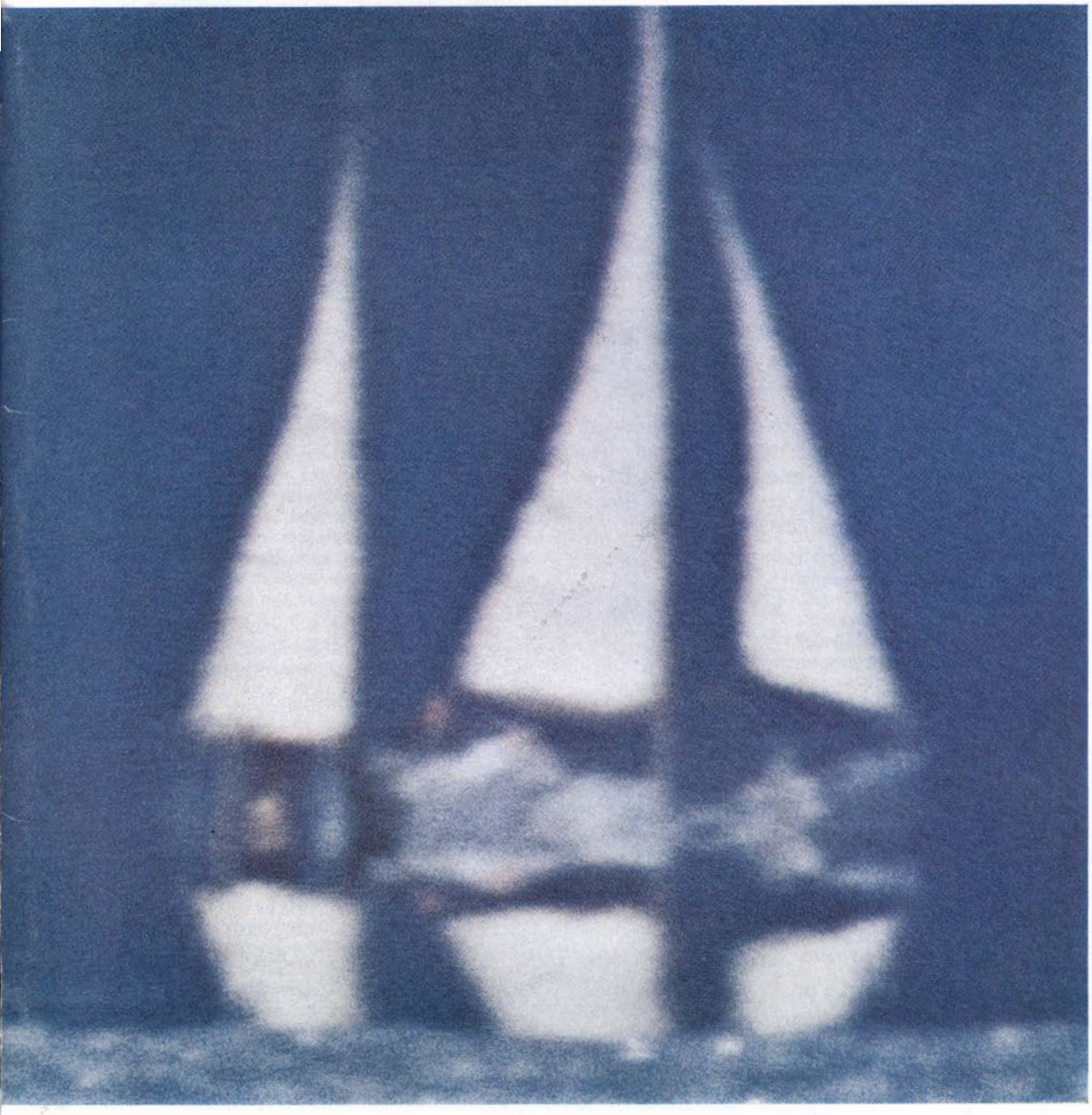
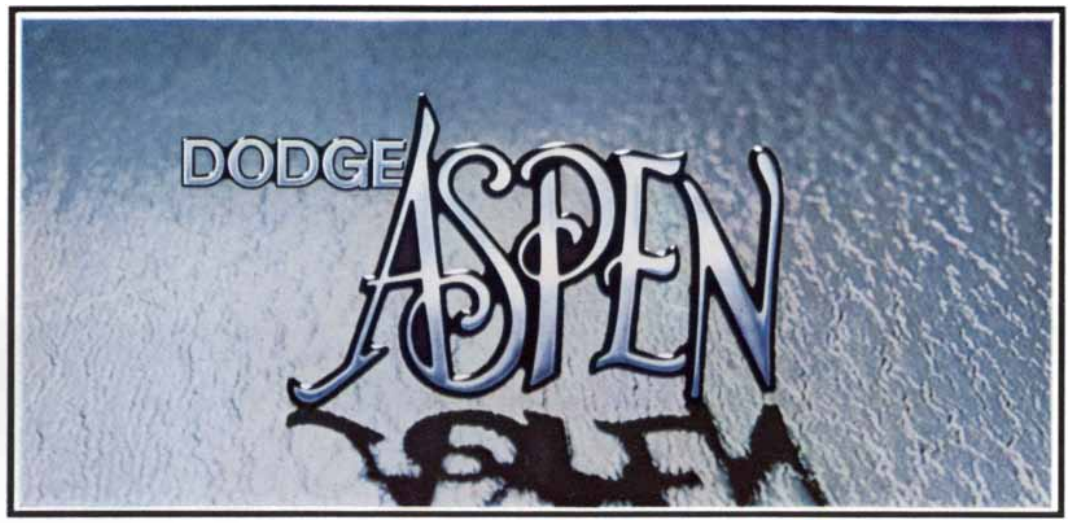
SCIENTIFIC AMERICAN



MIRAGES

$1.25

January 1976



We took the things you said and engineered them into

This new kind of car was developed for those people who wanted the economy and maneuverability of a small car, but did not want to sacrifice comfort and ride.

It is available in two versions: Dodge Aspen and Plymouth Volaré.

To give you a big-car ride, we invented a new suspension.

This new kind of small car called for a new suspension. So, our engineers invented an Isolated Transverse front suspension (Patent Pending). It's truly unique. Road shock is kept as far away from the driver and passengers as possible.

For greater quiet, we engineered two noise-reducing systems.

Most people take noise for granted in a small car. But not our engineers. They developed two special sound systems. A sealing system that minimizes wind noise and an insulation system that reduces road noise.

Dodge Aspen and Plymouth Volaré are roomier than many larger cars.

We reserved smallness for the outside. Our sedan models offer more total headroom and legroom than many bigger cars. The two-door coupes seat five passengers, comfortably. The four-door and the station wagon models seat six, comfortably.

No small-car scrimping on comfort.

Getting in and out of some small cars can be a problem. In these cars, our engineers developed wide door openings that make it easy. On some models, there's even available a seat that not only gives the driver a greater amount of comfort adjustment than conventional seats, but also a pull-strap seat-back release that lets you get in and out of the back seat of the two-doors using only one hand.

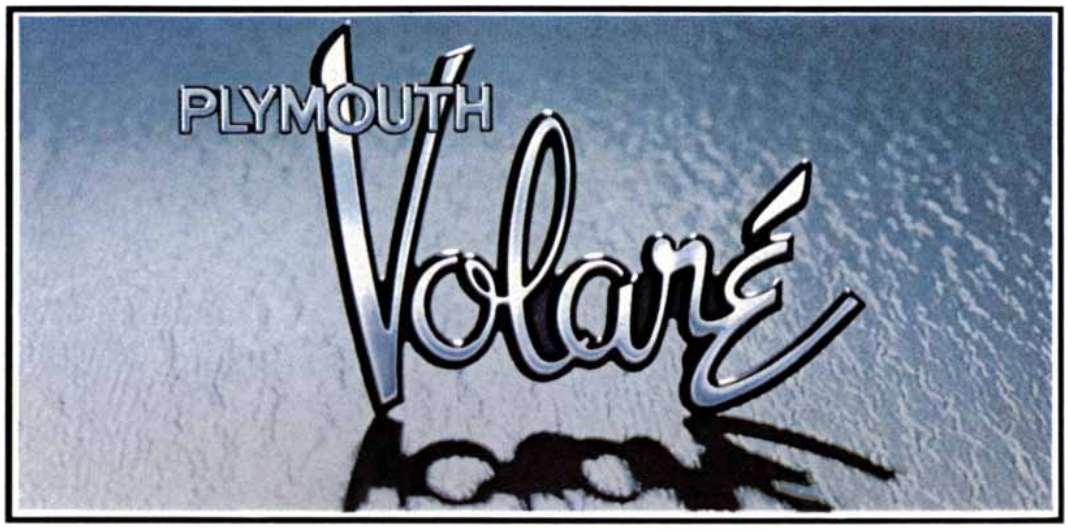
Plymouth Volaré and Dodge Aspen are designed to make it easier to see the road.

We evaluated the relationship between the driver, the steering wheel and the seat to put you in a comfortable position to see the road, and large windows and windshield make it even easier to see.

Aspen and Volaré achieved E.P.A. rated highway mileage of up to 30 m.p.g.

According to official E.P.A. estimated mileage results, Aspen and Volaré with Slant Six engines and manual transmissions got 30 on the highway and 18 in the city with their wagon models, and 27 on the highway and 18 in the city with their coupes and sedans.* We did this by designing the shape to reduce wind friction. Then we reduced weight to about a thousand pounds less than a full-size car. That all helps increase economy. Of course, your actual mileage may differ depending on your driving habits, the condition of your car and optional equipment.

*In California, see your dealer for E.P.A. mileage figures for California equipped cars.



were important in big cars a new kind of small car.

Plymouth Volaré and Dodge Aspen are covered by The Clincher.

This means that for the first 12 months of use, any Chrysler Motors Corporation dealer will fix, without charge for parts or labor, any part of your 1976 Dodge Aspen or Plymouth Volaré we supply (except tires) which proves defective in normal use, **regardless of mileage.** Of course, the owner is responsible for normal maintenance such as changing filters and wiper blades.

With three body styles, your choice is not restricted.

These new small cars are offered as a 2-door coupe, a 4-door sedan and a 4-door station wagon—in several versions. Our new-size wagon weighs up to 1,200 lbs. less than some bigger ones. That means you'll save gas and even though it's two feet shorter than a full-size wagon, you can carry 92% as much weight and up to 76% as much bulk. Compare what you can carry with competitive wagons.

SIZE	LENGTH OVER-ALL	TOTAL LOAD CAP.	CU.FT. CARGO SPACE	PAS-SENGERS
FORD COUNTRY SQUIRE	18.8 ft.	1,200 lbs.	94.6	6
CHEVELLE WAGON	18.0 ft.	1,100 lbs.	85.1	6
OUR NEW WAGON	**16.8 ft.**	**1,100 lbs.**	**71.9**	**6**
PINTO WAGON	14.9 ft.	850 lbs.	57.2	4

They're even easy to service.

We included all of our electronic advances. That means you can forget about replacing points and condensers. And, under normal driving conditions, there is no recommended transmission maintenance.

We wanted you to be able to order the options you wanted.

Many small cars are limited in the number of convenience options you can add. Ours can take all the important ones. For example, air conditioning engineered to provide comfort levels comparable to larger cars, power seats, power windows, electric door locks, tilt steering wheel, 4-speaker stereo, sun roof, power steering, power disc brakes, 50,000-mile battery, Sure-Grip axle, 4-speed manual overdrive transmission.

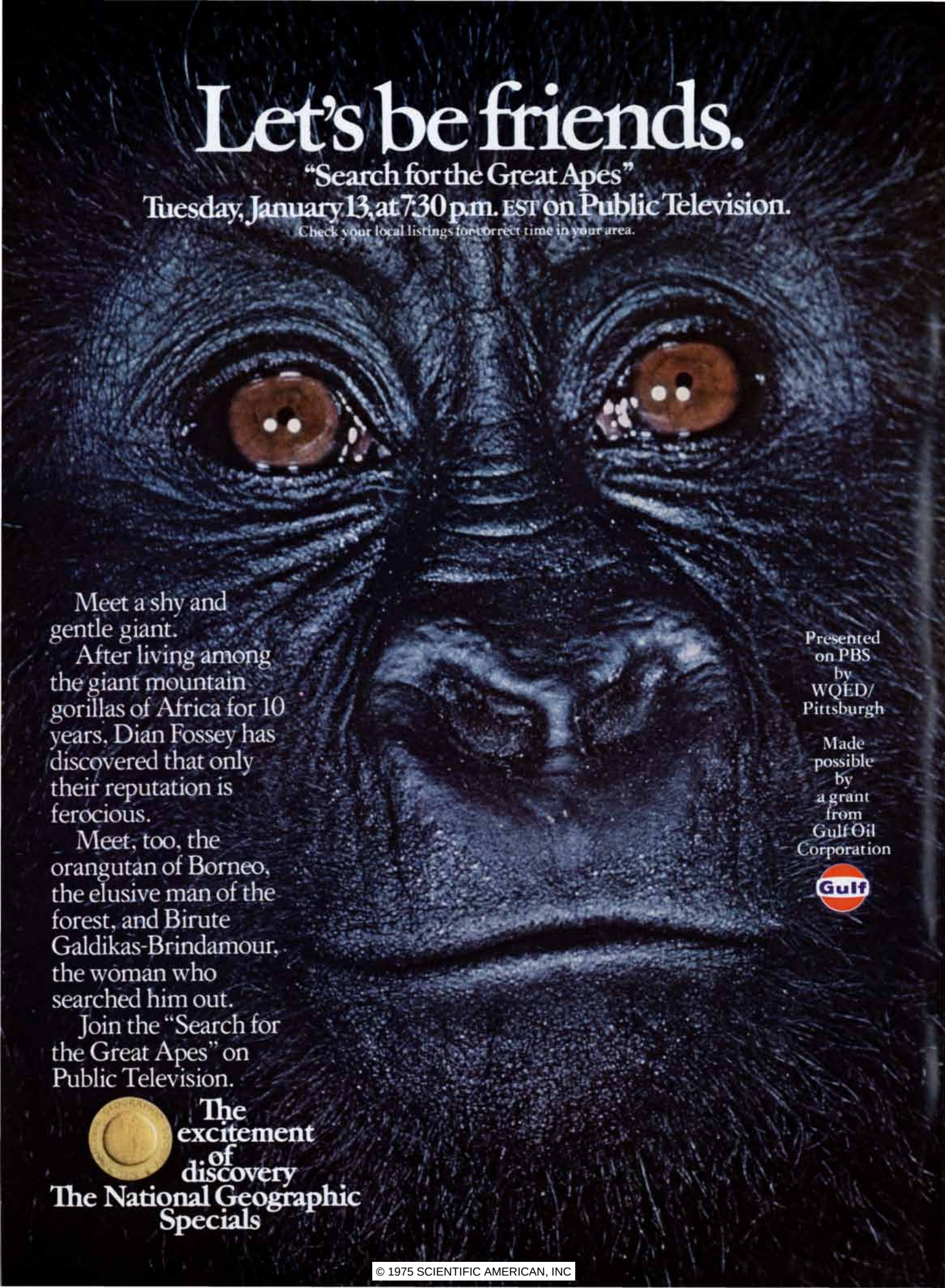
The price may be the biggest surprise of all.

