters outside holes; width of the house was between seven and eight meters.

No signs of human burial have yet been found in the mound, so that it appears to have been a ceremonial site, perhaps the focal point of a procession. The age of the mound is not known, but it had already assumed its present eroded shape when an Iron Age hut was built on its flank. Once the long mound had fallen into disuse the hilltop was the scene only of grazing and hunting. The topsoil has yielded early Bronze Age (post-Neolithic but pre-Iron Age) flint arrowheads and scraps of pottery but no other signs of occupation for nearly 2,000 years. Then, in the sixth century B.C. or shortly before, some unknown need brought new settlers to the hill and the first of two successive hill forts was begun.

The new fortification was larger than the old enclosures, and most of the settlement lay outside the area of the Neolithic occupation. Behind a strong rampart with dry-stone walls bonded together by a complex framework of lacing timbers we have uncovered the large postholes of six rectangular houses, the largest house more than 24 meters long. The houses formed a regular pattern

along a road that continued the line of an entrance passageway. Two more houses of the same type were built on the last Neolithic ditch, by then almost silted up, where the remnant of its bank might afford some shelter from the prevailing southwest wind.

Other Iron Age houses, more scattered, lay inside the Neolithic enclosure. At least three had a hearth, and the rest may also have had one; most of the floors were eroded away during the final occupation of the fort. In addition there were at least 27 smaller structures, most of which were built on four posts in a square array. Some of them belonged to the first hill fort and others to its successor; none had a hearth, and two were full of barley when they were burned. Storage pits are common in prehistoric settlements in Britain built on chalk, but none has been found in the Crickley Hill limestone. Therefore the "four-posters" seem to be granaries or storage huts. In the first hill fort they were scattered in a broad zone surrounding the main rectangular buildings. Three four-posters close to the entrance of the fort may also have been granaries, but a defensive purpose appears more likely: a small pit beside the two northern huts contained sling stones.

Whatever the defensive systems of this first fort may have been, they clearly failed against its last attackers, who thoroughly burned it, threw down its walls and ignited the timber-laced core of the bank. Before the fort was rebuilt there was enough time for some layers of debris to wash down from the ruins but not enough for a weathered soil to form. Since the formation of such a soil takes more than months but less than decades, it is clear that the second fort was built soon after the destruction of the first one. The builders radically changed the old fort's defenses and its internal planning. The new entrance was of massive stone with solid round-fronted bastions and an outer work with its own gate. The ramparts, thicker than before, rose in an impressive tier. Inside the fort a huge round house, about 15 meters in diameter, dominated the area where the regularly aligned houses had once stood. Because the round house stood on a line with the entrance, the cobbled roadway through the entrance turned south to flank a row of small round houses, each one about eight meters in diameter. All the round houses had a well-built hearth. As in the first fort, the four-posters were without heating and were clustered at a slight distance from the living area.

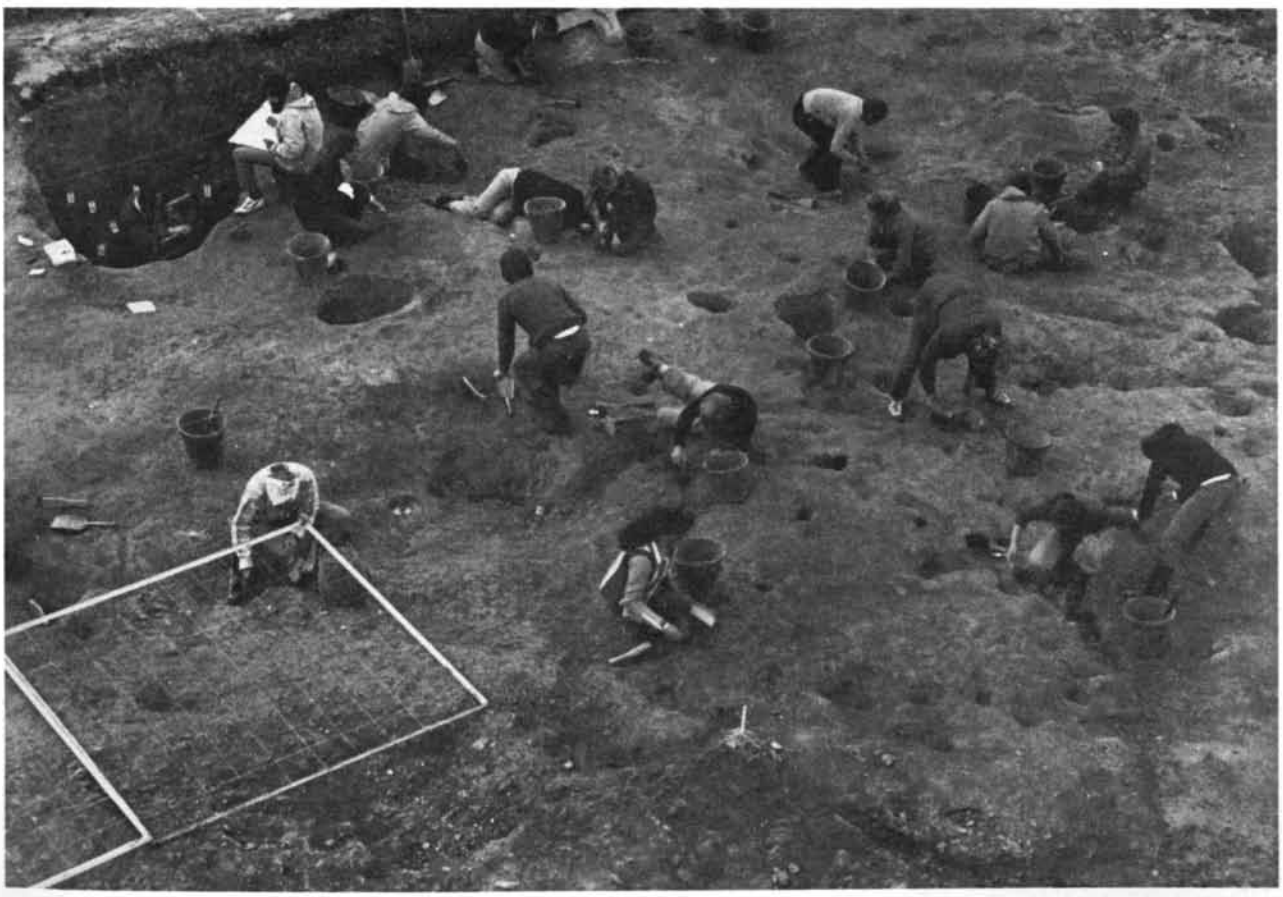
In spite of the sophisticated defenses of the second hill fort, it fared no better than the first. The evidence is that it was attacked before there was time for its new roads to become rutted and its tim-

bers to decay. The extent of the destruction is indicated by thick layers of charcoal at the entrance of the fort and by the fire-reddened walls of the bastions. As the great round house burned, a west wind impressed a red streak on the limestone.

It is unlikely that the identity of the attackers or the defenders of either hill fort will ever be known. The pottery of the second fort resembles that of the early Iron Age of the upper Thames valley in Oxfordshire, about 20 miles away. The pottery of the first fort is harder to identify and seems to belong to a different tradition. These artifacts, together with the evidence of an abrupt change in the design of houses and ramparts, strongly indicate that the populations of the two phases were quite distinct from each other. Of the two forts the earlier one is the more notable. Indeed, its long rectangular houses are without parallel among the Iron Age buildings excavated in England, which otherwise have been round houses and small square or rectangular structures of uncertain purpose. In that period the long rectangular house was common on the European continent, but it is unnecessary to hypothesize that the Crickley Hill builders were immigrants from across the English Channel; a growing body of evidence attests to the existence of large rectangular houses in the Neolithic and early Bronze Age of Britain. Further excavation elsewhere may uncover more examples to place beside the ubiquitous native round house.

As far as we know Crickley Hill was never again the site of a settlement. Nearby hilltops were fortified in the fourth century B.C. and later, and villages clustered in the lowlands below them, but the Crickley Hill ramparts were left to collapse without interference. Less than a mile away, on the slope overlooking Crickley Hill, a woman of the highest status was buried with her jewelry shortly before the Roman conquest, but nothing of this date has been found in our excavations. The hilltop seems to have reverted to pasture, its solitude continuing to the present, interrupted only by a metalworker who built his shelter in the entranceway of the last fort. It may indeed have been his pig that was crushed when the bastion walls finally collapsed in the second century A.D. or later.

Ten years of digging at Crickley Hill has yielded a remarkable series of discoveries, including the clearest evidence in Britain of an organized and fortified Neolithic settlement, a curious blend of ritual and functional elements. There is no reason, however, to suppose the site is unique in the complexity of its occupation: it is the scale of our project that has afforded the opportunity for discovery. Of the thousands of acres within the ramparts of the hill forts of southern England fewer than 1 percent have been excavated so far. From such a small sample firm conclusions about settlement patterns cannot be drawn. Even at Crickley Hill, where work on about 40 percent of the interior area is now complete, answers to the significant questions about the social structure, population and economy of the various occupations have only begun to emerge. Coherent arguments bearing on land use and the relations between Crickley Hill and other centers of occupation in the region are premature before the contemporaneity of Crickley Hill and the other sites is established. Much exciting work remains to be done. With the help of our volunteer labor force the summers of the next 10 years may prove as fruitful as those of the past 10 in deciphering the achievements and motives of the prehistoric settlers of Crickley Hill.



VOLUNTEER DIGGERS under the author's direction have excavated nearly half of the Crickley Hill site in the past 10 years. Here they clean the earliest Neolithic surface, pockmarked by postholes. At the top left diggers uncover a section of a causewayed-camp ditch.

WHAT'S BETTER THAN SPEED READING?

SPEED LEARNING

(SPEED PLUS COMPREHENSION)

Speed Learning is replacing speed reading because it's easy to learn . . . lasts a lifetime . . . applies to everything you read . . . and is the only fully accredited course with the option of college credits.

Do you have too much to read and too little time to read it? Do you mentally pronounce each word as you read? Do you frequently have to go back and re-read words or whole paragraphs you just finished reading? Do you have trouble concentrating? Do you quickly forget most of what you read?

If you answer "yes" to any of these questions — then here at last is the practical help you've been waiting for. Whether you read for business or pleasure, school or college, you will build exceptional skills from this major breakthrough in effective reading, created by Dr. Russell Stauffer at the University of Delaware.