You will find it hard to believe that you can get so much car, so much value for such a low price.

We made two versions . . . with two names . . . Dodge Aspen and Plymouth Volaré.

Aspen at your Dodge dealer's. Volaré at your Chrysler-Plymouth dealer's. See them now.

CHRYSLER CORPORATION



Established 1845 **SCIENTIFIC AMERICAN** January 1976 Volume 234 Number 1

ARTICLES

DEPARTMENTS

PUBLISHED MONTHLY BY SCIENTIFIC AMERICAN, INC., 415 MADISON AVENUE, NEW YORK, N.Y. 10017. SECOND-CLASS POSTAGE PAID AT NEW YORK, N.Y., AND AT ADDITIONAL MAILING OFFICES. AUTHORIZED AS SECOND-CLASS MAIL BY THE POST OFFICE DEPARTMENT, OTTAWA, CANADA, AND FOR PAYMENT OF POSTAGE IN CASH. SUBSCRIPTION RATE: $15 PER YEAR, U.S., POSSESSIONS AND CANADA; $18 PER YEAR, ALL OTHER COUNTRIES. EUROPEAN SUBSCRIPTION COPIES MAILED C/O PUBLICATIONS DISTRIBUTION SERVICE, ST. KATELIJNEVEST 14, B2000 ANTWERP, BELGIË.

THE COVER

The photograph on the cover appears at first glance to show a two-masted sailboat and its reflection in the water. In fact, however, the photograph shows a two-image mirage (see "Mirages," page 102). In the upper image the hull of the boat has vanished, a common characteristic of a two-image mirage. Moreover, a close look at the surface of the water reveals that it is too rough for the second, inverted image of the sailboat to represent a reflection. The mirage, which was photographed on Puget Sound, was caused by the fact that the temperature over the surface decreased with height, so that the atmosphere acted as a huge lens, bending the rays of light between the sailboat and the camera. A different kind of mirage would appear if the temperature of the air increased with increasing height above the surface.

THE ILLUSTRATIONS

Cover photograph by Alistair B. Fraser

Page	Source
22–31	Dan Todd
32	U.S. Geological Survey
34–35	Ilil Arbel
36–37	U.S. Geological Survey and Ilil Arbel
38–39	Jet Propulsion Laboratory
40	U.S. Geological Survey and Ilil Arbel
41–43	Jet Propulsion Laboratory
45–54	Allen Beechel
64–66	George V. Kelvin
71	Mark Ptashne
72	George V. Kelvin
73	Tom Maniatis (*left*), George V. Kelvin (*right*)
74–76	George V. Kelvin
80	Roy L. Caldwell, University of California at Berkeley
82–85	Tom Prentiss
86	L. Jackson, University of California at Berkeley
88	Tom Prentiss
91–100	Lorelle M. Raboni
102	Washington State Ferry System
103–104	Alistair B. Fraser, Pennsylvania State University
105–107	Adolph E. Brotman
108–109	Alistair B. Fraser, Pennsylvania State University
113	World Health Organization
114	Yale Medical Library, Clements C. Fry Collection
115	Burndy Library
116–117	National Library of Medicine
119–123	Ilil Arbel
125–128	Jerome Kuhl

QUALITY PAPERBACK BOOK SERVICE

The easy, economical way to buy the best paperbacks in print

Prices shown are publishers' list prices

More than 7000 distinguished new paperbacks are published each year. No wonder you miss so many good ones.

WE THINK YOU'LL ENJOY this easy way to acquire the best paperbacks published—even those hard-to-find books you've probably spent hours looking for—and you'll save money too. As a trial member of QPB, you'll receive our informative catalog, the *QPB Review*, each issue of which contains approximately one hundred titles to choose from. You'll also immediately qualify for our Free Book Plan. All QPB books are large-size, durable, handsome editions printed on quality paper with modern, readable typefaces. Join QPB today.

Any 3 books or sets for only $1 each

You simply agree to buy six more within the next year and as a member you'll immediately qualify for our Free Book Plan

HOW MEMBERSHIP WORKS

1. You will receive the *QPB Review* fifteen times a year—about every 3½ weeks. This informative catalog describes the Main Selection plus approximately one hundred Alternates. **The Main Selection is always offered at a 20% discount.**

2. If you want the Main Selection, do nothing. It will be shipped to you automatically. If you want one or more of the Alternates—or no book at all—just indicate your decision on the reply form always enclosed with the *Review* and return it by the date specified.

3. Free books. For every book or set you buy (exclusive of your three introductory choices for $1 each) you earn at least one Bonus Point, and the Bonus Points you accumulate entitle you to free books and sets. (You pay only shipping charges.)

4. Return privilege. If your *Review* is delayed in the mail and therefore you receive the Main Selection without having had ten days to notify us if you did not want it, you may return that Selection at our expense.

QUALITY PAPERBACK BOOK SERVICE, INC.
Middletown, Pennsylvania 17057

QUALITY PAPERBACK BOOK SERVICE, INC.
Middletown, Pennsylvania 17057 **6-QB170-1**

Please enroll me as a member of the Quality Paperback Book Service and send me the three books or sets whose numbers I've indicated in the boxes below. Bill me only $3 for all three, plus shipping. My only obligation is to purchase six more books or sets during the coming year, receiving a 20% discount on every Main Selection I take. My membership is cancelable any time after I buy the six additional books or sets. A moderate shipping charge is added to all shipments.

INDICATE BY NUMBER YOUR THREE BOOKS OR SETS [ ] [ ] [ ]

5

Mr.
Ms. _____
(Please print plainly)

Address _____ Apt. _____

City & State _____ Zip _____

Q-13

LETTERS

Sirs:

We should like to take this opportunity to reply to the comments of Martin Gardner in his critique of our NASA research report, "Development of Techniques to Enhance Man/Machine Communication" ["Mathematical Games," SCIENTIFIC AMERICAN, October, 1975].

The research pertains to the use of an automatic, solid-state machine that randomly selects from among four hidden targets while a subject tries to choose which target was selected. The machine provides immediate feedback as to the machine state, and rings a bell for correct subject responses, to allow him to try to use this feedback and reinforcement to improve his scores. Of the 147 volunteer subjects, six were identified whose learning performance was significant at the 0.01 level or better; the binomial probability of this occurring by chance is less than 0.004. At the other extreme, no subject had a negative learning slope of equal significance. In our report we took these preliminary findings to indicate that "there is evidence for paranormal functioning from our work with the ESP teaching machine." This evidence includes one subject who achieved scores at the $p < 10^{-6}$ level of significance in his 2,800 trials.

Gardner's major criticism of the experiments is based on an error in fact, namely his misconception of the manner in which data were collected. Subjects made runs of 25 trials. These trials were automatically printed on continuous fanfold paper tape, which carries a permanent record of every trial, machine state and trial number. After a series of eight to 10 runs the subject would bring the continuous fanfold to one of the experimenters for entry in the experimental log. The tapes were always delivered to us intact with all runs recorded. They were never torn into "disconnected bits and pieces," as Gardner asserts (implying that an individual could *post hoc* select which runs he turned in). Since we were interested in evidence of learning within each day's session, it was of particular importance to us to have the complete intact tape.

RUSSELL TARG

HAROLD PUTHOFF

Electronics and Bioengineering Laboratory
Stanford Research Institute
Menlo Park, Calif.

Scientific American, January, 1976; Vol. 234, No. 1. Published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017; Gerard Piel, president; Dennis Flanagan, vice-president; Donald H. Miller, Jr., vice-president and secretary; George S. Conn, treasurer; Arlene Wright, assistant treasurer.

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

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

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

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

NAME

NEW ADDRESS

OLD ADDRESS

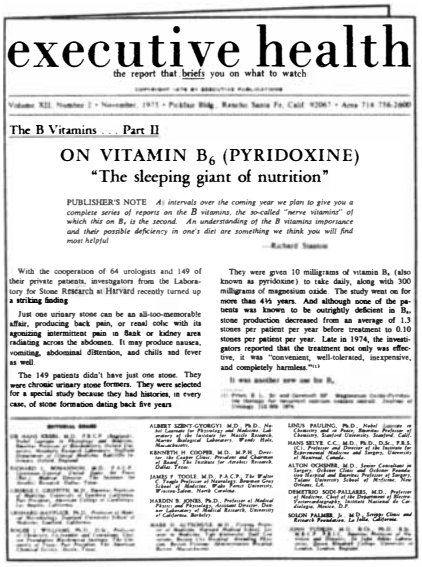
Sirs:

As the scientist who actually conducted the research, I should like to respond to Martin Gardner's critique of ESP teaching-machine research. . . .

Gardner's attack on the research was based on a number of misconceptions. First, data were not collected from tapes torn at the whim of subject and/or experimenter into "disconnected bits and pieces." Fanfold paper data tapes were torn off only at the end of a daily session; each session record was delivered to me intact. Second, Phase 3 research with Mr. Elgin was conducted as for all other subjects in Phase 3: I sat beside the subject and recorded scores by reading the total number of hits per run of 25 trials from the digital display on the face of the machine. Third, prior to the initiation of each Phase 3 session Mr. Elgin specified how many "practice" runs of 25 trials he wished to do. In no instance was such a decision capriciously based on the outcome of a run. Mr. Elgin's average score over the 2,500 experimental trials in Phase 3 had a probability of 5×10^{-4}, hardly "slight ESP" as Gardner asserts. We consider this satisfactory replication of his earlier work with a

YOU HAVE BUT ONE LIFE

— doesn't it make sense to find out how to take the best care of it you can?



executive health

the report that briefs you on what to watch

The B Vitamins . . . Part II

ON VITAMIN B_6 (PYRIDOXINE)

"The sleeping giant of nutrition"

PUBLISHER'S NOTE: *At intervals over the coming year we plan to give you a complete series of reports on the B vitamins, the so-called "nerve vitamins" of which this on B6 is the second. An understanding of the B vitamins importance and their possible deficiency in one's diet are something we think you will find most helpful*

—Richard Traum

With the cooperation of 64 urologists and 149 of their private patients, investigators from the Laboratory for Stone Research at Harvard recently turned up a striking finding

Just one urinary stone can be an all-too-memorable affair, producing back pain, or renal colic with its agonizing intermittent pain in flank or kidney area radiating across the abdomen. It may produce nausea, vomiting, abdominal distention, and chills and fever as well.

The 149 patients didn't have just one stone. They were chronic urinary stone formers. They were selected for a special study because they had histories, in every case, of stone formation dating back five years

They were given 10 milligrams of vitamin B_6 (also known as pyridoxine) to take daily, along with 300 milligrams of magnesium oxide. The study went on for more than 4½ years. And although none of the patients was known to be outrightly deficient in B_6, stone production decreased from an average of 1.3 stones per patient per year before treatment to 0.10 stones per patient per year. Late in 1974, the investigators reported that the treatment not only was effective, it was "convenient, well-tolerated, inexpensive, and completely harmless."[1]

It was another new use for B_6.

[1] Prien, E. L., Jr. and Gershwin, M.E. Magnesium Oxide-Pyridoxine Therapy for Recurrent Calcium Oxalate Urinary Stones. *The Journal of Urology*, 112: 844, 1974.

What keeps well people well? Medical men have long concentrated upon sick people and how to get them well . . . not upon well people and how to keep them well. Now many top research scientists are concentrating their efforts on *preventive medicine . . . how to keep well people well.*

May we suggest that you get the benefits of this new research for yourself? Do as so many thousands of executives do. Subscribe to Executive Health Report. The members of our Editorial Board are among the world's most distinguished authorities on *preventive medicine.* Their wise advice can help you not only live longer but enjoy those extra years!

EDITORIAL BOARD

SIR HANS KREBS, M.D., F.R.C.P. (England), *Nobel Laureate in Physiology and Medicine. Emeritus Professor of Biochemistry, Oxford University, Metabolic Research Laboratory, Nuffield Department of Clinical Medicine, Radcliffe Infirmary, Oxford, England.*

RICHARD L. BOHANNON, M.D., F.A.C.P., *Lieutenant-General, United States Air Force (Ret.); Medical Director, The Institute for Aerobics Research, Dallas, Texas.*

LEONARD HAYFLICK, Ph.D., *Professor of Medical Microbiology, Stanford University School of Medicine, Stanford, California.*

ROGER J. WILLIAMS, Ph.D., D.Sc., *Professor of Chemistry, Co-founder and Consultant, Clayton Foundation Biochemical Institute, The University of Texas; Past President, The American Chemical Society, Austin, Texas.*

ALBERT SZENT-GYORGYI, M.D., Ph.D., *Nobel Laureate for Physiology and Medicine, Laboratory of the Institute for Muscle Research, Marine Biological Laboratory, Woods Hole, Massachusetts.*

KENNETH H. COOPER, M.D., M.P.H., *Director, the Cooper Clinic; President and Chairman of Board, The Institute for Aerobics Research, Dallas, Texas.*

JAMES F. TOOLE, M.D., F.A.C.P., *The Walter C. Teagle Professor of Neurology, Bowman Gray School of Medicine, Wake Forest University, Winston-Salem, North Carolina.*

HARDIN B. JONES, Ph.D., *Professor of Medical Physics and Physiology, Assistant Director, Donner Laboratory of Medical Research, University of California, Berkeley.*

MARK D. ALTSCHULE, M.D., *Visiting Professor of Medicine, Harvard Medical School; Lecturer in Medicine, Yale University; Staff Consultant, Boston City Hospital; Attending Physician, Boston Veterans Administration Hospital, Boston, Massachusetts.*

LINUS PAULING, Ph.D., *Nobel Laureate in Chemistry and in Peace; Emeritus Professor of Chemistry, Stanford University, Stanford, Calif.*

HANS SELYE, C.C., M.D., Ph.D., D.Sc., F.R.S. (C), *Professor and Director of the Institute for Experimental Medicine and Surgery, University of Montreal, Canada.*

ALTON OCHSNER, M.D., *Senior Consultant in Surgery, Ochsner Clinic and Ochsner Foundation Hospital and Emeritus Professor of Surgery, Tulane University School of Medicine, New Orleans, LA.*

DEMETRIO SODI-PALLARES, M.D., *Professor of Medicine, Chief of the Department of Electrocardiography, Institute Nacional de Cardiologia, Mexico, D.F.*

SOLON PALMER, Jr., M.D., *Scripps Clinic and Research Foundation, La Jolla, California.*

JOHN YUDKIN, M.D., B.Ch., Ph.D., B.Sc., M.R.C.P., F.R.I.C., *Emeritus Professor of Nutrition and Dietetics, Sir John Atkins Laboratories, Queen Elizabeth College, University of London, London, England.*

Executive Health Report is not sold on newsstands but only by private subscription at $18 a year in the U.S.A. and its possessions. $19 per year in Canada and Mexico. All other countries $21 by surface mail, $24 by air mail. Individual reports (back issues) $1.50 per copy.