Not just "speed reading" — but speed reading-thinking-understanding-remembering-and-learning

The new *Speed Learning Program* shows you step-by-proven-step how to increase your reading skill and speed, so you understand more, remember more and use more of everything you read. The typical remark made by the 75,000 slow readers who completed the *Speed Learning Program* was: "Why didn't someone teach me this a long time ago?" They were no longer held back by the lack of skills and poor reading habits. They could read almost as fast as they could think.

What makes Speed Learning so successful?

The new *Speed Learning Program* does not offer you a rehash of the usual eye-exercises, timing devices, costly gadgets you've probably heard about in connection with speed reading courses or even tried and found ineffective.

In just a few spare minutes a day of easy reading and exciting listening, you discover an entirely new way to read and think — a radical departure from anything you have ever seen or heard about. Research shows that reading is 95% *thinking* and only 5% eye movement. Yet most of today's speed reading programs spend their time teaching you rapid eye movement (5% of the problem) and ignore the most important part (95%) *thinking*. In brief, *Speed Learning* gives you what speed reading *can't*.

Imagine the new freedom you'll have when you learn how to dash through all types of reading material *at least* twice as fast as you do now, and with greater comprehension. Think of being able to get on top of the avalanche of newspapers, magazines and correspondence you have to read . . . finishing a stimulating book and retaining facts and details more clearly and with greater accuracy than ever before.

Listen-and-learn at your own pace

This is a practical, easy-to-learn program that will work for you — no matter how slow a reader you think you are now. The *Speed Learning Program* is scientifically planned to get you started quickly . . . to help you in spare minutes a day. It brings you a "teacher-on-cassettes" who guides you, instructs, encourages you, explaining material as you read. Interesting items taken from *Time* Magazine, *Business Week, Wall Street Journal, Family Circle, N.Y. Times* and many others, make the program stimulating, easy and fun . . . and so much more effective.

Executives, students, professional people, men and women in all walks of life from 15 to 70 have benefited from this program. Speed Learning is a fully accredited course . . . costing only 1/5 the price of less effective speed reading classroom courses. Now you can examine the same, easy, practical and proven methods at home . . . in spare time . . . without risking a penny.

Examine Speed Learning FREE for 10 days

You will be thrilled at how quickly this program will begin to develop new thinking and reading skills. After listening to just one cassette and reading the preface you will quickly see how you can achieve increases in both the speed at which you read and in the amount you understand and remember.

You must be delighted with what you see or you pay nothing. Examine this remarkable program for 10 days. If, at the end of that time you are not convinced that you would like to master Speed Learning, simply return the program and owe nothing. See the coupon for low price and convenient credit terms.

Note: Many companies and government agencies have tuition assistance plans for employees providing full or partial payment for college credit programs.

In most cases, the entire cost of your Speed Learning Program is Tax Deductible.

EARN PROFESSIONAL & COLLEGE CREDITS

Speed Learning is approved for credit by the following professional and educational institutions:

- **Foundation for Accounting Education** 20 CPE Credit Hours
- **American Management Association** 1.9 Continuing Education Units
- **National Society of Public Accountants** 20 Continuing Education Hours
- **College Credit** 3 credits from the National College of Business

Details and registration forms included with each program.

OFFERED INTERNATIONALLY BY

- **Institute of Electrical & Electronics Engineers**
- **American Chemical Society**
- **National Association of Life Underwriters**

learn INCORPORATED 113 Gaither Drive, Mount Laurel, N.J. 08054

31SAN

Please send me the Speed Learning Program at $89.95 plus $3.00 for handling and insured delivery.
Please check the method of payment below:

☐ Check or money order enclosed for items ordered. New Jersey Residents add 5% sales tax.
☐ Please charge my credit card under the regular payment terms: ☐ Bank Americard
☐ Master Charge Interbank No. _____ ☐ American Express ☐ Diners Club
Card No. _____ Exp. Date _____

I understand that if after 10 days I am not delighted in every way, I may return the materials and obtain a full refund with no questions asked.

Name _____
Address _____
City _____ State _____ Zip _____
x Signature _____

If you don't already own a cassette player, you may order this Deluxe Cassette Recorder for only $49.95. (Includes handling and delivery.)
Check here to order ☐

Outside U.S.A. $99.95 + $3 surface mail — airmail extra

THE AMATEUR SCIENTIST

Flames in which air is introduced into a flammable gas rather than vice versa

by Jearl Walker

Looking at the familiar sight of a flame of natural gas burning in air, one might wonder what would happen if the gas and the air were interchanged, that is, if a jet of air were introduced into an atmosphere of natural gas. Would one see some kind of reverse flame, with the same shape, height, color and temperature distribution as a normal flame? These questions occurred to Stuart Travis of Akron, Colo., who with his investigation of a reverse flame of air in an atmosphere of methane won second place in the physics division of this year's International Science and Engineering Fair. The reverse flame is similar to a normal one but displays some surprising differences that neither Travis nor I can fully explain.

The normal and reverse flames investigated by Travis are diffusion flames: the burning requires the mutual diffusion of the fuel and the oxidizer into the flame front, the surface where the burning takes place. In the other general type of flame, premixed flames, the fuel and the oxidizing agent are mixed before they reach the flame, as in a Bunsen burner. Diffusion flames and premixed flames can be either laminar (that is, smoothly flowing) or turbulent, depending on the rate of flow of the participating gases. Travis limited his attention to laminar diffusion flames. The commonest example of this type of flame is a candle flame, which I discussed in this department for April, 1978.

One might think that laminar diffusion flames are well understood, since candle flames have been investigated for a long time, but it is not so. Flames are exceedingly difficult to understand even in general terms, much less in detail, because of the wide variety of phenomena occurring in them. Many chemical reactions take place, and many of them are complex. Light is emitted by both chemical and thermodynamic phenomena. Gases expand as they are heated. Heat is transferred across the surface of the flame in several ways. The flow is in three dimensions. Probing the flame usually distorts it in inexplicable ways.

Students of flames concentrate on several key features: color, shape, temperature distribution and burning velocity. Burning velocity is the rate at which the flame propagates perpendicular to the flame front. The propagation can be either through a stationary gas or, as in a flame on a stove, through a flame that remains stationary while the gas flows. It is easy to calculate the burning velocity when the flame front is flat but more difficult when the flame front is curved, as it is with most burners.

Another matter of interest in the investigation of a flame is the reaction zone, the region where the fuel and the oxidizer meet. Part of the light from the flame emerges from the chemical reactions in this zone. In a normal diffusion flame the reaction zone is blue, largely because of the blue emissions of the molecules excited by the chemical reactions in the zone. If the rate of flow of fuel is sufficiently low, the entire flame may be blue. An example is the flame of a candle with a thin wick.

With a somewhat greater flow of fuel a yellow or white tip develops near the top of a normal diffusion flame. As the flow of fuel is increased further the yellow or white region grows, eventually dominating the flame and relegating the blue reaction zone first to the sides of the flame and then finally to just the side regions near the base. The yellow-white tip consists of carbon particles that have become hot enough to emit radiation more or less continuously across the entire visible spectrum.

The combustion chamber employed by Travis for his investigation was made out of a 500-milliliter round-bottom flask. The flask was held upside down on a ring stand by a clamp. Projecting upward from the round surface was a tube that had been annealed to the flask. A similar tube projected sideways from the neck. A stopper in the neck carried a short length of metal tubing that served as the burner. To make a normal diffusion flame Travis supplied methane to the burner tube and air to the side tube. For a reverse flame he interchanged the supplies. A small hole in the side of the flask gave access for a thermocouple to probe the temperature of the flame. Also passing through the stopper in the neck were two electrical leads to the Nichrome wire Travis employed to start the flame inside the flask. His air and methane came from tanks in his school laboratory.

To light a reverse flame Travis placed the length of Nichrome wire across the mouth of the burner tube and then turned on the methane. After about 15 seconds the flask was filled with methane and he could light the gas escaping through the vent at the top of the flask. (He needed the 15 seconds to avoid having a mixture of a small amount of methane and a large amount of air, which would explode.) He adjusted the flow of methane until the flame at the vent was four or five inches high. After the reverse flame was created inside the chamber the flame at the vent burned off the excess methane escaping from the chamber.

The Nichrome wire was connected by two leads to a current source, which was then turned on enough to make the wire glow red. Next the air tank was turned on to supply air at a low rate (about .6 liter per minute). The flow of air had to be low at first or the amount of oxygen supplied by it might have caused the flask to explode. The hot wire ignited a flame at the mouth of the burner; this was the reverse flame. The current on the wire was then turned off and the leads through the stopper were wiggled until the wire was brought down from its position over the mouth. The flame was more distinguishable in the dark, so that most of these things were done in dim light. To extinguish the flame Travis turned off the air first to avoid any danger of explosion.

Travis lighted the normal flame by a similar procedure. Again he carefully avoided lighting the flame at the top of the flask while a small amount of methane was in a large amount of air. He turned on the methane first to fill the flask before he turned on the air. To turn off the flame he first turned off the air. When I asked him about the danger of explosion, he replied that he had had no problem with the reverse flame but that with the normal flame he had occasionally had an explosion that blew the stopper out of the flask. To guard against a more serious accident anyone doing this kind of experiment should wear a face shield. Moreover, a combustion chamber larger than Travis' 500-milliliter flask should not be used or the hazard of explosion will be greater.

The distribution of temperature in the flames was measured with the thermocouple probe slipped into the combustion chamber through the small hole. Travis got his thermocouple from the Edmund Scientific Company (6975 Ed-

During the last 50 years, we've been working to answer a few simple questions.

Questions like how? and why? and what? are the essential questions every researcher must ask. Since 1929, we've been asking – and answering – these simple questions to help develop a wider range of complex technologies – from jet engines to lasers, helicopters to radar. And in the last five years, we've been doing intensive research on one of our most basic questions: How can we make better use of our energy?

How does coal burn? Our coal devolatilization project uses a specialized infrared spectrometer to help determine the molecular structure of different coals during thermal decomposition. Once we know what really happens when coal burns we can process those coals that have excessive pollutants or seek out those which are best for liquefaction programs.

How can we look inside a flame? Coherent Anti-Stokes Raman Spectroscopy (CARS) is one way. We mix laser beams at different frequencies. Tune them to specific molecules inside a flame or combustion chamber. The beams interact and generate a new beam at a different frequency. By analyzing its color spectrum we get temperature, and by looking at its strength we measure concentration. CARS is helping us chart how fuel and air really mix before and during actual ignition. It's giving us information that will lead to engines and fuels that run cleaner and more efficiently.

How can we put the laser to work? We're just beginning to realize the potential of the multi-kilowatt CO_2 laser. Our researchers were the first to develop a laser that makes the kind of fast, deep penetration weld shipbuilders and automakers need. We've also experimented with Layerglazing™; a process for melting and rapidly solidifying material into a useful shape 1/50,000 of an inch at a time. This technique will help improve the properties of different kinds of metals and alloys.

For half a century, we've been in the business of asking the questions that can help shape the future. That's why we're always looking for people who look to the future. If you're interested in knowing more about the research we do, please write United Technologies Research Center, Silver Lane, East Hartford, CT 06108.

It'll help give you an idea of the kinds of questions we'll be asking in the next fifty years.

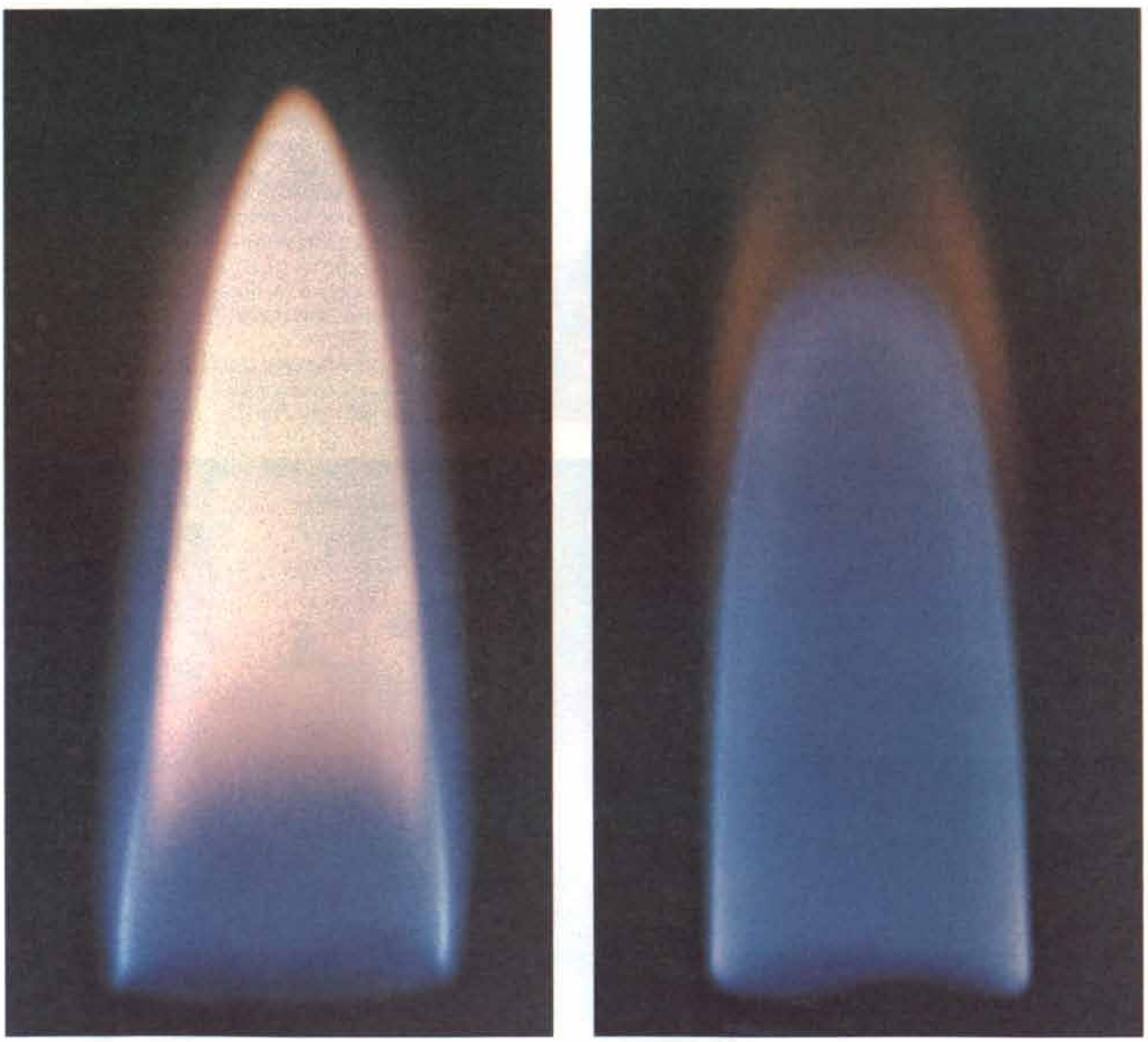
scorp Building, Barrington, N.J. 08007). He ground down the tip of the Chromel-Alumel wire, rebrazed it and ground it down again to make it as small as possible. Still the probe was relatively large compared with the flame and therefore could measure only average temperatures. One must also bear in mind two unmeasurable effects of the probe: it distorts the flame, and by carrying off heat it distorts the temperature distribution.

To record the location of the probe in the flame Travis mounted a ruler outside the chamber in such a way that the distance from the tip of the thermocouple to the hole in the side of the chamber equaled the distance from the hole to the ruler. Hence when the tip was moved a certain distance within the flame, the stiff wire moved an equal distance across the ruler, enabling him to measure the displacement in the flame.

Travis measured the rate of flow of gas with two relatively inexpensive flow meters (Dwyer brand) that worked in the range of a few liters per minute. (Local distributors for flow meters are listed in the yellow pages of the telephone book.) For very low gas flows Travis approximated the rate of flow by bubbling the gas through a container of water. He estimated the average bubble size and then counted the number of bubbles passing through the water each minute. A simple calculation gave him the volume flow rate.

The spherical combustion chamber distorted photographs of the flames, and so Travis replaced it with a rectangular chamber made out of ordinary plate glass. When he lighted a flame, he had to work fast because the heat soon cracked the glass. More expensive glass designed for thermal stress could have been substituted to avoid the problem. Travis made his photographs with a 35-millimeter camera fitted with a 55-millimeter lens and a one-power extender. He made black-and-white pictures, using Kodak Tri-X film at an aperture of $f/2.8$ and an exposure time of a thirtieth of a second.

To measure the burning velocity of the flames Travis employed a technique known as Gouy's method, which is based on the total surface area of the flame front. The burning velocity is assumed to be constant over the entire flame front in spite of the front's curvature. The volume flow rate of gas from the burner is equal to the area of the mouth of the burner multiplied by the velocity of the gas issuing from the mouth (as measured by a flow meter). Since the flame is stationary, the volume flow rate must also equal the burning velocity multiplied by the total surface



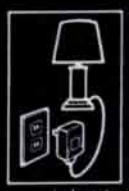
A normal flame **(left)** *and a reverse flame* **(right)**

Who could blame you?

CHIVAS REGAL • 1.75 LITER BOTTLE • 12 YEARS OLD WORLDWIDE • BLENDED SCOTCH WHISKY • 86 PROOF • GENERAL WINE & SPIRITS CO., N.Y., N.Y.

For the man who has everything. Something to control it with.

The BSR System X-10™ The Controller.



JUST PLUG IT IN.
Just plug the system into any wall outlet. Wherever you choose. (No special wiring is needed.) And you can control up to sixteen lights and appliances just by pushing a couple of buttons.

As a matter of fact, the X-10 is so easy to operate, you can do it with your eyes shut. So, if the late show is beginning to turn you off, you can turn it off without budging out of bed. And keep your eyes shut until morning.

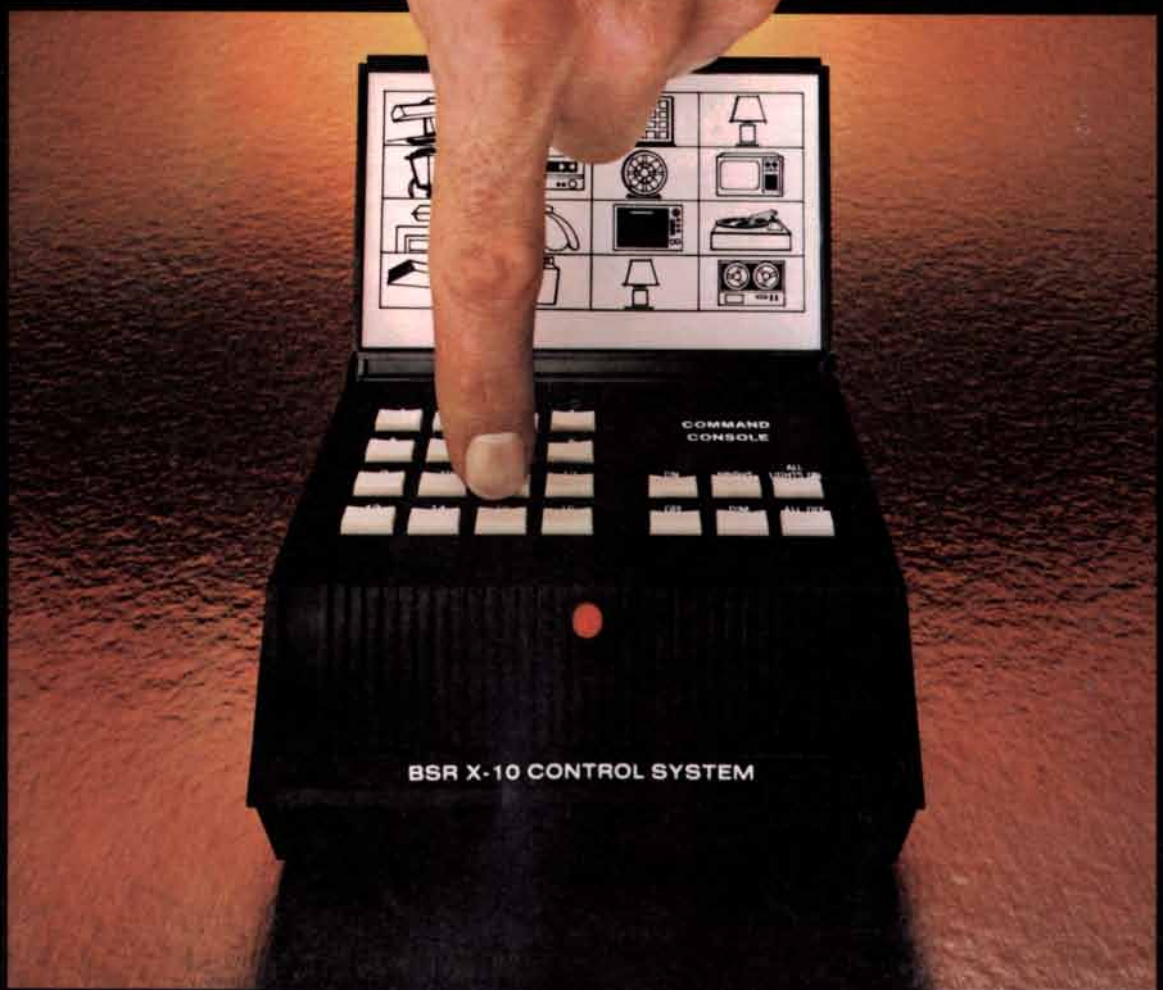
Or, say you hear a mysterious noise from downstairs in the middle of the night. Just reach for your Controller and turn on every light in the house. That'll scare the daylights out of an intruder.