Subscribe now under this unusual introductory offer:

(1) **Your choice of any three of the reports listed below ($1.50 each) FREE!**

(2) **Your money back at anytime during the entire year** if you do not find our reports live up to your expectations.

Never forget: *"Men's lives are chains of chances"* but as Euripides saw clearly so long ago: *"Chance fights ever on the side of the prudent."* Your only insurance against "tomorrow" is what you do today. You have but one life . . . doesn't it make sense to find out how to take the best care of it you can?

Please study the reports listed here and **circle your three choices:**

Sir Hans Krebs, M.D.: On the overuse and misuse of medication. Too many patients are being made ill by drugs due to their own or their physician's carelessness.

James F. Toole, M.D.: On Strokes and "Little Strokes" . . . their causes and what you should know about them to help protect yourself!

Alton Ochsner, M.D.: On "The Chair Disease" . . . Why blood clots in your veins are a little-realized occupational hazard of desk-bound executives.

Dr. Linus Pauling, "On Vitamin C Against Disease." Why your health may be benefited in many ways by a proper intake of this essential vitamin.

Professor John Yudkin: On "This Slimming Business" . . . The truth about the prevention and cure of overweight!

Miles H. Robinson, M.D.: On Sugar and White Flour . . . the Dangerous Twins. How, with the best of intentions, we have managed to process natural foods into appetite-tempting, disease-breeding trouble makers . . .

Trace Minerals . . . Part I. On Chromium Deficiency and Atherosclerosis.

On The Arthritis Mystery. Can what you eat — or don't eat — make you arthritis-prone or arthritis-resistant?

Dr. Hans Selye: On Stress Without Distress. Your mind can make or break you!

Trace Minerals . . . Part II: On Your Danger From Cadmium in the water you drink and the food you eat — and how to protect yourself!

The B Vitamins . . . Part I. On B_{12} . . . Some puzzling physical and behavioral problems . . . and when to suspect a deficiency of this remarkable vitamin in your diet.

The B Vitamins . . . Part II. On Vitamin B_6 (Pyridoxine) "The Sleeping Giant of Nutrition."

Please use the coupon below, under our money-back at any time guarantee.

EXECUTIVE HEALTH, Pickfair Bldg., Rancho Santa Fe, Calif. 92067

Gentlemen: Enclosed is my check for $_____ for a year's subscription to Executive Health to start with this month's issue. I have circled the three $1.50 reports I am to receive free. It is understood that I am to get my money back if at any time during the entire year I become dissatisfied with your reports. In addition, I would appreciate your sending me a complete list of your 70 other reports already in print because among them may be some from which I might greatly benefit and would otherwise miss.

NAME (please print)_____

ADDRESS_____

STATE/COUNTRY_____ ZIP/POSTAL CODE_____ SA-4

printer as a recording device, in which $p < 10^{-6}$....

PHYLLIS M. COLE

Palo Alto, Calif.

Sirs:

The account by Targ and Puthoff of how data were collected for Phase 1 of their ill-fated test of Targ's ESP teaching machine differs radically from information provided by a reliable source who had access to the original records. Accordingly I wrote to Dr. Puthoff on October 16, asking that he allow *Scientific American* to arrange for a statistician acceptable to all parties to inspect the original tapes. There has been no reply.

Even in cases where a subject turned in an unbroken tape there is no way to determine if other runs have been omitted, because the machine's printer does not count *runs*. It counts only guesses from 1 to 25 within a run. I have no doubt that Phase 1 was designed with the expectation that complete tapes of all runs would be turned in by all 147 participants. I do question that it worked out that way. I am told that some participants, annoyed by the printer's clatter, stopped using it and turned in handwritten records. Until the original tapes can be checked by an outside expert I stand by my statement that they are in "bits and pieces."

Ms. Cole corrects an important error. I had said that in Phase 3, when Duane Elgin was "rehabilitated," his scores were printer-recorded. The facts are: Elgin scored significantly in Phase 1, when the printer was used. In Phase 2, when the computer provided safeguards against unconscious bias, his score dropped to chance. In Phase 3 no printed records of *any kind* were made. Ms. Cole sat beside him and kept a handwritten tally of his 7,000 guesses, of which 4,500 were decided (in advance) to be practice runs that were not counted.

We have here a classic instance of one firm believer being observed by another firm believer, both with a strong vested interest in a favorable outcome. The test was not even single-blind. None of the three investigators has made any comment on the central point of my article: that the results of Phase 2, the only phase with adequate safeguards, were entirely negative.

MARTIN GARDNER

Hastings-on-Hudson, N.Y.

You've invested a lot in your career.

Invest 80¢ a week more.

Think of the years and the effort you've spent to get where you are in business.

Now think of where you want to be.

And now read about a way to help you get there: The Wall Street Journal.

The Journal is a national, daily publication written and edited for men and women in business who want to be more valuable than they are now.

So The Journal is filled, page after page every business day, with solid business information, gathered from all over the world, concerning anything anywhere that may be pertinent, interesting, informative and useful to business-minded people. The key word is "useful."

That's why The Wall Street Journal is printed in nine different plants across the United States. So that the information in it will get into the hands of Journal subscribers early enough each day to be used by them that day.

Try The Journal on our special introductory offer. Just 80¢*a week will get it delivered to your home or office. Fill out and mail in the attached card. If someone else has already used the card, call this toll-free number 800-257-0300. (In New Jersey call 800-462-2000.) For new subscriptions only. Take us up on our 80¢-a-week offer, for any period from 13 weeks to a year.

80¢ a week. After all you've already invested in yourself, it's very little.

The Wall Street Journal

2A501 *Price good in United States and possessions and Canada.

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

JANUARY, 1926: "Professor Alfred L. Wegener, of the department of mineralogy and climatology at the University of Graz, says that millions of years ago the two Americas, as well as Europe, Asia, Australia, Antarctica and all the islands of the present-day world, were one continent centered around Africa. Tidal forces—the attraction of the sun and moon for the earth's solid mass (not ocean tides)—broke this super-continent up, and the pieces slowly dispersed in various directions, like the blocks of a great, flat cake of floating ice that is broken up by the waves. Some of these pieces, the present continents and islands, are drifting still, gradually moving away from the nuclear Africa. This theory is startling. To many it seems absurd. It may prove to be erroneous. It may gain final acceptance among geologists. At the present time it is strongly heterodox, but there is something about it that seems to captivate the interest of scientists."

"Thousands of surgical operations on cancers diagnosed early in their course have yielded positive proof that complete removal at such a stage prevents recurrence and not only greatly prolongs life but also may even yield a cure. In the attack on cancer physical agents also have been called into play. The apparently specific effects of high-voltage X rays may be applied to destroy cancer cells and to inhibit the growth of new cells, and radium in the form of the substance itself and of radium emanation may be used in certain instances to produce actual cures of cancer. The primary emphasis must be laid, however, on the fact that cancer must be seen and diagnosed early. After the tissues have been invaded and after the cancer has spread from the first point of origin to neighboring tissues and in some instances throughout the body, remedial attempts are usually hopeless."

"Incontestably the greatest blot upon the government and people of the United States is their treatment of the Indian. It is a story of brutal tyranny, broken pledges and cynical contempt. Strong words? Yes. But so far as may be, Congress must atone for the past by swift and adequate legislation for the unhappy descendants of those proud races whom our forefathers dispossessed."

"Dr. J. Allen Gilbert, Ph.D., of Portland, Ore., has been conducting a simple but effective test with the cooperation of *Scientific American*. It has been a test of the truth or falsity of communications through spirit mediums. More than 300 mediums have sent messages. Of that number 284 revealed what they believe to be a secret countersign agreed upon by Dr. Gilbert and Mrs. Gilbert before her death. Every one of the entire 284 was wide of the mark."

SCIENTIFIC AMERICAN

JANUARY, 1876: "The year 1875 will ever be a memorable date in the history of geographical discovery. Within the twelvemonth two of the most important questions of African geography have been settled, and in the far North the demonstration of an open water way between Europe and the countries drained by the great Siberian rivers is perhaps the most important addition to science that could be made in polar regions. The source of the Nile! For twenty centuries the boldest spirits have essayed its discovery, only to be turned back by insuperable obstacles. Its conquest waited for the plucky energy and the relentless push of Stanley. In the meantime Cameron has taken up the unfinished work of Livingstone and has crossed Africa through its most difficult and dangerous part, from Tanganyika to the mouth of the Congo. Less significant geographically but of far greater promise economically is Professor Nordenskjöld's discovery of an open passage by sea between Europe and northern Asia."

"Only 36 years ago the mails between England and Halifax, Nova Scotia, were conveyed in sailing vessels. The possibility of steam navigation had been demonstrated 12 or 14 years earlier, and coasting steamers were then actually engaged in trade. Thoughtful men had already begun to see that this steam coasting trade would probably be extended to the ocean. Mr., afterwards Sir Samuel, Cunard, threw himself heartily into the enterprise, came to England from Canada for the purpose of promoting it, and accepted the Government tender for the carriage of mails across the Atlantic. Today the Cunard fleet consists of 49 vessels having an aggregate of 90,506 tons. In the year 1874 the ocean vessels of the Atlantic fleet made 123 trips from Liverpool to New-York and Boston, and the same number in the opposite direction. The outward bound steamers carried 9,198 cabin passengers and 26,570 steerage passengers (mostly immigrants) from England to America."

"MM. Albert and Gaston Tissandier made a balloon ascent in Paris lately, and after a three hours' trip alighted near Illiers, about six miles from Paris. At 800 meters above the ground they entered a solid stratum of cloud 700 meters thick, the temperature being four degrees Centigrade below zero. At 1,500 meters altitude they passed through a succession of ice crystals, a galaxy of little hexagonal stars, which danced round the car and sparkled in the sun. These did not exist in the lower stratum of cloud, but were suspended in the atmosphere over an expanse from 150 to 200 meters thick. The temperature here was at zero, and higher still it was at six degrees. Cumuli clouds were perceived overhead at about 2,300 meters altitude, but the aeronauts did not go higher than 1,700 meters, about one mile."

"The prospects of the Philadelphia Centennial Exhibition are so gloomy just now that its friends are filled with alarm. Every effort has been made to stimulate the people of the United States to enthusiasm in its favor, but there is a very widespread feeling among potential exhibitors across the Atlantic that exhibitions are a nuisance, and that if the requested funds cannot be obtained for the completion of the undertaking from a people who refuse to respond to the most pressing invitations for subscription, the entire enterprise may not be worth the candle. As a last resource the promoters of the Centennial Exhibition are going to Congress for a vote of $1,500,000."

"The number of deaths in New-York (population 1,060,000) in the four weeks ending October 23 was 2,226. Of this number 932 were deaths of children under five years. The remaining deaths were distributed among the following causes: violence, 97; small-pox, 41; diphtheria, 186; scarlatina, 16; measles, 4; croup, 75; whooping-cough, 32; typhoid fever, 50; typhus fever, 4; puerperal diseases, 17; diarrhoeal diseases, 224; consumption, 318; lung diseases other than consumption, 230."

Scientific Programmable

$79.95

Sinclair's new Scientific Programmable is no ordinary calculator. It has only 19 keys—and a very low price—but its problem-solving capability far exceeds that of any ordinary scientific calculator. Sinclair's big plus is programmability—the ability to remember a calculation sequence of up to 24 steps entered directly from the keyboard. Once stored in the program memory, a calculation sequence can be recalled at the touch of a single key, and applied to new numbers to produce new results. So if you carry out repetitive calculations, the Scientific Programmable will save you key strokes and lots of time. It will also save you from errors, since the calculation sequence will be recalled exactly as you entered it, over and over again.

Programming Entering a calculation sequence into the program memory couldn't be easier. Press the BE key to tell the calculator to remember a sequence. Then key in the calculation almost exactly as you normally would. Use VAR at the points where you'll want the program to stop, so that you can enter new numbers or display partial results. During entry, the number of steps is displayed, so you won't exceed the program memory's capacity. When you've finished, press BE again to tell the calculator that the sequence is complete. The program is now available until you overwrite it with another program or until you switch the calculator off. Just enter new numbers, press EXEC, and get new results. Of course, the Scientific Programmable can be used as an ordinary scientific calculator. Even half-way through the execution of a program, you can stop, carry out a calculation from the keyboard, then press EXEC to continue execution of the program when you're ready.

Sinclair Program Library You don't have to be a programmer to use the Scientific Programmable's full problem-solving capability. Each calculator comes with a fully documented library of hundreds of programs to solve standard problems, and complete instructions on how to use them.

Applications The Scientific Programmable has many applications in areas of computation other than repetitive calculations. It can be used for the analysis of experimental data, the evaluation of integer functions, and with a variety of methods for the iterative solution of equations. They are dealt with in detail in the instruction book and Program Library. And, although the Scientific Programmable is no toy, it does play games—and wins!

Display 5 digit mantissa, 2 digit exponent. Number entry is floating decimal point and/or scientific notation; results in scientific notation.

Logic Reverse Polish Notation.

Functions Sine, cosine, arctangent (radians), log, alog (base 10), square root, reciprocal, change sign, clear/clear entry.

Memory Store, recall, exchange.

The Scientific Programmable is small enough to hold in your hand and big enough to use on your desk. It measures 6″ x 2-7/8″ x 1-1/4″, has non-slip rubber feet and a big green display. It uses a small, inexpensive 9v battery or the Sinclair AC adapter.

$79.95 includes battery, AC adapter, carrying case, instruction book and the Sinclair Program Library.

10 Day Trial Offer If the Scientific Programmable doesn't meet your requirements return it to us within 10 days for a prompt refund.

Warranty The Scientific Programmable is backed by Sinclair's full one-year parts and labor warranty.

sinclair™ **375 Park Avenue, New York, NY 10022**
(212) 688-6623

Please send me___Scientific Programmable(s) at $79.95 (NY residents add sales tax).

☐ Enclosed is my check for_____ ☐ Please charge my credit card account

Credit card_____Account#_____Expires_____

Name_____Signature_____

Address_____

_____

Mail to: Sinclair Radionics Inc.,375 Park Avenue, New York, NY 10022 Or call: (212) 688-6623

In England: Sinclair Radionics, Ltd., London Road, St. Ives, Huntingdon, Cambs. (0480) 64646

The amazing Mercedes-Benz 450SE. Take a good, close look. You've never seen anything like it.

Here's a walk-around checklist for one of the most important sedans ever engineered by Mercedes-Benz – or by anyone else. It'll help you see, firsthand, why the 450SE defies imitation ...and why it deserves your closest consideration.

1.
Start here. Look at the trunk space. An uncluttered 18.2 cubic feet. Spare tire is stored out of the way, flat under the floor, where it can also serve as an additional impact-absorbing device.