Or even if you just want to dim your lights, you can do it with the Controller.

UNDER $90.*
And get this. An X-10 Control System with enough modules to control lights and appliances at three different locations in the house, costs under $90. So the man who has everything doesn't have to have a pot of money to own one.

BSR System X-10.

Everything's Under Control.



The Controller

BSR (USA) LTD. Blauvelt, N.Y. 10913

*Manufacturer's suggested retail price.

area of the flame front. If the surface area can be determined, the only unknown is the burning velocity.

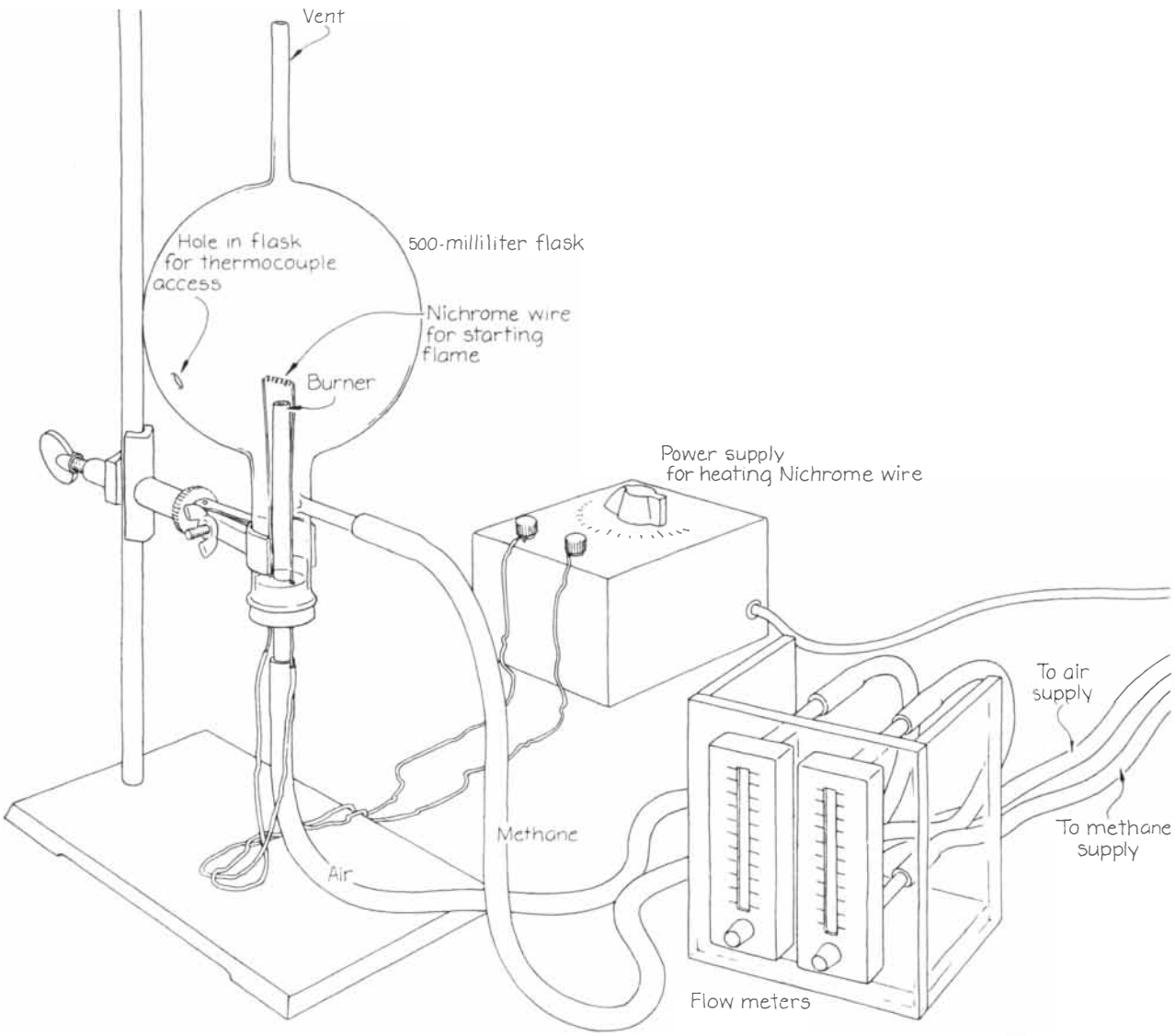
To determine the total surface area of the flame front a photograph is made of the flame and then the flame in the picture is divided in half along its central vertical axis. One of the halves is split into several sections as is shown in the illustration on the next page. For each such section two sides are radii from the central axis out to the flame front and another side is the sloping side along the flame front. The surface area along the circumference of the flame corresponding to one of these sections is calculated by multiplying pi by the average of the two radii for the section and then multiplying by the length along the slope. The calculation is done for each of the sections. The total area obtained is approximately the total surface area of the flame front.

How accurate is the result? The calculation can be in error for several subtle reasons. One is that determining the position of the flame front can be difficult if the reaction zone is thick, as it is likely to be in flames with low burning velocities. Another source of error stems from the difficulty of calculating the surface area at the top of the flame, where the flame front may be substantially curved. A larger error may arise because the base of the flame is often poorly defined in a photograph. Finally, the burning velocity of the flame may be distorted by the thermal expansion of the gases when they enter the reaction zone and are heated.

Predicting the shape or height of a normal diffusion flame is no easy matter either. Both depend on a complex interaction of heat transfer, diffusion and the rates of the chemical reactions. In a simple analysis the height of a diffusion flame should be proportional to the volume flow rate and inversely proportional to the average rate of diffusion of the gases.

The reverse flames Travis created in his apparatus are just as difficult to analyze as normal diffusion flames, and perhaps even more so. They are noticeably different from the normal flames in size, shape and color. A reverse flame is usually rounded at the top instead of having the conical tip of a normal flame. A normal flame with a certain volume flow rate of methane is much higher than a reverse flame with the same volume flow rate for its air. The normal flame usually has a poorly defined base, whereas the reverse flame has a distinct base. Moreover, its base is separated from the mouth of the burner by a relatively wide dead space.

The outstanding difference between the two flames is in their color. A normal flame with a small flow of fuel is blue throughout. As the flow of methane is increased a yellow or white tip ap-



The combustion chamber designed by Stuart Travis for a reverse flame

pears at the top of the flame, and the blue regions are reduced and shifted downward to the base. A reverse flame is entirely blue until a large amount of air is forced into the flame. Then the top turns orange.

The temperature distribution of the two flames is slightly different. The normal flame tends to be hottest along the side of the inner cone of darkness; the reverse flame has somewhat hotter areas over the inner cone. Bear in mind that these measurements are affected by the introduction of the thermocouple probe into the flame and by the fact that the thermocouple averages the temperatures over a region, even though the region can be a small one.

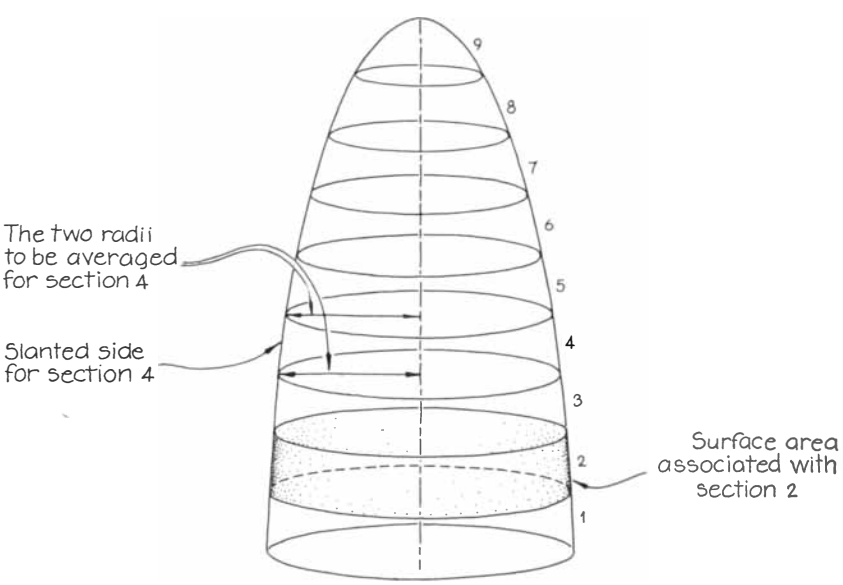
The reverse flame also has a higher burning velocity. When Travis adjusted a normal flame and a reverse one so that they were the same height, the reverse flame had about twice the burning velocity of the normal one. When he adjusted the flames to have equal flow rates (methane for the normal flame and air for the reverse flame), the ratio was about 10. In both types of flame the burning velocity decreased (from a few centimeters per second) as he increased the flow rate because the area of the flame front did not increase as much as the flow rate.

Why does the reverse flame differ from the normal flame? I am not sure, because even the normal diffusion flame is not completely understood. The reverse flame and the normal one are similar when the flow rates of each are low. Then both flames are rounded at the top and blue in the reaction zone. As the methane flow rate for the normal flame is increased, the yellow or white tip quickly appears. The corresponding orange tip of the reverse flame develops less rapidly because the flow of air must be increased considerably to bring it out.

It is puzzling that the colors of the tops are not the same. They should be (as far as I know) because in both flames the color of the top is due to the visible emissions from incandescent particles of carbon. Since the air supply in the reverse flame must be relatively high to create an orange tip, the formation of incandescent carbon particles may also require a relatively large supply of air.