2.
Heated rear window. Automatic timer prevents wasted energy. Tinted glass all around. Standard equipment.

3.
Rain channels use airflow to divert rainwater from side and rear windows.

4.
Chrome strip/side protection molding is rubber-mounted to eliminate metal-to-metal contact.

5.
Interior is fully padded for safety and sumptuousness. Sealed against noise and heat. Four different sound-absorbing materials.

6.
Maximum-visibility windshield. All-around visibility in the 450SE totals 313°. Overlapping windshield wipers clear 73 percent of the entire windshield.

7.
Steel-belted radial tires. Standard equipment.

8.
Safety bumpers, rubber protected, hydraulic-regenerative. Front and rear.

9.
Halogen fog lights. Standard equipment.

14.
Aerodynamically ribbed rear lights use airflow to stay clear in foul weather or on dusty roads.

13.
Cockpit. Control panel has full instrumentation. Tachometer, quartz crystal clock, cruise control, AM/FM stereo radio, electrically operated windows are all standard equipment.

12.
Safety cone door locks. So strong that one of them can support the weight of the entire 450SE.

11.
Central locking system. Locks all doors, trunk *and* gas filler port at the turn of a key.

10.
The outside rearview mirror. Distortion-free, folds full forward *and* rear for pedestrian safety. Large, inside rearview mirror folds and breaks away on impact for driver and passenger safety.

The amazing Mercedes-Benz 450SE. Now, look at 8 things you don't see. You've still never seen anything like it.

You'll experience and appreciate some of them during your test drive. (Your Dealer can arrange it at your convenience.) The others are special Mercedes-Benz bonuses. All of them are standard equipment.

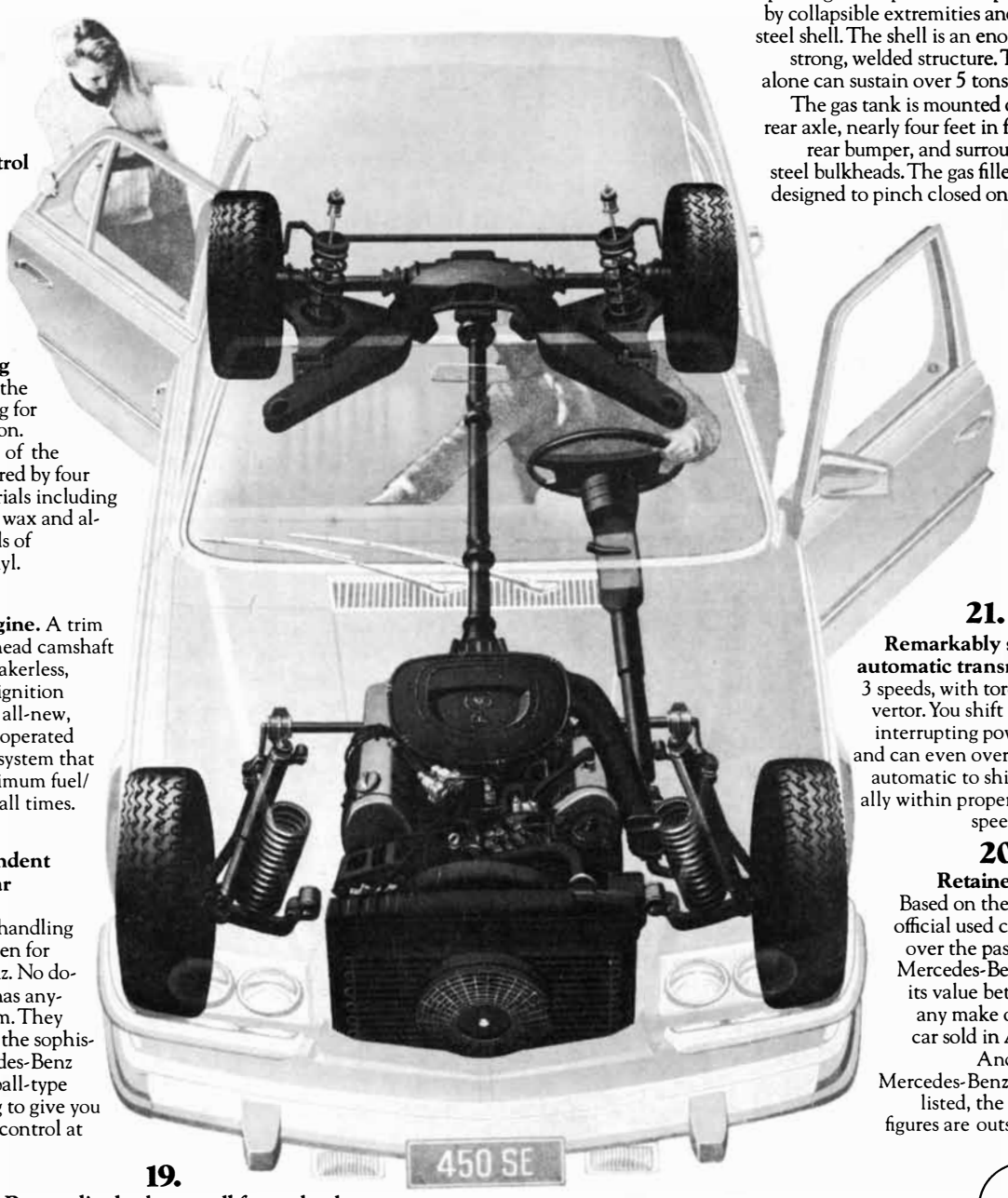
22.
Safety advances. The entire passenger compartment is protected by collapsible extremities and a rigid steel shell. The shell is an enormously strong, welded structure. The roof alone can sustain over 5 tons weight. The gas tank is mounted over the rear axle, nearly four feet in from the rear bumper, and surrounded by steel bulkheads. The gas filler pipe is designed to pinch closed on impact.

15.
Automatic Climate Control maintains a preselected temperature; dehumidifies, heats and cools.

16.
Undercoating extends up to the rocker molding for extra protection. The underside of the 450SE is buffered by four different materials including anti-corrosion wax and almost 24 pounds of plasticized vinyl.

17.
A unique engine. A trim 4.5 liter, overhead camshaft V-8 with a breakerless, transistorized ignition system and an all-new, mechanically operated fuel injection system that maintains optimum fuel/air mixture at all times.

18.
Fully independent front and rear suspension. They set new handling standards—even for Mercedes-Benz. No domestic sedan has anything like them. They combine with the sophisticated Mercedes-Benz recirculating ball-type power steering to give you extraordinary control at all times.

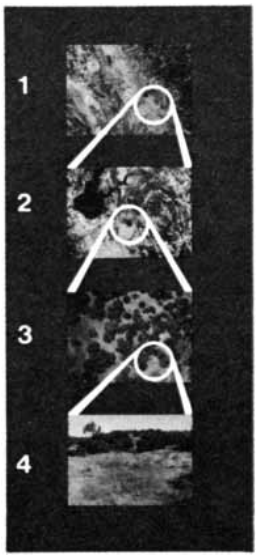


21.
Remarkably smooth automatic transmission, 3 speeds, with torque convertor. You shift without interrupting power flow and can even override the automatic to shift manually within proper engine speed ranges.

20.
Retained value. Based on the average official used car prices over the past 5 years, Mercedes-Benz holds its value better than any make of luxury car sold in America. And among Mercedes-Benz models listed, the 450SE's figures are outstanding.

19.
Power disc brakes on all four wheels. The brakes are ventilated—then further cooled by special turboblades. They are designed to stop you smoothly, securely time after time—without disconcerting fade.

Mercedes-Benz
Engineered like no other car in the world.



Broad-coverage, small-scale data from satellites (1) are evaluated to select more limited areas for the acquisition of more costly (per acre), larger-scale imagery data (2, 3) from high-, then low-flying aircraft. These data, in turn, are used to select specific ground plots (4) for on-site measurement of individual plants and detailed assessment of species mix, size, age, and condition.

New information gathering and processing techniques hold high hope for improved management of our renewable resources.

Every nine days, two satellites circle the earth, recording all surface data between 82° N to 82° S latitudes. Handling and interpreting these massive amounts of data are two of the keys to improved resource management. ESL Incorporated, using an HP 3000CX computer-based system, is helping to handle the data and process it quickly and economically into multilevel, cross-comparative forms.

The effective management of renewable natural resources over any but small areas depends directly on frequent, accurate, large-scale inventories, and the ability to compare them. Heretofore, adequate data were difficult, sometimes impossible, to acquire on a timely basis. And where such acquisition *was* possible, the sheer volume of data needed was costly and discouraged adequate interpretation—let alone comparison over time.

Now, with multispectral scanner data of almost any locale available from NASA's LANDSAT satellites, complemented by optimal analysis of small- and large-scale photography from high- and low-flying aircraft, and relevant ground data such as tree size and condition, a cost-effective alternative to existing information gathering and analysis procedures is possible.

To exploit this volume of data from complementary sources, ESL Incorporated of Sunnyvale, California, has developed a complete systems approach, the hub of which is an interactive digital image processing system (IDIMS), utilizing an HP 3000CX as the host computer for several data-handling and processing subsystems.

By applying techniques facilitated by the ESL system, thematic maps, inventory statistics, assessments, and predictions can be made and updated to meet a variety of objectives: timber volume estimates, wild land studies, censuses of people and animals, costal zone management, and range and forest inventory, to name a few.

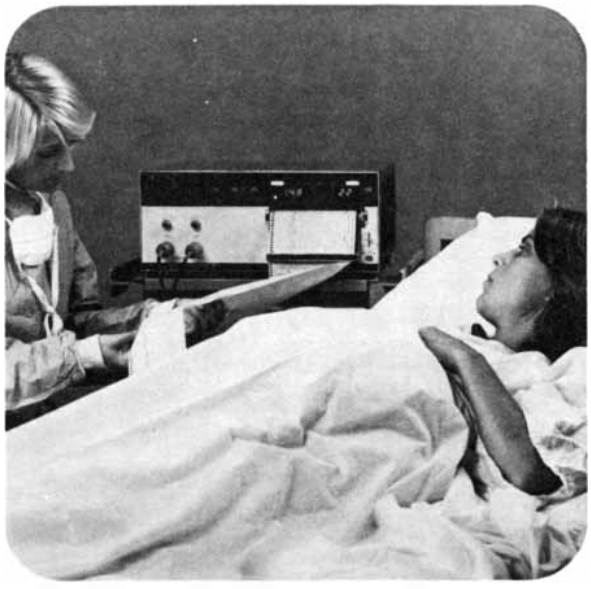
The advantages of the system are impressive. Spatial and spectral resolution characteristics of the digital data appear to be near optimum for computer mapping of surface vegetation. Geometric fidelity, spatial resolution, and computer compatability of sampling and analysis combine to simplify the generation of map overlays demarcating various boundaries—of administrative, management, ownership, and political areas—to meet diverse informational needs. Satellite coverage is both systematic and periodic, providing a broad, uniform data base and the opportunity for frequent updating of the resource information base through identification and mapping of surface changes.

The ESL system is used by NASA at the Ames Research Center and the Johnson Space Flight Center, and by the EROS Data Center maintained by the U.S. Department of the Interior in Sioux Falls, South Dakota. ESL maintains a service bureau in Sunnyvale, California, and another is planned for the Washington, D.C. area. For more information, write or call ESL Incorporated.

For more information on the versatile HP 3000CX, talk to us. Hewlett-Packard, 1504 Page Mill Road, Palo Alto, California 94304.

System prices for the HP 3000CX start at $99,500*.

This fetal monitor frees the obstetrical team to care for mother and unborn child.



Obstetricians have been relying on cardiotocograms, especially during difficult deliveries, for more than a decade of fetal monitoring. By continuously measuring fetal heart rate and maternal labor contractions, the fetal monitor (cardiotocograph) gives the obstetrical team valuable information about the stress of birth on the fetus, from which they can better manage labor and deliver healthier babies.

Virtually automatic in its operation, the new HP 8030A Cardiotocograph frees the delivery team from the task of managing the instrument for the much more important one of caring for the mother and unborn child. Specifically, the instrument can measure fetal heart rate by any of four methods (direct or abdominal ECG, ultrasound, or heart sound) and labor activity by internal or external methods. The nurse or physician need only plug the desired transducer into the appropriate connector—and the instrument automatically switches to the correct mode, calibrates the input signal, adjusts its sensitivity and makes the measurement accurately.

A built-in scope continuously shows the fetal input waveform and thus helps the obstetrical team to place the transducer for the best signal. The trace itself does not fade and its intensity adjusts automatically to changes in room light; it can also be "frozen" for inspection by pressing the STOP button.

The computed beat-to-beat heart rate and maternal uterine pressure are displayed simultaneously on large, clear digital readouts while status lamps clearly identify the type of measurement method in use. An acceptance lamp lights with every valid heart rate measurement, and an INOP indicator immediately signals if there is any problem in the ECG electrode or transducer.

The instrument checks its own operation and circuits at the touch of a button; its built-in silent recorder prepares a permanent cardiotocogram on smudgeless thermal paper; and it is easily configured to various monitoring modes simply by plugging printed circuit boards into the instrument chassis.

From any point of view, the 8030A meets the needs and preferences of the obstetrical team. Prices start at $6,250*.

For more information on these products, write to us, Hewlett-Packard, 1504 Page Mill Road, Palo Alto, California 94304.

HEWLETT hp PACKARD

Sales and service from 172 offices in 65 countries

Mail to: Hewlett-Packard, 1504 Page Mill Road, Palo Alto, CA 94304.
Please send me further information on

( ) HP 3000CX Computer System
( ) HP 8030A Cardiotocograph

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

*Domestic USA prices only.

00640

THE AUTHORS

H. A. BETHE ("The Necessity of Fission Power") has devoted much of his time and effort in the past few years to a study of the role of nuclear power in the context of the world energy problem, particularly since his retirement last July from active teaching at Cornell University. Born and educated in Germany, Bethe joined the Cornell faculty in 1935 after working at universities in Germany and Britain. His years at Cornell were interrupted by World War II, when he worked first at the Radiation Laboratory of the Massachusetts Institute of Technology and later as head of the theoretical-physics division of the Los Alamos Scientific Laboratory. His basic research on the theory of the atomic nucleus has won him many honors, among them the Nobel prize for physics, awarded in 1967 for his discovery in the late 1930's of the nuclear reactions that power the stars. Over the years he has also served as a consultant to both Government and industry on the design of nuclear power reactors. His current "one-man campaign for increasing our energy sources other than oil and gas," the latest in a series of activities that have involved him in national issues, stems from a conviction that "the general public is not well enough informed about science and technology and their role in our society. This allows any number of nuts to dispense misinformation couched in noble rhetoric. I only take on issues for which I feel my professional background qualifies me; I don't want to make a career of public advising."