The fact that the methane diffuses from the outside of the reverse flame rather than from the inside as it does in the normal flame may be a key to the differences between the two types. Methane needs a fairly high temperature for decomposition. If the methane comes from inside the flame, it enters the very hot regions of the flame rather quickly. The means by which it eventually forms solid carbon particles is not understood, although the process is thought to be incomplete until the gases have risen to the top of the flame. When methane diffuses into a reverse flame, it enters from the cooler regions and is gradually heated, and so the reactions that form the particles may be different. My speculations on these matters can be only rough guesses, however, until more work has been done on reverse flames.

A normal diffusion flame that is stable will sit slightly above a burner opening at a height where the burning velocity matches the gas velocity. If one velocity changes significantly, the flame will move either away from the mouth or into it. The gas velocity decreases somewhat as the gas stream spreads from the mouth of the burner, so that the flame front settles at the point above the mouth where the two velocities match. Certain flames (such as ethylene jets)



How a flame is sectioned to determine its surface area

Dahlem Workshop Reports

Dahlem Workshops derive their name from the section of the City of West Berlin where they were founded. Four times annually top experts from the international scientific community are invited to spend a week in Berlin in order to do hard and creative work: They discuss an interdisciplinary subject of high topicality until an authentic and well balanced report on the state of the art can be given.

These unique reports are subsequently publicized as ***Dahlem Workshop Reports.***

Besides taking stock of contemporary knowledge, every Dahlem Workshop is expected to review new concepts and new experimental techniques. Here the task of the group is to recommend future directions of research, which are both promising and urgently needed in order to fill gaps of knowledge. Again, these recommendations constitute part of the Dahlem Workshop Reports.

Each Dahlem Workshop Report is published within six months after the end of the conference. In this way the international community of scientists can benefit from the results at the earliest possible point in time.

From the two series ***"Physical and Chemical Sciences Research Reports"*** and ***"Life Sciences Research Reports"*** 13 titles are available altogether. For more and detailed information please use the attached Info-Coupon.

New

Life Sciences Research Report 12
H. Gerischer, J. J. Katz
Light-Induced Charge Separation in Biological and Chemical Systems
1979. 553 pages. Paperback DM 75,-- $44,20

Life Sciences Research Report 13
M. Shilo
Strategies of Microbial Life in Extreme Environments
1979. 514 pages. Paperback DM 70,-- $42,40

Life Siences Research Report 14
J. G. Nichols
The Role of Intercellular Signals: Navigation, Encounter, Outcome
1979. 309 pages. Paperback DM 44,-- $25,90

Life Sciences Research Report 15
Biomedical Pattern Recognition and Image Processing
In preparation

verlag chemie
Weinheim · New York

Info-Coupon

for Dahlem Workshop Reports

Address:

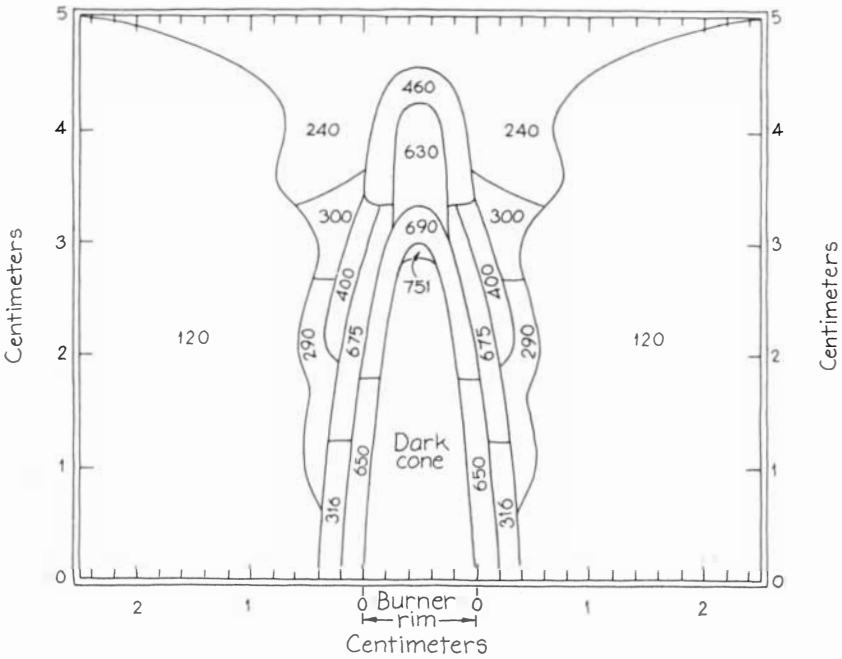
Please return to Verlag Chemie.
P.O. Box 1260/1280. D-6940 Weinheim
For USA and Canada:
Verlag Chemie International Inc.
175 Fifth Avenue New York. N.Y. 10010 USA.

may lift a good deal higher above the burner and then stabilize for a time. Such lifted flames are not easy to achieve because they call for a perfect balance between the gas velocity and the burning velocity that is difficult for the experimenter to achieve.

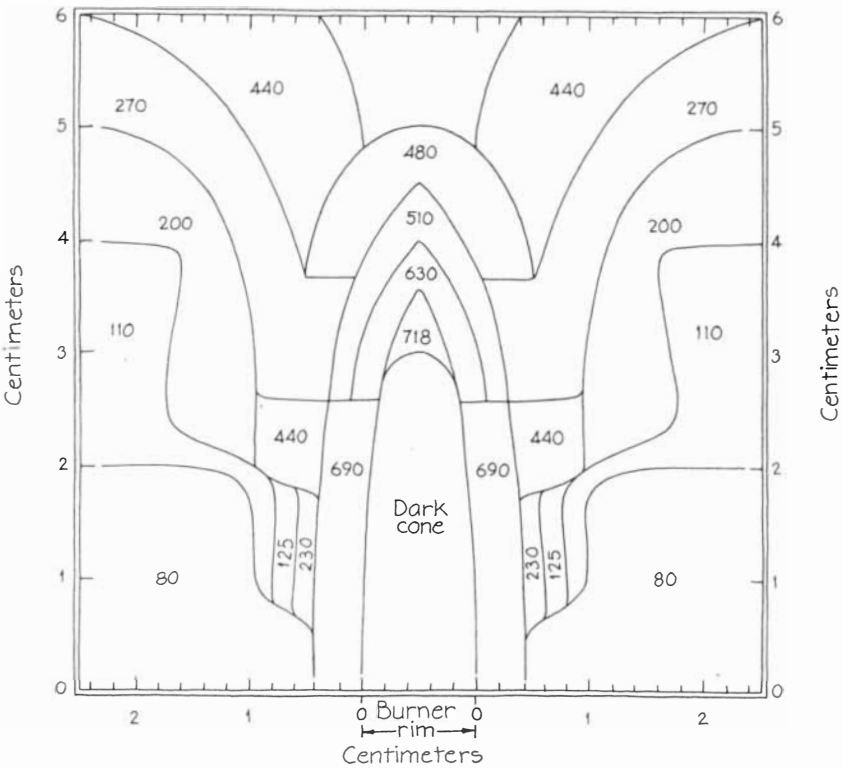
Travis could not make his normal flame lift off in this way but could easily make the reverse flame do it. He turned the methane supply off and then quickly turned it back on. The flame lifted several centimeters above the burner mouth, remaining there for up to four minutes. I am not sure why the reverse flame does this, but I suspect that the relatively high burning velocity may be the key.

When the methane supply in a normal flame is turned up high, the flame becomes turbulent. The reverse flame does not show such turbulence. When its air supply is turned up high, the flame moves away from the mouth of the burner and dies.

I can find no mention of reverse flames in the printed material on flames. Travis may be the first person to have investigated them. If you want to pursue the study, much further experimentation could be done. Try hydrocarbon gases other than methane. (Be careful if they are toxic or highly explosive.) Travis also experimented with propane; you might start there. In my earlier discussion of a candle flame I described how a simple spectroscope can serve for observing the line emissions of the chemical radicals CH and C_2 from the reaction zones. You may want to make similar observations on the reaction zones of reverse flames. If the emissions are missing, the chemistry of the reaction zones in the two types of flame would be quite different.



The temperatures (in degrees Celsius) around a normal flame



The temperatures around a reverse flame

Several people have written to me about the similarity between falling trees and the falling chimneys I discussed in this department for February. James D. Plimpton of Albuquerque, N.M., pointed out that a falling tree with full foliage may float briefly just before hitting the ground because of the large air resistance encountered by its branches. A more dangerous aspect of falling trees was described by Paul R. Burnett of Temple Hills, Md. A tree that is to be felled is first notched on one side and then given a single horizontal cut on the side opposite the notch. When the tree begins to fall, it hinges around the inch or so of wood remaining between the notch and the straight cut. If the tree is young, it may be noticeably bent by air resistance as it falls, but a live tree will not snap because of the high tensile strength of the living wood. The hinge, however, does snap, hurling the butt end of the tree upward with considerable force. Then that end may move in the direction of fall of the tree, as the lower end of a falling chimney does. If the tree has many full branches, much energy may be stored in the branches as they bend while striking the ground. The branches may then suddenly push the trunk back toward the stump with great force. Burnett has seen stumps almost knocked out of the ground by the impact. This sudden, violent recoil is particularly dangerous for anyone who remains at or even near the stump after felling the tree.

The falling behavior of snags (dead and partly decayed trees) is even curiouser. John D. Engels of North Bend,

The Gas Option

Gas energy: It can reduce our dependence on foreign oil.

Every barrel of oil we import increases our balance of payments deficit. It exports jobs and opportunity. It hampers our ability to control inflation here at home. And it impairs our national security.

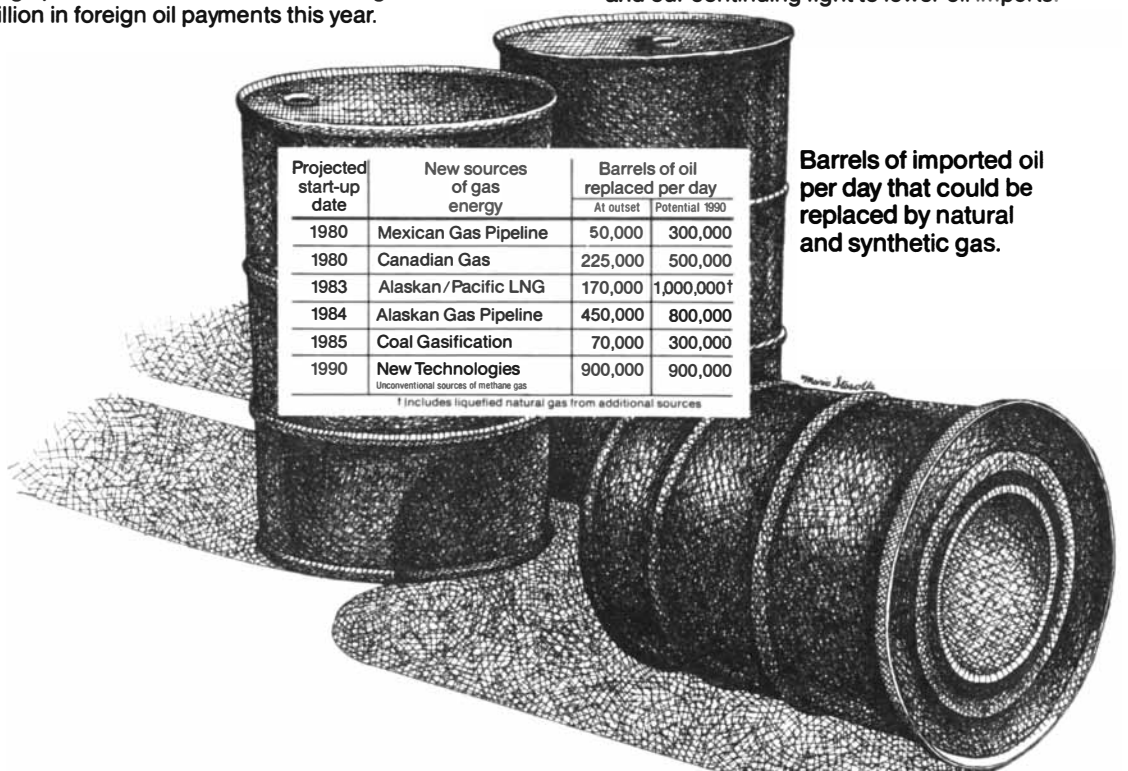
But we could be importing less oil. And the increased use of gas energy, from both conventional and new sources, can help.

Already, important progress is being made. Increased industrial use of natural gas this year is expected to offset an average of 450,000 barrels a day of imported oil. At an average price of $18 a barrel, this is saving the United States $3 billion in foreign oil payments this year.

Shown below are the target dates for various new sources of natural and synthetic gas and the potential daily barrels of foreign oil they can replace. Individually and in total, they represent a potential savings to the United States of nearly 4 million barrels a day of imported oil by 1990.*

Add to this the efficiency of the gas pipeline network, the cost of gas to the consumer, the environmental benefits of the gas energy system—and gas emerges as one of our most sensible energy options.

Gas energy: It can contribute to the stability of the dollar, and our continuing fight to lower oil imports.



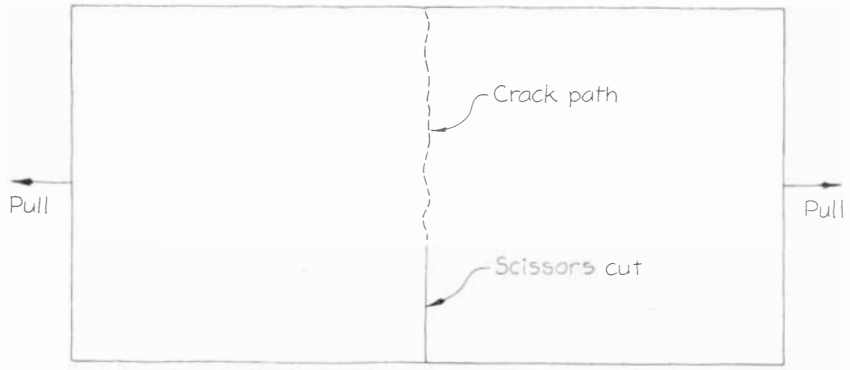
Projected start-up date	New sources of gas energy	Barrels of oil replaced per day	
		At outset	Potential 1990
1980	Mexican Gas Pipeline	50,000	300,000
1980	Canadian Gas	225,000	500,000
1983	Alaskan/Pacific LNG	170,000	1,000,000†
1984	Alaskan Gas Pipeline	450,000	800,000
1985	Coal Gasification	70,000	300,000
1990	New Technologies Unconventional sources of methane gas	900,000	900,000

† Includes liquefied natural gas from additional sources

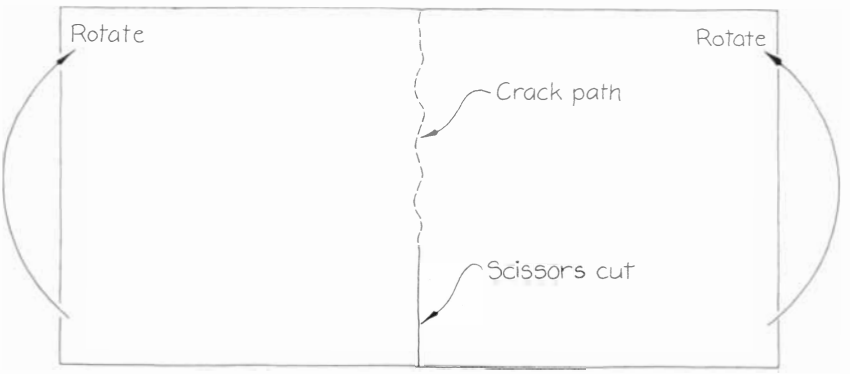
Barrels of imported oil per day that could be replaced by natural and synthetic gas.

Gas: The future belongs to the efficient.

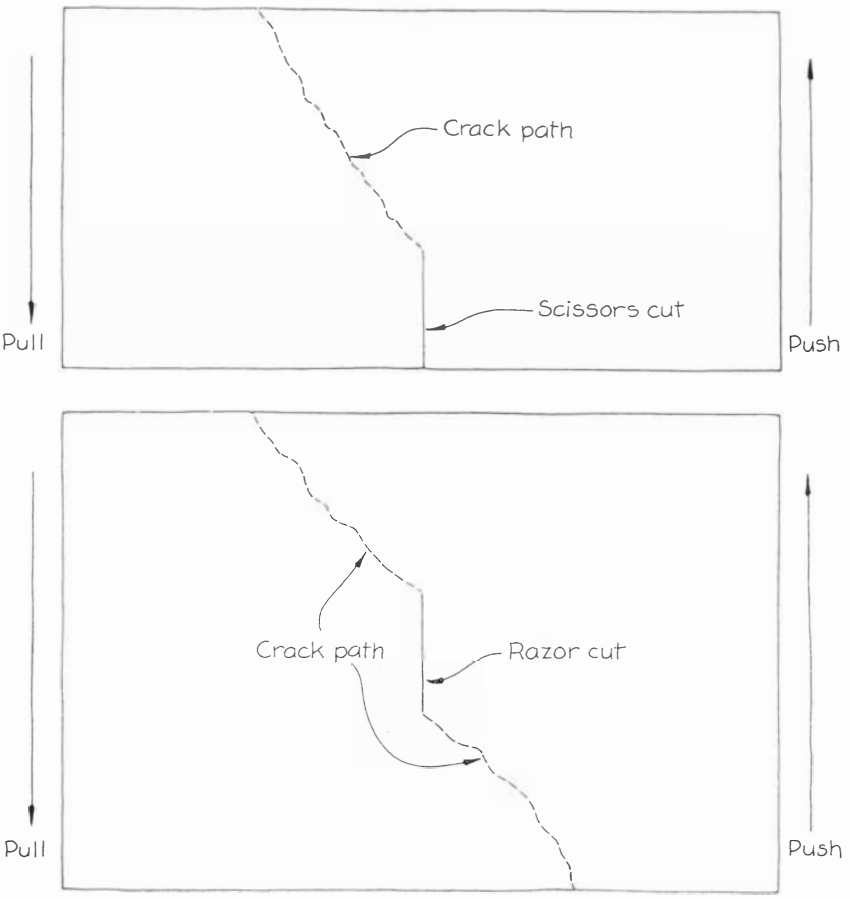
*For more information, write to American Gas Association, Department 1114-19 SC, 1515 Wilson Blvd., Arlington, VA 22209



Mixed-mode fracturing, Mode I: pure tensile stress



Mode I: pure bending stress



Mode II: the propagation of a shear crack

Ore., has seen falling snags fracture much the way chimneys do. Highly decayed snags sometimes break in half, and the two pieces land on opposite sides of the stump, presenting a serious problem for the logger trying to decide which way to run. Engels has also seen trees with much green foliage bend as they fall. In some instances the tree straightens rapidly as it nears the ground, so that the upper half reaches the ground before the lower half.

After reading my description (in the same article) of how pencil points break Gerald R. Martin of Napa, Calif., sent a description of a strange effect chemistry students first noticed with glass stirring rods. If the rods are dropped, they often break into three pieces of equal length and a small number of chips. This breaking pattern was dubbed the "pi rule," since the rods break into approximately 3.1 pieces. Martin and other mathematicians generalized the result (as mathematicians are likely to do) to include lengths of fresh chalk. Are there other examples of the pi rule?

I tested the rule by dropping pieces of chalk onto the floor after marking each piece so that I could reconstruct the fragments and locate the point of impact. Although the rule was not obeyed exactly, the chalk usually broke into three sections (rarely of the same length) with fractures similar to those in pencil points that have been broken. My work drew several comments from my colleagues at Cleveland State University. (A professor repeatedly dropping pieces of chalk on the floor is bound to attract attention.) Karl Casper discovered that if the chalk is reduced in length by a third, it will usually break into two sections rather than three. It could be that a full-length piece first snaps about a third of the way along its length and then the remaining two-thirds of a piece snaps approximately in half when the top section hits a moment later. You might try high-speed photography to find out whether my speculation is right or whether the entire piece of chalk vibrates and breaks in two places simultaneously.

Another example of fracturing, which seems at first to be unrelated, was sent to me by J. G. Nandris of the University of London. In about 2000 B.C. a huge menhir (an upright monolith), Le Grand Menhir Brisé, was erected at what is now Carnac in France. The stone was some 70 feet high and weighed more than 300 tons. How men with primitive tools could fashion and then erect such a stone is not understood. The stone stood until about 700 years ago, when it toppled and broke into four pieces. The upper half now lies in three pieces aligned east and west. The lower half is aligned toward the northwest, apparently after having fallen in a direction almost opposite to that taken by the upper half.