MICHAEL H. CARR ("The Volcanoes of Mars") is chief of the astrogeologic-studies branch of the U.S. Geological Survey in Menlo Park, Calif. A graduate of the University of London, he came to the U.S. to attend Yale University, acquiring his M.S. and Ph.D. there in geochemistry. He then went to the University of Western Ontario, where he worked on the shock deformation of rocks. That led to an interest in impact craters and hence to lunar geology. Since joining the Geological Survey in 1962, he writes, "I have been involved in a variety of lunar and planetary projects. As leader of the Viking Orbiter imaging team I am currently involved in plans for extending our orbital photographic coverage of Mars when the Viking spacecraft arrives there this summer. On those rare occasions when I am home I like to rouse my neglected horse and go riding through the adjacent hills."

DAVID B. CLINE, ALFRED K. MANN and CARLO RUBBIA ("The Search for New Families of Elementary Particles") pursue their collaboration in spite of geography. Cline is professor of physics at the University of Wisconsin, where he received his Ph.D. in 1965. Mann is professor of physics at the University of Pennsylvania, to which he moved in 1949, two years after obtaining his Ph.D. from the University of Virginia. Rubbia is professor of physics at Harvard University; a native of Italy, he completed his graduate training at the University of Pisa and worked for eight years at the European Organization for Nuclear Research (CERN) before becoming a member of the Harvard faculty in 1970.

TOM MANIATIS and MARK PTASHNE ("A DNA Operator-Repressor System") did most of the research described in their article in Ptashne's laboratory at Harvard University. Maniatis, who now works at the Cold Spring Harbor Laboratory, went to Harvard as a National Institutes of Health Fellow in 1971 after obtaining his Ph.D. in molecular biology from Vanderbilt University. Before that he was an undergraduate at the University of Colorado. Ptashne, now professor of biochemistry and molecular biology at Harvard, was graduated from Reed College in 1961 and received his Ph.D. in molecular biology from Harvard in 1968.

ROY L. CALDWELL and HUGH DINGLE ("Stomatopods") first became acquainted with these fierce marine crustaceans in the course of a zoological field trip to Bermuda in the summer of 1966. At that time both of them were associated with the department of zoology at the University of Iowa, Caldwell as a graduate student and Dingle as an instructor in zoology. Caldwell, who obtained his Ph.D. from Iowa in 1969 for his work on the population dynamics of insects and the control of insect migration, has since joined the faculty of the University of California at Berkeley. He has continued his interest in the behavior of stomatopods, explaining that "although they are certainly not the easiest animals to study, the rewards have been great, since most of my work has had to be conducted in such lovely spots as Curaçao, Eniwetok and Thailand." Dingle, who is now professor of zoology at Iowa, is a graduate of Cornell

University and holds an M.S. and a Ph.D. from the University of Michigan. He also admits to indulging a "passion for travel," noting that "I am in addition an avid birdwatcher, a foible that is endured stoically by Caldwell and my family."

HARRY J. JERISON ("Paleoneurology and the Evolution of Mind") is professor of psychiatry and psychology at the University of California School of Medicine in Los Angeles. A native of Poland, his degrees are from the University of Chicago: a B.S. in 1947 and a Ph.D. in 1954. He was a research psychologist at the Aeromedical Laboratory from 1949 until 1957, when he was appointed director of the behavior-research laboratory at Antioch College, a position he held until he moved to U.C.L.A. in 1969. Jerison is the author of *Evolution of the Brain and Intelligence,* published by Academic Press in 1973.

ALISTAIR B. FRASER and WILLIAM H. MACH ("Mirages") are at Pennsylvania State University; Fraser is associate professor of meteorology and Mach is a research assistant in the department of meteorology. Fraser, who is Canadian, was graduated from the University of British Columbia and then began work as a weather forecaster with the Canadian government. He obtained his Ph.D. from the Imperial College of Science and Technology in London in 1968. "From my earliest days as a weather forecaster in Vancouver," he writes, "I have been particularly fascinated with meteorological information that could be deduced from naked-eye observation." Mach received his bachelor's and master's degrees at the University of Washington in 1969 and 1972 respectively. He is a Ph.D. candidate at Penn State.

WILLIAM L. LANGER ("The Prevention of Smallpox before Jenner") is Archibald Cary Coolidge Professor of History emeritus at Harvard University. His active association with Harvard extended over many years: he received his B.A. there in 1915, his M.A. in 1920 and his Ph.D. in 1923, and he was a member of the faculty from 1926 to 1964. A past president of the American Historical Society, Langer is perhaps best known as the editor of *An Encyclopedia of World History,* which was first published in 1940 and is now in its fifth edition, and the series "The Rise of Modern Europe," of which 18 volumes have been published to date.

WE'VE BEEN WORKING ON FOR THE LAST 100 YEARS.

And we're still working on it.

You see, the invention of the telephone didn't stop with Alexander Graham Bell. It just started.

Because the telephone is just the beginning of a telephone call. It's part of an intricate network of a trillion parts and nearly a billion miles of circuits.

To build this network, we at Bell Labs and Western Electric have long worked as a team with AT&T and your Bell telephone company.

As a result, America has the best telecommunications system in the world.

And the world has the benefits of such Bell System innovations as the transistor, the coaxial cable and direct distance dialing.

Working together, we've created entirely new communication systems. Like our latest switching machine that can route 550,000 calls an hour.

At the same time, we're constantly improving existing systems. Like tripling the capacity of our major microwave radio system in the last ten years.

Even the standard telephone that you probably think never changes has had virtually every major part improved since 1972.

In fact, we've made more than 2,500 improvements in your phone in the last twenty-five years.

Not just to make it look different. But to make it work better.

And to keep its cost down.

Improvements like these don't just happen.

The Bell System invests more than $750 million a year in research and development.

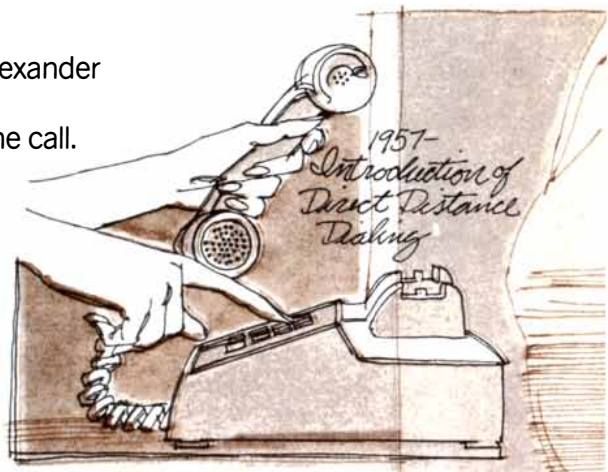
As an outgrowth, we receive an average of more than two patents every working day. And nearly half the things Western Electric will make this year didn't even exist four years ago.

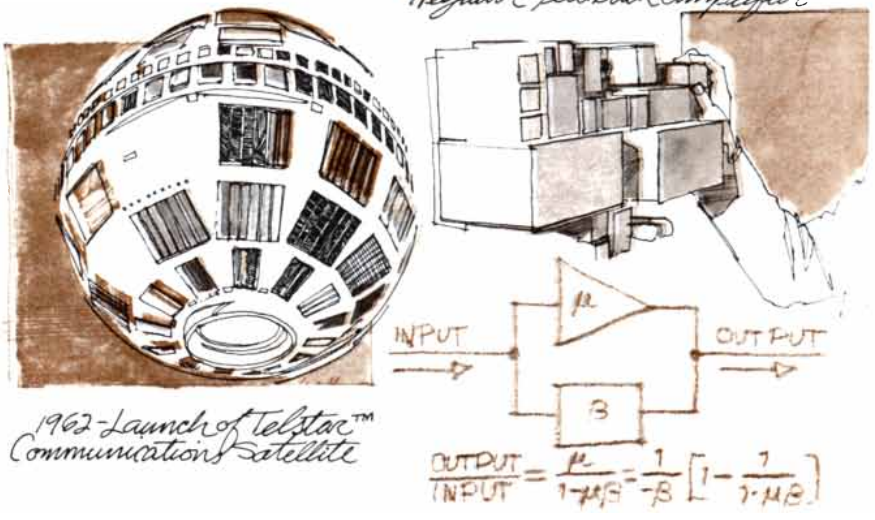
In the next 10 years, we plan to expand the capacity of the telephone network as much as we have in the past 100 years.

To keep this network operating and growing takes the innovative teamwork of Bell Labs and Western Electric.

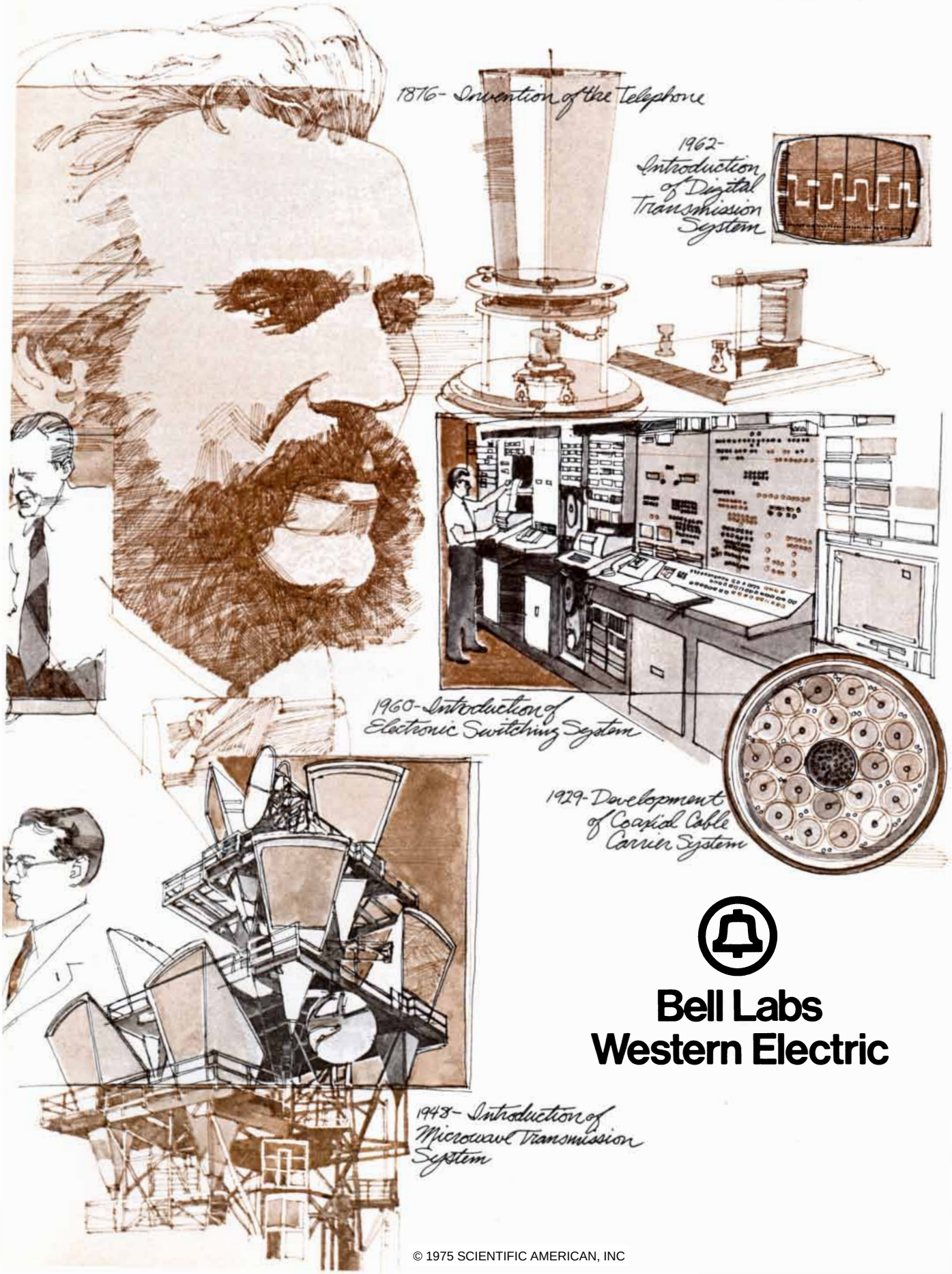
The kind of innovative teamwork that makes us say:

One Bell System. It works.

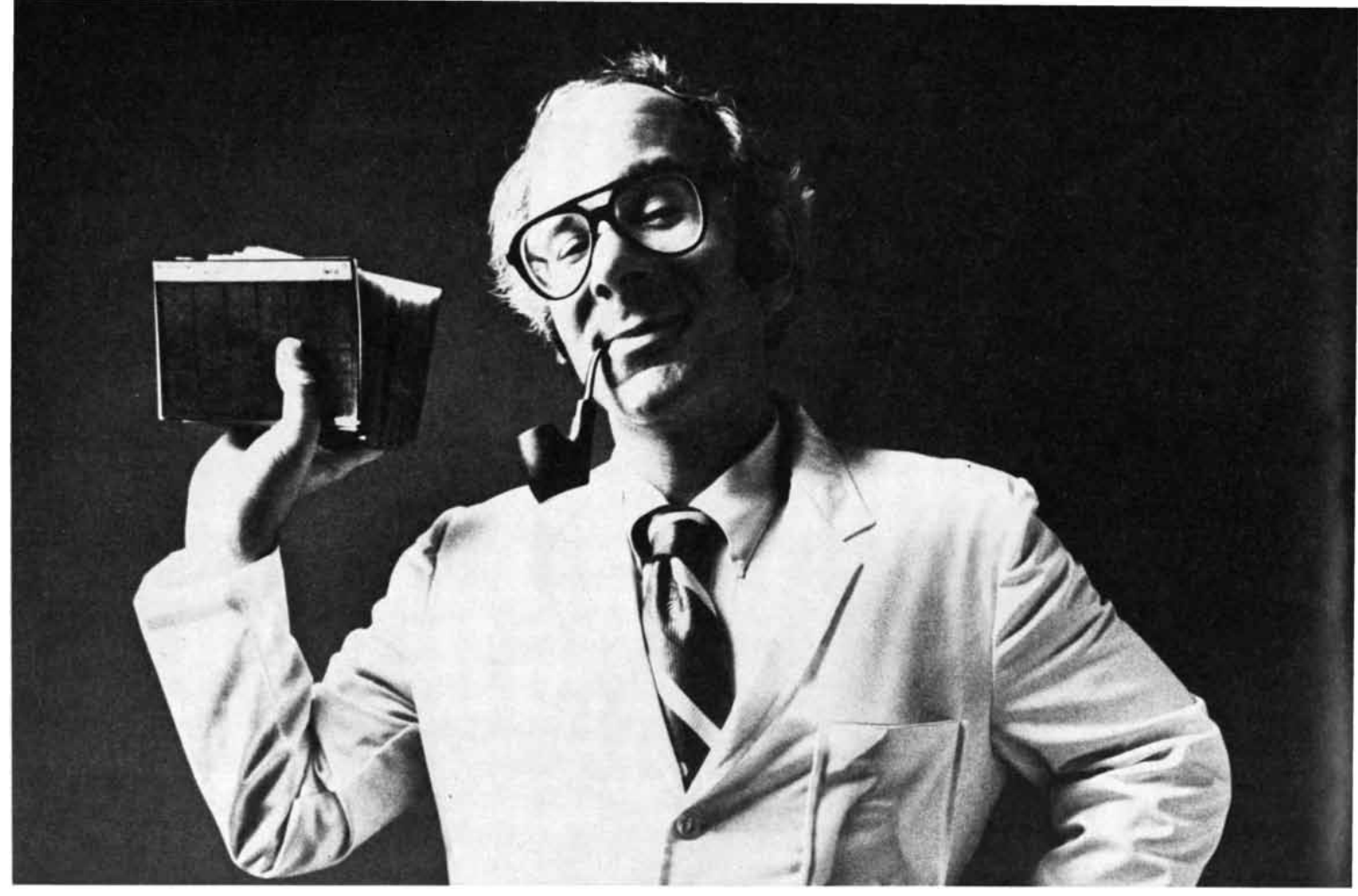




YOUR NEXT PHONE CALL



Bell Labs
Western Electric



"How We Solved The Paper Shortage."

More than 100 years ago, during the Franco-Prussian war, René Dagron photographed some 1,000 messages on a film two inches wide. He then strapped his film to a pigeon's leg and his message flew to Paris where it was projected by lantern on a screen and read.

That was the beginning of microfilm.

Today's microfilm systems allow you to store information in 1/56th the space you're used to.

1400 lbs. of paper will fit in your hand as 5 lbs. of microfiche.

A warehouse-full of computer printouts can be stored in a single room.

And as paper becomes increasingly more expensive and more difficult to get, the use of computer output microfilm becomes a necessity.

The MMS COM System

It's CalComp's way of putting together a COM printer, a controller with built-in magnetic tape unit and tape-cartridge drive, a keyboard printer, and the software you need for a microfiche management system. It stands alone. It does not require a host computer.

The CalComp system uses either 105 mm. microfiche or 16 mm. roll film. It can reduce images 24, 42, or 48 times. And the output emerges titled and indexed.

Job instructions can be typed in from the keyboard printer. This saves you costly re-programming.

Examine the CalComp system. We think you'll find the best performance for price in the industry.

Call your local CalComp office, or contact California Computer Products, Inc., SA-M1-76, 2411 West La Palma Avenue, Anaheim, California 92801. (714) 821-2011.

CALCOMP

Established 1845 SCIENTIFIC AMERICAN January 1976 Volume 234 Number 1

The Necessity of Fission Power

If the U.S. must have sources of energy other than fossil fuels, the only source that can make a major contribution between now and the end of the century is nuclear fission

by H. A. Bethe

The quadrupling of the price of oil in the fall of 1973 came as a rude but perhaps salutary shock to the Western world. It drew attention to the fact that oil is running out, and that mankind must turn to other fuels, to strict energy conservation or to both.

The price increase was not entirely unjustified. From 1950 to 1973 the price of oil, measured in constant dollars, had declined steadily. Moreover, it has been estimated that if world oil production were to continue to increase at the same rate that it has in the past two decades, the upward trend could persist only until about 1995; then the supply of oil would have to drop sharply [*see illustration on next page*]. Accordingly the oil-producing countries must see to their own economic development while their oil lasts so that they can rely on other sources of revenue thereafter. At the same time the rest of the world must take measures to become less dependent on oil—particularly imported oil—while there is still time.

What would it take for the U.S., which currently gets more than 15 percent of its energy in the form of imported oil, to become "energy independent?" In a report issued last June the Energy Research and Development Agency (ERDA) outlined its plans for the U.S. to achieve this goal. The ERDA projections are expressed in terms of quads, or quadrillions (10^{15}) of British thermal units (B.t.u.). According to ERDA, the drive to achieve energy independence calls for a two-pronged approach. First, the U.S. must be technologically geared not only to expand the production of its existing principal energy resources (oil, gas, coal and uranium) but also to develop several new energy sources. Second, a major energy-conservation effort must be initiated both to reduce total energy consumption and to shift consumption to sources other than oil. Only if both remedies are successfully applied can energy independence be achieved—and then it can be achieved only by 1995 [*see illustration on page 23*]. Without any new initiatives the need for imported oil will rise steadily from about 12 quads at present to more than 60 in the year 2000. At current prices the importation of that much oil would cost about $120 billion, compared with $25 billion in 1974, an increase of $95 billion.

Now, $95 billion may not sound like a gigantic sum when this fiscal year's Federal budget deficit is projected to be about $70 billion. The economics of international trade, however, is a different matter. Even a $10 billion trade deficit has a major effect on the stability of the currency. It is almost impossible to think of exports that could bring in an additional $95 billion. Besides, if current trends are allowed to continue, the U.S. would take about 30 percent of the world's oil production when that production is at its maximum. Clearly it is critical that the U.S. not follow this course.

What is critical for the U.S. is a matter of survival for Japan and the countries of western Europe. After all, the U.S. does have substantial amounts of oil and gas and plenty of coal. Japan and Italy have none of those fuels. England and Norway will have a limited domestic supply of oil in a few years, but other countries of western Europe have no natural oil resources of their own and have limited amounts of coal. If the U.S. competes for scarce oil in the world market, it can only drive the price still higher and starve the economies of western Europe and Japan. The bankruptcy of those countries in turn would make it impossible for the U.S. to export to them and thus to pay for its own imports.

For the next five years or so there is only one way for the U.S. to make measurable progress toward the goal of energy independence, and that is by conserving energy. There are two kinds of energy conservation. One approach is to have the country lower its standard of living in some respects, for example by exchanging larger cars for smaller ones. This measure has been widely accepted, probably at some cost in safety. To most Americans, however, it appears undesirable to continue very far in this direction.

The other approach to conservation is to improve the efficiency with which energy is consumed. A number of useful suggestions have been made, such as insulating houses better, increasing the efficiency of space-heating and water-

heating systems, improving the way steam is generated for industry and upgrading other industrial processes. Conversions of this type require substantial investment, and their cost-effectiveness on a normal accounting scheme is not clear. Much leadership, public education and tax or other incentives will be needed to realize the potential for increased efficiency. If all these things are provided, the total energy consumption of the U.S. in the year 2000 could be reduced from 166 quads to 120.

ERDA predicts that if at the same time the generation of electricity from coal and nuclear fuel is allowed to expand as it is needed, the U.S. can achieve an intermediate trend in oil imports: a satisfactory decline in the first 10 years, followed by a rise until oil imports are higher in 2000 than they are now. Energy independence will not have been achieved by that course either.

In all three ERDA projections it is assumed that the U.S. will move gradually from liquid fuels (oil and gas) to solid fuels (coal and uranium). For example, in President Ford's State of the Union Message in January, 1975, the actual contribution of various fuels to our energy budget in 1973 was presented along with the President's aims for 1985 and the expected situation in 1985 if no action is taken [*see illustration on page 24*]. The latter situation would require the importation of 36 quads of oil, in fair agreement with ERDA's prediction of 28 quads for 1985.

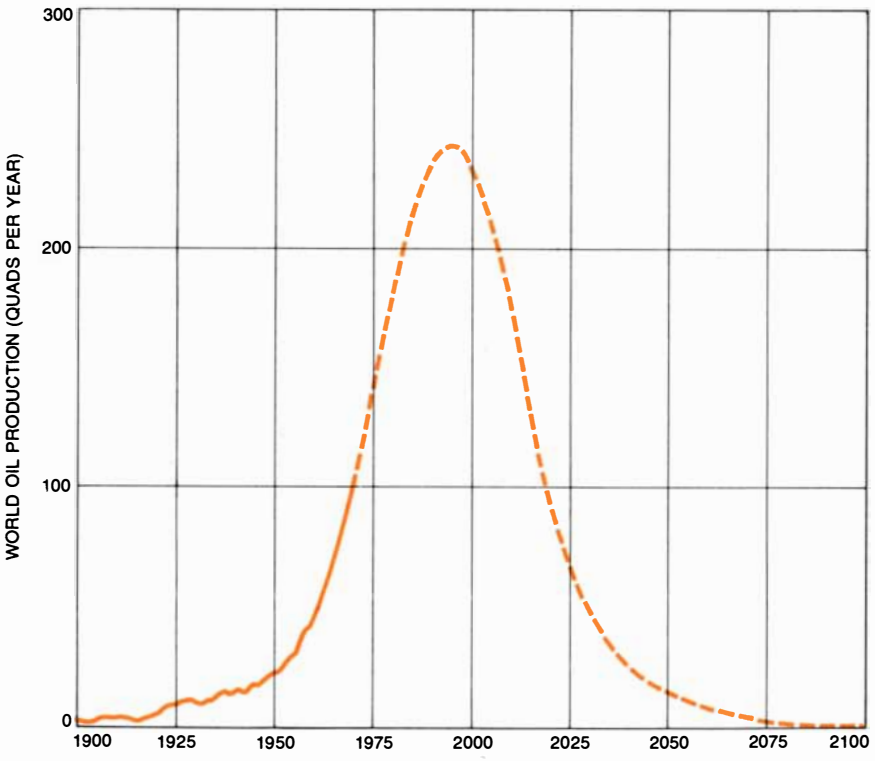
The Ford projection envisions a total U.S. consumption of 103 quads in 1985, 28 quads more than in 1973. Since much of the added energy would go into the generation of electricity, with a thermal efficiency of 33 to 40 percent, however, consumable energy would increase by only 17 quads, or 26 percent. Taking into account an expected 22 percent increase in the working population during that period, the consumable energy per worker would stay roughly constant.

The Ford message projects that domestic oil production will increase by seven quads by 1985 and that natural-gas production will decrease by only two quads, in spite of the fact that in the U.S. oil production has declined in the past two years and natural-gas discoveries have run at less than half of consumption for the past eight years. The ERDA report agrees that by stimulating the domestic production of oil and gas the U.S. could attain just about the total production figure used by the President, 53 quads, with gas somewhat higher than his estimate and oil lower.

Of course, the country would be depleting its resources more rapidly and would have to pay for it by having less domestic oil and gas in the years after 1985. The proposed stimulation of domestic oil and gas production, however, would provide the breathing space needed to bring other forms of energy into play. The only energy resources the U.S. has in abundance are coal and uranium. Accordingly President Ford calls for a massive increase in coal production, from 600 million tons in 1973 to 1,000 million tons in 1985. Meanwhile the Administration's energy program calls for the building of 200 nuclear-fission reactors with an energy output equivalent to about 10 quads.

Coal should certainly be substituted for oil and gas in utilities and in other industrial uses wherever possible. The conversion of coal into synthetic gas or oil is essential; demonstration plants for these processes and price guarantees should be given the highest priority. The same applies to oil from shale.

Coal cannot do everything, however, particularly if it is used intensively for making synthetic fuel. The U.S. needs another, preferably nonfossil, energy source. The only source that is now sufficiently developed to play any major role is nuclear fission. Thoughtful people have raised a number of objections to nuclear-fission reactors, which I shall discuss below, but first let me review some of the alternative energy sources that have been suggested.



FINITE SUPPLY OF OIL is responsible for the shape of this curve representing world oil production over a two-century span. The projection is based on the work of M. King Hubbert of the U.S. Geological Survey, who estimates that if world oil production were to continue to increase at the same rate that it has in the past two decades, output would peak in about 1995 and then drop sharply. Energy content of various fuels discussed in this article is expressed in quads, short for quadrillions (10^{15}) of British thermal units (B.t.u.).

Nuclear fusion is the energy source that has most strongly captured the imagination of scientists. It is still completely unknown, however, whether useful energy can ever be obtained from the fusion process. It is true that both stars and hydrogen bombs derive their energy from the fusion of light atomic nuclei, but can such energy be released in a controlled manner on the earth? The requirements for accomplishing the task are tremendous: a mixture of heavy-hydrogen gases must be brought to a temperature of about 100 million degrees Celsius and kept there long enough for energy-releasing reactions between the hydrogen nuclei to take place at a rate sufficient to yield a net output of energy.

The most obvious way to try to satisfy this condition is by magnetic confinement. At 100 million degrees hydrogen is completely ionized, and the positively charged nuclei and negatively charged electrons can be guided by magnetic

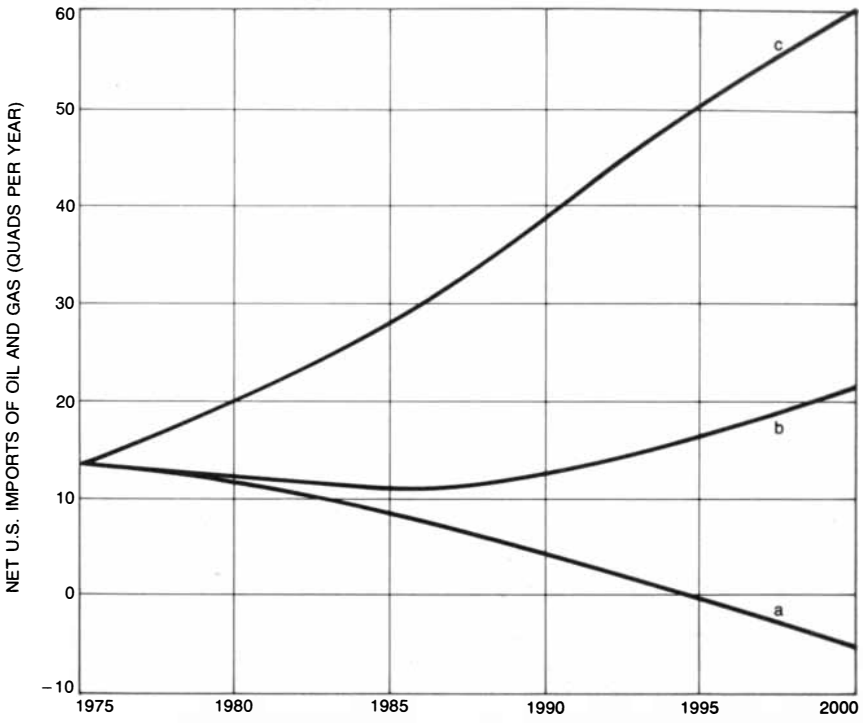
fields. Since the early 1950's physicists in many countries have designed many intricate magnetic-field configurations, but they have not succeeded in attaining the break-even condition. Great hopes have alternated with complete frustration. At present the prospects seem better than ever before; a few years ago Russian experimenters developed the device named Tokamak, which has worked at least roughly according to theoretical expectations. This device has been reproduced in the U.S. with comparable success. More than $200 million has now been committed by ERDA for a much larger device of the Tokamak type, to be built at Princeton University; if that machine also fulfills theoretical expectations, we may know by the early 1980's whether or not power from fusion is feasible by the Tokamak approach.