Why did the stone fall? Was it broken

SCIENCE/SCOPE

Fiber-optic cables allow sensitive information to be transmitted across unsecured areas without the need for encryption. A new communication link built by Hughes for the U.S. government connects command centers to computers in separate buildings with hair-thin glass fibers. Alarms are triggered anytime there is an unauthorized attempt to obtain access to the link. The new link, which may transmit at rates as high as 100 million bits per second, is less costly and allows considerably more system flexibility than protected wire systems using conventional security techniques.

Improved processing techniques have created silicon solar cells that produce 15 percent more power than their predecessors in converting sunlight into electric power for spacecraft operations. Advanced cells developed by Spectrolab, a Hughes subsidiary, generate 20 milliwatts of power per square centimeter. The increase in power is attributed to improving the thermal and optical properties of the cell and to a chemical etching method that sculpts microscopic tetrahedrons (pyramids) on the cell's surface. The result provides a dull-surfaced cell that abosorbs more sunlight than cells with polished surfaces.

An exotic chip that would alert a pilot when he has been detected by enemy radar promises to open a new arena in modern electronic warfare. The unique wafer, called an integrated optic spectrum analyzer (IOSA), would allow a pilot to prepare for a dogfight, turn on jamming equipment, or take any other appropriate action. The device works by having a surface acoustic wave device convert processed radar signals into sound waves. These sound waves interact with light from a tiny solid-state laser and cause the beam to bend toward a detector array made of charge-coupled devices. The amount of deflection indicates the frequency of the radar signal. The IOSA, only recently made possible by breakthroughs in diverse technologies, is being developed by Hughes for the U.S. Air Force.

Career growth opportunities exist at all levels at Hughes Support Systems for a variety of engineers qualified by degree or extensive work experience. They include systems engineers and software and hardware design engineers for major simulation and test equipment programs. Also, field engineering posts throughout the U.S. and the world offer travel, autonomy, and responsibility for the life cycle of Hughes electronic systems. For immediate action, phone collect (213) 670-1515, Ext. 5444. Or send your resume to Professional Employment, Dept. SE, Hughes Aircraft Company, P.O. Box 90515, Los Angeles, CA 90009.

Weather satellite pictures are worth at least $172 million in annual savings to industry and government, according to the National Oceanic and Atmospheric Administration. Among the beneficiaries are Florida citrus growers. Imagery from a NASA satellite is processed every half hour into a temperature map showing the southward progress of crop-killing frost during winter nights. The data allows growers to wait until the last moment to begin the costly heating of groves. The satellite's "camera" is a visible-infrared spin-scan radiometer built by the Santa Barbara Research Center, a Hughes subsidiary. This device makes pictures using either reflected sunlight or heat radiated from land, sea, and cloud tops.

by lightning or pulled down by peasants? Why did the two halves fall in opposite directions? Most likely the stone was felled by an earthquake that snapped it at its midpoint. When the upper half struck the ground, it broke into three pieces much like the chalk lengths I dropped. The fracture surfaces on all four pieces even appear to be similar to the chalk fractures. The lower half of the stone continued to ride the heaving ground (if indeed the stone was toppled by an earthquake) until it too was knocked down, but it happened to fall in a different direction from that of the upper half. Both the menhir and the earthquake that shattered it must have been awesome.

The type of fracturing common to pencil points, chimneys, chalk and even the menhir is called mixed-mode fracturing because it involves two general modes of fracture, as was explained to me recently by Anthony R. Ingraffea of Cornell University. The illustrations on page 202 show how to cut and tear sheets of paper to demonstrate these modes. A fracture of the type named Mode I occurs when the material is under either pure bending stress or pure tensile stress, neither of which will cause the two sides of the resulting crack to slide over each other. To demonstrate this phenomenon cut a slit in a sheet of paper and pull or rotate the ends as shown in the top two illustrations on page 202. The crack propagates in a straight line parallel to the scissors cut.

If the two faces of a crack are made to slide over each other without being separated, the fracture is of the type named Mode II. Cut two sheets of paper as shown in the bottom illustration on page 202 and then push and pull as required. This time the fracture propagates in a curved manner.

The cracks that propagate across the width of a falling chimney or a pencil point under stress are of mixed modes: bending (Mode I) and shearing (Mode II). Therefore the crack displays some of the fracture pattern of each mode. The direction in which the crack curves during the latter part of the propagation is different for chimneys and pencil points because of the direction of shearing. For example, if the chimney falls to the left, a clockwise shearing makes the crack curve downward as it travels across the chimney. In a pencil point the direction of shearing sends the crack curving upward, away from the point.

There was an error in the illustration of the amplifier circuit for the seismograph described in this department for July. The line from the .4-microfarad capacitor to the pin labeled "2" on the second gate should be connected to the line between the one-kilohm resistor and the 100-kilohm resistors.

BIBLIOGRAPHY

Readers interested in further explanation of the subjects covered by the articles in this issue may find the following lists of publications helpful.

MATHEMATICAL GAMES

ALGORITHMIC INFORMATION THEORY. G. J. Chaitin in *IBM Journal of Research and Development,* Vol. 21, No. 4, pages 350–359, July, 1977, and errata, Vol. 21, No. 5, page 496, September, 1977.

LAND-BASED INTERCONTINENTAL BALLISTIC MISSILES

STRATEGIC VULNERABILITY: THE BALANCE BETWEEN PRUDENCE AND PARANOIA. John D. Steinbruner and Thomas M. Garwin in *International Security,* Vol. 1, No. 1, pages 138–181; Summer, 1976.

THE MX MISSILE: AN ARMS CONTROL IMPACT STATEMENT. M. Callaham, B. T. Feld, E. Hadjimichael and K. M. Tsipis. Program in Science and Technology for International Security, Massachusetts Institute of Technology, March, 1978.

PROGRESS IN ARMS CONTROL? READINGS FROM *SCIENTIFIC AMERICAN.* Introductions by Bruce M. Russett and Bruce G. Blair. W. H. Freeman and Company, 1979.

THE CAUSES OF DIABETES

REPORT OF THE NATIONAL COMMISSION ON DIABETES TO THE CONGRESS OF THE UNITED STATES. United States Department of Health, Education, and Welfare. Government Printing Office, 1976.

RECEPTORS FOR PEPTIDE HORMONES: NEW INSIGHTS INTO THE PATHOPSYCHOLOGY OF DISEASE STATES IN MAN. Richard C. Eastman and Jeffrey S. Flier in *Annals of Internal Medicine,* Vol. 86, No. 2, pages 205–219; February, 1977.

TYPE I DIABETES MELLITUS. A. G. Cudworth in *Diabetologia,* Vol. 14, No. 5, pages 281–291; May, 1978.

VIRUS-INDUCED DIABETES MELLITUS: ISOLATION OF A VIRUS FROM THE PANCREAS OF A CHILD WITH DIABETIC KETOACIDOSIS. Ji-Won Yoon, Marshall Austin, Takashi Onodera and Abner Louis Notkins in *The New England Journal of Medicine,* Vol. 300, No. 21, pages 1173–1179; May 24, 1979.

SHAPE-MEMORY ALLOYS

NITINOL HEAT ENGINES. Ridgway Banks in *Shape Memory Effects in Alloys,* edited by Jeff Perkins. Plenum Press, 1975.

SHAPE MEMORY EFFECTS AND APPLICATIONS: AN OVERVIEW. Walter S. Owen in *Shape Memory Effects in Alloys,* edited by Jeff Perkins. Plenum Press, 1975.

THE NEUTRAL THEORY OF MOLECULAR EVOLUTION

THEORETICAL ASPECTS OF POPULATION GENETICS. Motoo Kimura and Tomoka Ohta. Princeton University Press, 1971.

MOLECULAR POPULATION GENETICS AND EVOLUTION. Masatoshi Nei. North-Holland Publishing Co., 1975.

STATISTICAL STUDIES ON PROTEIN POLYMORPHISM IN NATURAL POPULATIONS, I: DISTRIBUTION OF SINGLE LOCUS HETEROZYGOSITY. Paul A. Fuerst, Ranajit Chakraborty and Masatoshi Nei in *Genetics,* Vol. 86, No. 2, Part 1, pages 455–483; June, 1977.

LACK OF EXPERIMENTAL EVIDENCE FOR FREQUENCY-DEPENDENT SELECTION AT THE ALCOHOL DEHYDROGENASE LOCUS IN *DROSOPHILA MELANOGASTER.* Hiroshi Yoshimaru and Terumi Mukai in *The Proceedings of the National Academy of Sciences of the United States of America,* Vol. 76, No. 2, pages 876–878; February, 1979.

MODEL OF EFFECTIVELY NEUTRAL MUTATIONS IN WHICH SELECTIVE CONSTRAINT IS INCORPORATED. Motoo Kimura in *The Proceedings of the National Academy of Sciences of the United States of America,* Vol. 76, No. 7, pages 3440–3444; July, 1979.

PRIMEVAL GALAXIES

ARE YOUNG GALAXIES VISIBLE? R. B. Partridge and P. J. E. Peebles in *Astrophysical Journal,* Vol. 147, No. 3, pages 868–886; March, 1967.

THE OPTICAL APPEARANCE OF MODEL PRIMEVAL GALAXIES. David L. Meier in *Astrophysical Journal,* Vol. 207, No. 2, pages 343–350; July 15, 1976.

YOUNG GALAXIES, QUASARS AND THE COSMOLOGICAL EVOLUTION OF EXTRAGALACTIC RADIO SOURCES. M. S. Longair and R. A. Sunyaev in *Radio Astronomy and Cosmology: International Astronomical Union Symposium No. 74,* edited by David L. Jauncey. D. Reidel Publishing Company, 1977.

RECENT THEORIES OF GALAXY FORMATION. J. Richard Gott III in *Annual Review of Astronomy and Astrophysics,* Vol. 15, pages 235–266; 1977.

THE ECOLOGY OF THE AFRICAN DUNG BEETLE

WHY HAVE SOME ANIMALS EVOLVED TO REGULATE A HIGH BODY TEMPERATURE? Bernd Heinrich in *American Naturalist,* Vol. 111, No. 980, pages 623–640; July–August, 1977.

ENDOTHERMY IN AFRICAN DUNG BEETLES DURING FLIGHT, BALL MAKING, AND BALL ROLLING. George A. Bartholomew and Bernd Heinrich in *The Journal of Experimental Biology,* Vol. 73, No. 2, pages 65–83; April, 1978.