There have been too many disappointments, however, to allow any firm predictions. Work on machines of the Tokamak type is also going forward in many other laboratories in the U.S., in the U.S.S.R. and in several countries of western Europe. If the problem can be solved, it probably will be. Money is not the limiting factor: the annual support in the U.S. is well over $100 million, and it is increasing steadily. Progress is limited rather by the availability of highly trained workers, by the time required to build large machines and then by the time required to do significant experiments. Meanwhile several alternative schemes for magnetic confinement are being pursued. In addition there are the completely different approaches of laser fusion and electron-beam fusion. In my own opinion the latter schemes are even further in the future than Tokamak.

Assume now that one of these schemes succeeds in the early 1980's. Where are we then? The problem is that the engineering of any large, complex industrial plant takes a long time, even after the principle of design is well known. Since preliminary fusion-power engineering is already under way, however, it is a reasonable hope that a prototype of a commercial fusion reactor could operate in about 2000, and that fusion might contribute a few percent of the country's power supply by 2020.

Solar power is very different. There is no doubt about its technical feasibility, but its economic feasibility is another matter. One should distinguish clearly between two uses of solar power: the heating of houses and the production of all-purpose power on a large scale.

Partial solar heating of houses may become widespread, and solar air-conditioning is also possible. ERDA is sponsoring the development of model solar-heated houses. Private estimates for solar-heating systems, for a "standard" house of 1,200 square feet, run between $5,000 and $10,000 in mass production, compared with about $1,000 for conventional heating systems. With such an installation one might expect to supply about 50 percent of the house's heating requirements (more in the South, less in the North, particularly in regions of frequent cloud cover). In any case an auxiliary heating system supplied with gas, oil or electricity must be provided; otherwise the cost of the solar-heating system becomes exorbitant.



THREE PROJECTIONS of the extent to which the U.S. will continue to be dependent on imported oil and gas for the rest of this century were made in a 1975 report by the Energy Research and Development Administration (ERDA). Curve *a*, which shows the U.S. ending its dependence on imported fuel in about 1995, assumes that all the recommendations of ERDA's National Plan for achieving "energy independence" are put into effect. Curve *b* shows the effect of an intermediate approach emphasizing energy conservation. Curve *c* shows the expected trend in oil imports in the absence of any major new initiatives. All three curves assume that the energy efficiency of automobiles will improve by about 40 percent over the next five years or so as a result of the trend toward smaller cars and that the number of coal-burning and nuclear power plants will expand to meet rising demand for electricity. Any slowdown on nuclear plants would shift all three curves sharply upward.

ERDA estimates that 32 million new households will be established between 1972 and 2000, and that they will then comprise about a third of all dwelling units. If all the new units are equipped with solar heating, it would require a private investment of $150 to $300 billion. The heating requirement for all residential units in 1973 was close to 10 percent of the country's total energy consumption, and that fraction is likely to remain about the same. Some of the new dwelling units will not use solar energy, but let us assume (optimistically) that an equal number of older houses will be converted to solar heat. In that case a third of all houses would derive on the average about half of their heat from the sun, which would then supply somewhat less than 2 percent of the country's total energy needs. This contribution would be helpful but clearly would not be decisive.

The use of solar heat on a large scale for power generation is something else again. (Here I shall assume electric power, but the situation would not be essentially different if the energy were to be stored in fuels such as hydrogen.) Of the many proposals that have been made, the most practical in my opinion is to have a large field (perhaps a mile on a side) covered by mirrors, all of which reflect sunlight to a central boiler. The mirrors would be driven by a computer-controlled mechanism; the boiler would generate electricity in the conventional manner. At least three separate groups, supported by ERDA, are working on this kind of project. The best esti-

mates I have heard give about $2,500 per installed kilowatt (power averaged over the 24-hour day) exclusive of interest and the effects of inflation during construction. On the same basis nuclear-fission reactors cost about $500 per kilowatt, so that solar power is roughly five times as expensive as nuclear power.

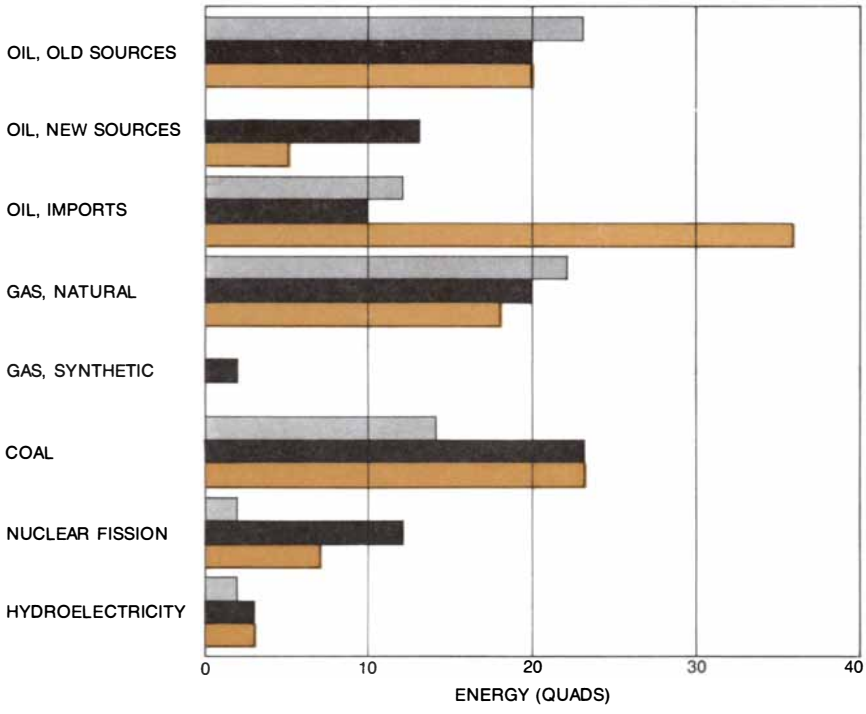
That cost estimate may sound high, but a little thought will show that it is not. First of all, the sun shines for only part of the day. On a sunny winter day in the southern U.S. one square mile of focused mirrors is just about enough to generate an average of 100 megawatts of electric power at a cost of about $250 million. To achieve that output the full heat of the sun must be utilized whenever it shines. At noon such a system would generate about 400 megawatts; near sunrise and sunset it would generate correspondingly less; at night it would generate none. To get an average of 100 megawatts one must have equipment to generate 400 megawatts, so that the generating equipment (boilers, turbines and so on) would cost roughly four times as much as they would in a comparable nuclear or fossil-fuel power plant. To this total cost must be added the cost of storing the energy that will be needed at night and on cloudy days. (The means of storage is so far a largely unsolved problem.)

Assume now that half of the cost is allotted to the mirrors and their electronic drive mechanisms; that would amount to $125 million for a plant of one square mile, or less than $5 per square foot. It is hardly conceivable that the mirrors and their drives could be built that cheaply, even in mass production, when a modest house costs $30 a square foot. I conclude therefore that all-purpose solar power is likely to remain extremely expensive.

Although it seems clear that solar power can never be practical for western Europe and Japan, the countries that need power most urgently, it might be just the right thing for certain developing countries, provided that the capital-cost problem can be solved. Many of those countries have large desert areas, rather modest total energy needs and abundant cheap manpower, which is probably required for the maintenance of any solar-power installation.

In addition to the alternative energy sources discussed above, a variety of other schemes have been suggested, such as harnessing the wind or the tides, burning garbage or agricultural wastes, converting fast-growing plants into fuels such as methane or tapping the earth's internal heat. Each of these approaches presents its own special difficulties, and at best each can make only a minor contribution toward the solution of the energy problem.

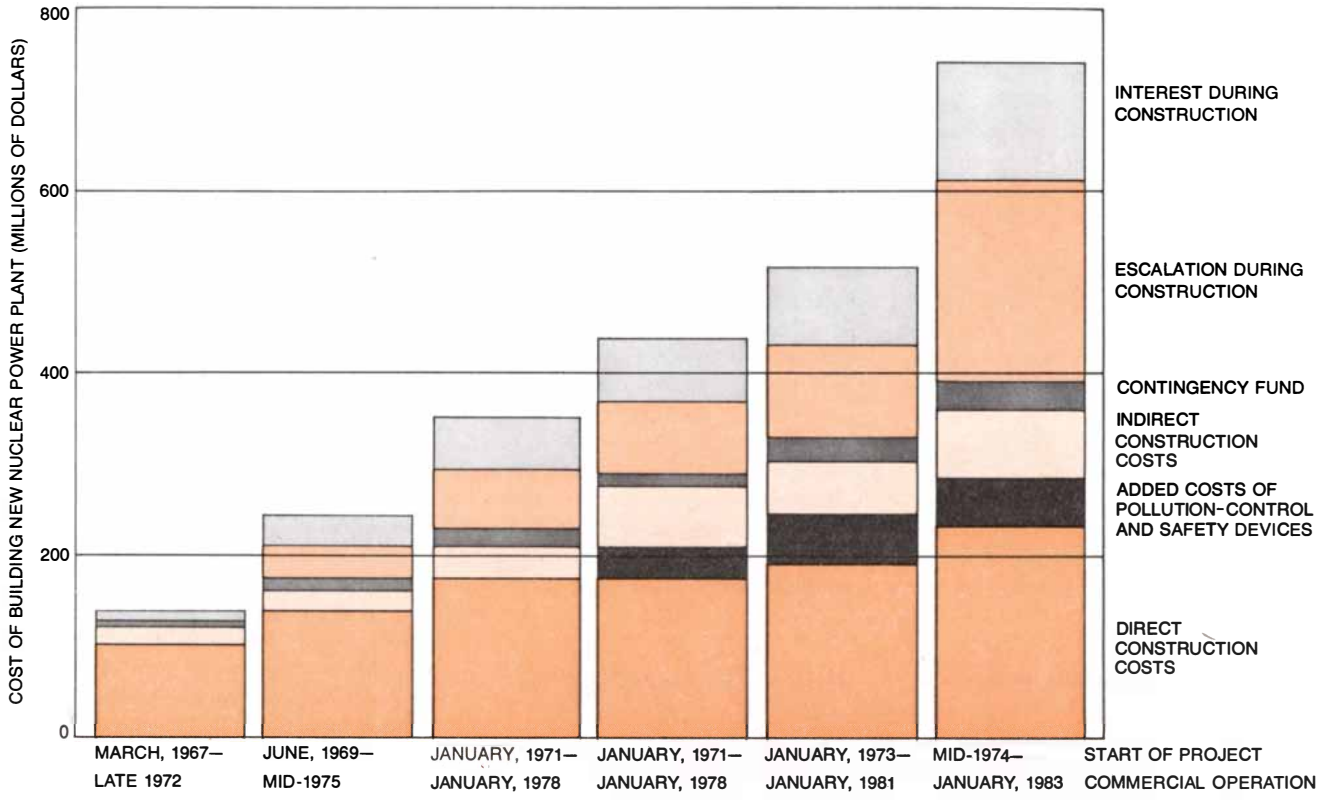
I do not mean to imply that work on alternative-energy projects is worthless. On the contrary, I believe that research and development on many of them should be pursued, and in fact ERDA is stepping up this type of work. I want to emphasize, however, that it takes a very long time from having an idea to proving its value in the laboratory, a much longer time for engineering development so that the process can be used in a large industrial plant and a still longer time before a major industry can be established. Certainly for the next 10 years and probably for the next 25 years the U.S. cannot expect any of the proposed alternative energy schemes to have much impact.



TWO ESTIMATES of the contribution of various fuels to the energy budget of the U.S. in 1985 are presented here along with the actual energy budget in 1973 (*light gray bars*). The chart is based on President Ford's State of the Union Message for 1975, in which he compared the expected impact of his administration's energy program (*dark gray bars*) with the expected situation if no action is taken (*colored bars*). The total U.S. energy consumption in 1973 was 75 quads; the total for 1985 in the absence of any major new programs is projected here to be about 112 quads, including some 36 quads of imported oil; Ford program, which includes a major energy-conservation effort aimed at saving about nine quads per year by 1985, envisions a total U.S. energy consumption of 103 quads for that year.

For all these reasons I believe that nuclear fission is the only major nonfossil power source the U.S. can rely on for the rest of this century and probably for some time afterward. Let us now examine the objections that have been raised against this source of power.

Some concern has been expressed over the fact that nuclear reactors in routine operation release radioactivity through outflowing liquids. According to the standards originally set by the Atomic Energy Commission and now administered by the Nuclear Regulatory Commission, these releases must be kept "as low as practicable," and under no circumstances must the additional radiation exposure of a person living permanently near the fence of the power plant be greater than five millirem per year. Most modern fission power plants release far less than this limit. For the purposes of comparison an average person in the U.S. receives 100 millirem per year in natural radiation (from cosmic rays, radioactivity in the earth and in buildings and radioactive substances inside his body) and an average of about 70 millirem per year from diagnostic medical X rays. It has been estimated that in the year 2000 a person living in the U.S. would on the average receive an additional tenth of a millirem from nuclear reactors if 1,000 of them are deployed. Chemical plants for reprocessing the nuclear fuel may add a couple of tenths of a millirem, but the Nuclear Regulatory



CONSTRUCTION COST of a 1,000-megawatt nuclear power plant of the light-water-reactor type has risen substantially for plants planned between 1967 and 1974 and expected to become operational between 1972 and 1983. As this ERDA bar chart indicates, however, a large fraction of the cost increase is due to inflation and interest during construction. Cost of building a coal-burning power plant has risen at a comparable rate during this period. (The bar for nuclear plants started in 1971 was revised upward in 1973.)

Commission is tightening the regulations further. In view of these very small numbers the controversy over the routine release of radioactivity, which was strong in the 1960's, has pretty much died down.

A more popular fear at present is that a reactor accident would release catastrophic amounts of radioactivity. Here it must be said first of all that a reactor is not a bomb. In particular, light-water reactors, which make up the bulk of U.S. reactors at present, use uranium fuel with a readily fissionable uranium-235 content of only 3 percent. Such material, no matter how large the amount, can never explode under any circumstances. (For breeder reactors, which can only come into operation about 20 years from now, the argument is slightly more complicated.)

It is, however, conceivable that a reactor could lose its cooling water, melt and release the radioactive fission products. Such an event is extremely unlikely, and one has never happened. There are at least three barriers to such a release. The radioactive fission products are enclosed in fuel pellets, and those pellets have to melt before any radioactivity is released. No such "meltdown" has occurred in nearly 2,000 reactor-years of operation involving commercial and military light-water reactors in the U.S. Moreover, even if there were to be a meltdown, the release of radioactivity would be retarded by the very strong reactor vessel, which typically has walls six to 12 inches thick. Finally, once this reactor vessel melts through, the radioactive material would still be inside the containment building, which is equipped with many devices to precipitate the volatile radioactive elements (mainly iodine, cesium and strontium) and prevent them from escaping to the outside. Only if very high pressure were to build up inside the reactor building could the building vent and release major amounts of radioactivity. The chance of that happening is extremely small, even in the event of a meltdown.