THE QUANTUM THEORY AND REALITY

CAN QUANTUM-MECHANICAL DESCRIPTION OF PHYSICAL REALITY BE CONSIDERED COMPLETE? A. Einstein, B. Podolsky and N. Rosen in *Physical Review,* Vol. 47, No. 10, pages 777–780; May 15, 1935.

ON THE EINSTEIN PODOLSKY ROSEN PARADOX. J. S. Bell in *Physics,* Vol. 1, No. 3, pages 195–200; November/December, 1964.

FOUNDATIONS OF QUANTUM MECHANICS. Edited by B. d'Espagnat. Academic Press, 1971.

EXPERIMENTAL CONSEQUENCES OF OBJECTIVE LOCAL THEORIES. John F. Clauser and Michael A. Horne in *Physical Review D,* Vol. 10, No. 2, pages 526–535; July 15, 1974.

USE OF INEQUALITIES FOR THE EXPERIMENTAL TEST OF A GENERAL CONCEPTION OF THE FOUNDATIONS OF MICROPHYSICS. B. d'Espagnat in *Physical Review D,* Vol. 11, No. 6, pages 1424–1435, March 15, 1975; Part 2, Vol. 18, No. 2, pages 349–358, July 15, 1978.

CONCEPTUAL FOUNDATIONS OF QUANTUM MECHANICS. Bernard d'Espagnat. W. A. Benjamin, Inc., 1976.

BELL'S THEOREM: EXPERIMENTAL TESTS AND IMPLICATIONS. John F. Clauser and Abner Shimony in *Reports on Progress in Physics,* Vol. 41, No. 12, pages 1881–1927; December, 1978.

A NEOLITHIC AND IRON AGE SITE ON A HILLTOP IN SOUTHERN ENGLAND

IRON AGE COMMUNITIES IN BRITAIN. Barry Cunliffe. Routledge & Kegan Paul, 1974.

THE IRON AGE IN LOWLAND BRITAIN. D. W. Harding. Routledge & Kegan Paul, 1974.

HILLFORTS: LATER PREHISTORIC EARTHWORKS IN BRITAIN AND IRELAND. Edited by D. W. Harding. Academic Press, 1976.

THE AMATEUR SCIENTIST

THE OPTICS OF FLAMES. Felix J. Weinberg. Butterworth & Co., 1963.

FLAME STRUCTURE. R. M. Fristrom and A. A. Westenberg. McGraw-Hill Book Company, 1965.

FLAMES: THEIR STRUCTURE, RADIATION AND TEMPERATURE. A. G. Gaydon and H. G. Wolfhard. Chapman and Hall, Ltd., 1970.

Only the best go to the Olympics.

Nothing in the world of competitive sport can match the Olympic challenge. It is a challenge that demands not only the best in human athletic achievement, but a determination that can be summoned up to overcome seemingly impossible obstacles. Yet with all the talent, skill and dreams the Olympic Games focus into crystal clarity for a brief instant, there can be only a few who wear the gold.

For Peggy Fleming and Jean-Claude Killy the intensity of their gold-medal-winning performances on the ice and the slopes passed through them for a few moments of heart-stopping action most of us never feel in a lifetime. But the memories of the day live for them forever. In photographs.

It is because of the vital importance of the lasting visual record of these events that Canon has been selected Official 35mm Camera of the 1980 Olympic Winter Games. Under conditions of utmost severity, in a situation that decries compromise, Canon photographic equipment will be expected to deliver images that comply with one unyielding standard. They must be the best attainable.

Canon's support for the 1980 Olympic Winter Games goes far beyond the intimate sorcery of eye, hand and camera. It extends to every aspect of the photographic obligations that the Games entail. Supporting photographers whose livelihood depends on the images they record for posterity. With professional service and repair, systems support and supplementary or emergency equipment. Standing behind our commitment to being best, by offering the best assistance money, skill and human dedication can provide.

The quality standard for all Canon photographic products is something you may not see on the outside, but you'll come to appreciate as the years go by. It's the big difference between Canon cameras and others that seem to offer equivalent performance. And it's something that simply can't be faked.

It's inevitable that considering Olympian achievement calls to mind superlative statements. At Canon, we don't use superlatives lightly. We take being "best" very seriously.

And we'll be at Lake Placid to prove it.

THE OFFICIAL 35MM CAMERA OF THE 1980 OLYMPIC WINTER GAMES

Canon®

Canon USA, Inc., 10 Nevada Drive, Lake Success, New York 11042 · 140 Industrial Drive, Elmhurst, Illinois 60126 · 6380 Peachtree Industrial Blvd., Norcross, Georgia 30071 · 123 Paularino Avenue East, Costa Mesa, California 92626 · Bldg. B-2, 1050 Ala Moana Blvd., Honolulu, Hawaii 96814 · Canon Optics & Business Machines Canada, Ltd., Ontario

The new 1980 Honda Civic. We did it all over again.

The Honda Civic. It is the car that has in seven short years brought Honda to the forefront as a designer and builder of fine automobiles.

More important, it is the car whose engineering achievements helped make it the darling of a new generation of energy conscious Americans.

But, although the Civic has been an extraordinary success, history cannot stand still. Technology cannot wait while we rest on our laurels.

So, from the wheels up, the 1980 Honda Civic has been completely restyled. If you passed the old Civic by because of its size, the new Civic has over thirteen percent more interior space, adding legroom and shoulder room. It has twenty percent more window area for better visibility. It has improved suspension and a longer wheelbase for a smoother ride. All this without adding so much as an inch to the overall length of the car. Remarkable!

Since we know a good thing when we build it, the 1980 Honda Civic has the same simple layout that made our first Civic so widely admired—and copied. You'll still find such features as front-wheel drive, transverse-mounted engine, rack and pinion steering, and (on all hatchbacks) four-wheel independent strut suspension.

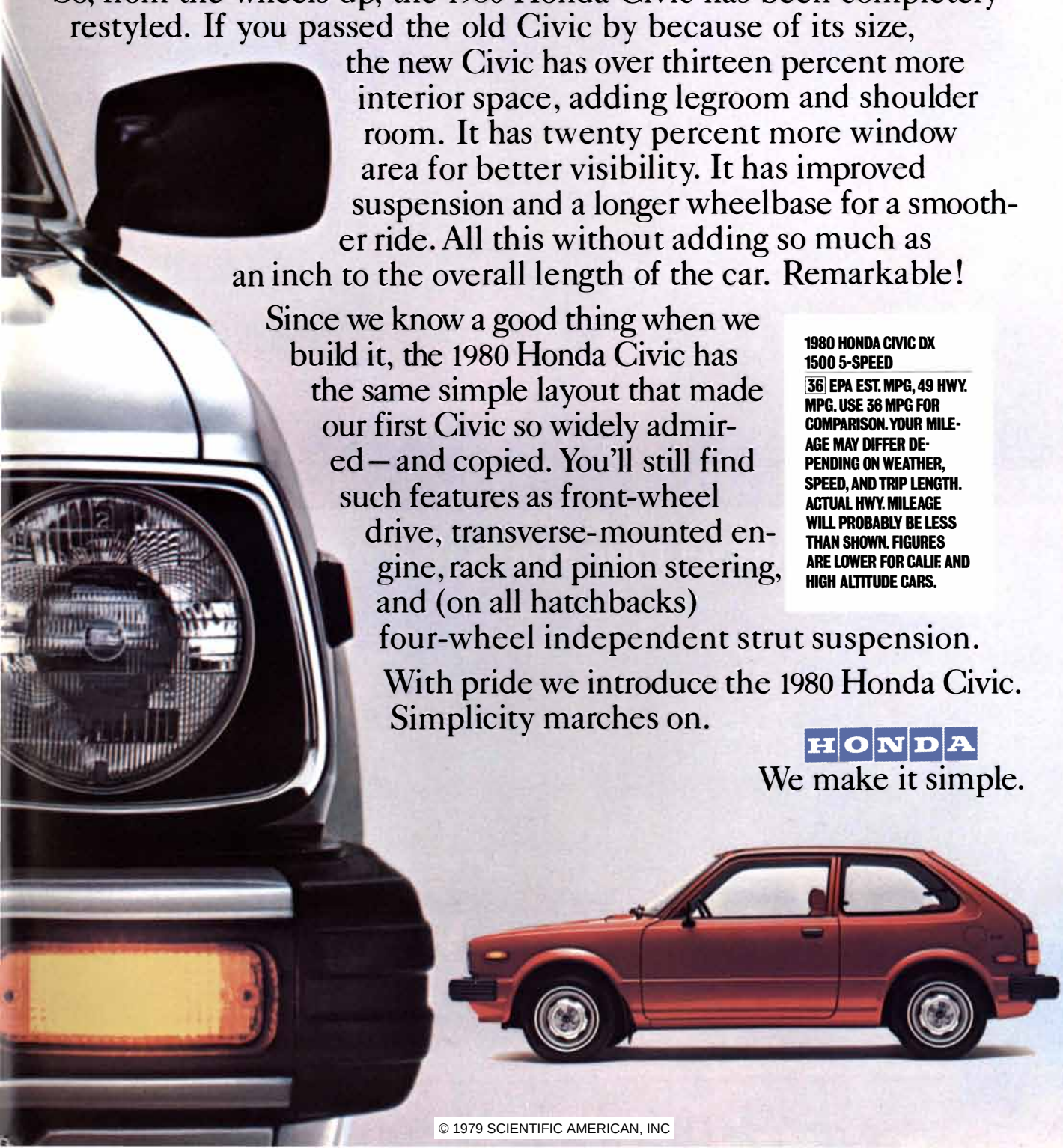
1980 HONDA CIVIC DX 1500 5-SPEED

36 EPA EST. MPG, 49 HWY. MPG. USE 36 MPG FOR COMPARISON. YOUR MILEAGE MAY DIFFER DEPENDING ON WEATHER, SPEED, AND TRIP LENGTH. ACTUAL HWY. MILEAGE WILL PROBABLY BE LESS THAN SHOWN. FIGURES ARE LOWER FOR CALIF. AND HIGH ALTITUDE CARS.

With pride we introduce the 1980 Honda Civic. Simplicity marches on.

HONDA

We make it simple.



Some people set their sights higher than others.

Seagram's V.O.

The symbol of imported luxury. Bottled in Canada.

Enjoy our quality in moderation.

Canadian whisky. A blend of Canada's finest whiskies. 6 years old. 86.8 Proof. Seagram Distillers Co., N.Y.C.