One may nonetheless ask: Exactly how likely is such a reactor accident? Obviously it is very difficult to estimate the probability of an event that has never happened. Fortunately most of the conceivable failures in a reactor do not lead to an accident. Reactors are designed so that in case of any single failure, even of a major part of the reactor, the reactor can still be safely shut down. Only when two or more essential elements in the reactor fail simultaneously can an accident occur. This makes a probabilistic study possible; an estimate is made of the probability of failure of one important reactor element, and it is then assumed that failures of two different elements are independent, so that the probability of simultaneous failure of the two is the product of the individual probabilities. This, however, is not always true. There can be "common mode" failures where one event triggers two or three failures of essential elements of the reactor; in that case the probability is the same as that of the triggering event, and one does not get any benefit from the multiplication of small probability numbers. The probability of such common-mode failures is of course the most difficult to estimate.

Working on the basis of these principles, a Reactor Safety Study commissioned three years ago by the AEC estimated the probability of various types of reactor accident. The results were published in draft form in August, 1974, in a document that has come to be known as the Rasmussen report, named for the chairman of the study group, Norman C. Rasmussen of the Massachusetts Insti-

tute of Technology. The final report was published last October.

The methods applied in the Rasmussen report have been used for several years in Britain to predict the probability of industrial accidents. Experience has shown that the predictions usually give a frequency of accidents somewhat higher than the actual frequency. Several groups, including the Environmental Protection Agency and a committee set up by the American Physical Society, have since studied various aspects of the problem and have come out with somewhat different results. Those differences have been taken into account in the final Rasmussen report; the most important of them will be discussed here.

The basic prediction of the Rasmussen report is that the probability of a major release of radioactivity is about once in 100,000 reactor-years. (Common-mode failures were found to contribute comparatively little to the total probability.) Such an accident would involve the release of about half of the volatile fission products contained in the reactor. A release of that scale would have to be preceded by a meltdown of the fuel in the reactor, an event for which the report gives a probability of once in 17,000 reactor-years. Finally, the report predicts that the water coolant from a reactor will be lost once in 2,000 reactor-years, but that in most cases a meltdown will be prevented by the emergency core-coolant system.

There is at least some check on those estimates from experience. For one thing, there has never been a loss of coolant in 300 reactor-years of commercial light-water-reactor operation. Furthermore, there has never been a fuel meltdown in nearly 2,000 reactor-years of commercial and naval light-water-reactor operation. If Rasmussen's estimate were wrong by a factor of 20 (in other words, if the probability of a meltdown were once in 850 reactor-years), at least one meltdown should have occurred by now.

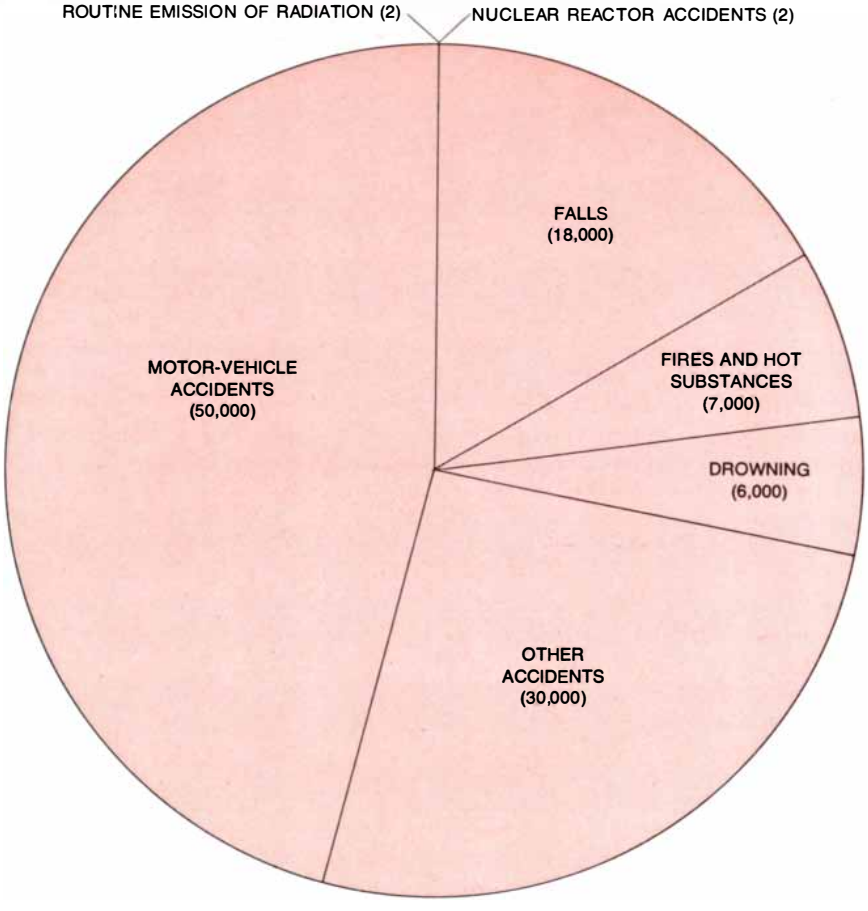
What would be the consequences in the extremely improbable event of a major release of radioactivity? The immediate effects depend primarily on the population density near the reactor and on the wind direction and other features of the weather.

For a fairly serious accident (one that might take place in a million reactor-years) Rasmussen estimates less than one early fatality but 300 cases of early radiation sickness. He also predicts that there could be 170 fatalities per year from latent cancers, a death rate that might continue for 30 years, giving a total of some 5,000 cancer fatalities. In addition there might be 25 genetic changes per year; counting the propagation of such changes through later generations, the total number of genetic changes could be about 3,000.

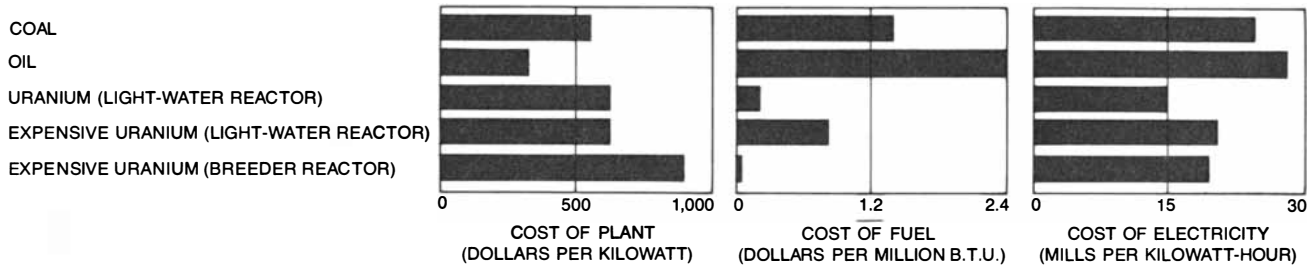
The number of latent cancers in the final version of the Rasmussen report is about 10 times as high as it was in the original draft report; that change was largely suggested by the study of the American Physical Society, as modified by a very careful study made by the Rasmussen group. A major release of radioactivity under average weather and population conditions (probably one in 100,000 reactor-years) would cause about 1,000 latent cancers, but it would not result in any cases of early radiation sickness.

It is obvious that 5,000 cancer deaths would be a tragic toll. To put it in perspective, however, one should remember that in the U.S. there are more than 300,000 deaths every year from cancers due to other causes. A reactor accident clearly would not be the end of the world, as many opponents of nuclear power try to picture it. It is less serious than most minor wars, and these are unfortunately quite frequent. Some possible industrial accidents can be more serious, such as explosions and fires in large arrays of gasoline storage tanks or chemical explosions. The danger from dam breaks is probably even greater.

The probability of a serious reactor accident was predicted in the Rasmussen report to be once in 10,000 years when there are 100 reactors, which is about the number expected for the U.S. in the year 1980. What if the number of reac-



ACCIDENT RISKS estimated for the entire U.S. population as the result of the operation of 100 nuclear power plants are compared here with the risks from several leading causes of accidents in terms of the average number of deaths per year attributable to each cause. (The averages for the latter categories are rounded to the nearest 1,000 fatalities.) The figure for the risk of death from nuclear accidents is based on the conservative assumption that there is likely to be one major release of radioactivity in the U.S. every 1,000 years, resulting in about 1,000 eventual deaths from cancer, and that once in 10,000 years there could be a more serious accident resulting in approximately 5,000 eventual deaths. The average risk from nuclear reactors is obviously extremely small compared with other risks that society accepts. It must be noted, however, that the nuclear-power risk can only be predicted, whereas the other risks are actuarial, that is, derived from statistics of actual events.



COST ESTIMATES summarized in this bar chart were made in late 1973 by the Philadelphia Electric Company for prospective electric-power plants consuming three different types of fuel: coal, oil and uranium. The first three entries in each set of bars were made on the assumption that plant construction would start in 1974. (The bars representing fuel costs for coal and oil have been updated to 1975.) The bottom two entries refer to an indefinite date in the future when uranium is expected to become much more expensive. According to this study, electricity from nuclear fuel will continue to be substantially cheaper than that from fossil fuel.

tors increases to 1,000, as many people predict for the year 2000 or 2010? The answer is that reactor safety is not static but is a developing art. The U.S. is now spending about $70 million per year on improving reactor safety, and some of the best scientists in the national laboratories are engaged in the task. I feel confident that in 10 years these efforts will improve both the safety of reactors and the confidence we can have in that safety. I should think that by the year 2000 the probability of a major release of radioactivity will be at most once in 10 million reactor-years, so that even if there are 1,000 reactors by that time, the overall chance of such an accident will still be no more than once in 10,000 years.

Taking into account all types of reactor accidents, the average risk for the entire U.S. population is only two fatalities per year from latent cancer and one genetic change per year. Compared with other accident risks that our society accepts, the risk from nuclear reactors is very small [*see illustration on opposite page*].

A special feature of possible reactor accidents is that most of the cancers would appear years after the accident. The acute fatalities and illnesses would be rather few compared with the 5,000 estimated fatalities from latent cancers in the foregoing example. The problem is that many more than the 5,000 victims will think they got cancer from the radiation, and it will be essentially impossible to ascertain whether radiation was really the cause. The average probability that the exposed population will get fatal cancer from the released radioactivity is only about 1 percent, compared with the 15 percent probability that the average American will contract fatal cancer from other causes. Will the affected people in the case of a reactor accident be rational enough to appreciate this calculation? Or would an accident, if it occurs, have a psychological effect much more devastating than the real one?

The problem of nuclear energy that is considered most serious by many critics is the disposal of nuclear wastes. Will such wastes poison the atmosphere and the ground forever, as has been charged? It is true that the level of radioactivity in a standard 1,000-megawatt reactor is very high: about 10 billion curies half an hour after the reactor is shut down. The radioactivity then decays quite quickly, however, and so does the resulting heat.

When the spent nuclear fuel is unloaded from a reactor, it goes through a number of stages. First the highly radioactive material, still in its original form, is dropped into a tank of water, where it is left for a period ranging from a few months to more than a year. The water absorbs the heat from the radioactive decay and at the same time shields the surroundings from the radiation.

After the cooling period the fuel will in the future be shipped in specially protected trucks or railcars to a chemical-reprocessing plant. (No such plant is currently in operation, but a large one is being built in South Carolina and could go into operation next year.) In the chemical plant the fuel rods will be cut open (still under water) and the fuel pellets will be dissolved. The uranium and the plutonium will be separated from each other and from the radioactive fission products. The uranium and plutonium can be reused as reactor fuel and hence will be refabricated into fuel elements. The remaining fission products are the wastes.

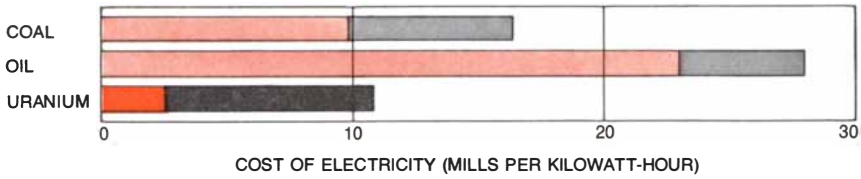
These substances are first stored in a water solution for an additional period to allow the radioactivity to decay further. Special tanks with double walls are now being used for that purpose in order to ensure against leakage of the solution.

After five years the wastes will be converted into solids, and after another five years they will be shipped to a national repository. Three different methods have been developed for solidifying wastes; one method now operates routinely at ERDA's reactor test station in Idaho to solidify the wastes from Government-owned reactors. The solid wastes can then be fused with borosilicate glass and fabricated into solid rods, perhaps 10 feet long and one foot in diameter. (Approximately 10 such rods will be produced by a standard 1,000-megawatt reactor in a year.) The rods are then placed in sturdy steel cylinders closed at both ends. It is difficult to see how any of the radioactive material could get out into the environment after such treatment, provided that the material is adequately cooled to prevent melting.

There are two possibilities for the national repository. One, for interim storage, would be in an aboveground desert area; the steel cylinders would be enclosed in a heavy concrete shield to protect the external world from the radiation. Cooling would be provided by air entering at the bottom of the concrete shield, rising through the space between the steel and the concrete and escaping at the top after having been heated by about 20 degrees C. Natural air circulation would be sufficient; no fans are required. The proposal for such a national repository has been studied and approved by a committee of the National Academy of Sciences.

The area required for such an interim-storage repository is not large. A standard reactor produces about two cubic meters of solid waste a year. The National Academy of Sciences committee estimated that all the wastes produced by U.S. reactors by 2010 could be stored on a tract of 100 acres. The cost is estimated at $1.5 billion, a small fraction of the probable cost of the reactors.

The second possibility for the national



ACTUAL POWER COSTS for the first quarter of 1975 were obtained by averaging the total costs, including fuel (*color*), for 22 utilities that operate nuclear reactors as well as other plants. These figures, compiled by the Atomic Industrial Forum, are in accord with those in illustration on preceding page. (The capital investment in these currently operating plants was of course much lower than that estimated for the future plants in that example.)

repository is permanent storage deep underground. The preferred storage medium here is bedded salt, which presents several advantages. First, the existence of a salt bed indicates that no water has penetrated the region for a long time; otherwise the salt would have been dissolved. Water trickling through the storage site should be avoided, lest it leach the deposited wastes and bring them back up to the ground, an extremely slow process at best but still better avoided altogether. Second, salt beds represent geologically very quiet regions; they have generally been undisturbed for many millions of years, which is good assurance that they will also remain undisturbed for as long as is required. Third, salt flows plastically under pressure, so that any cracks that may be formed by mechanical or thermal stress will automatically close again.

The first attempt by the AEC to find a storage site in a salt mine in Kansas was unfortunately undertaken in a hurry without enough research. (Drill holes in the neighborhood might have allowed water to penetrate to the salt bed and the waste.) Now ERDA is carefully examining other sites. A promising location has been found in southeastern New Mexico. There are roughly 50,000 square miles of salt beds in the U.S.; only three square miles are needed for disposal of all the projected wastes up to the year 2010.

The method of disposal is this: In a horizontal tunnel of a newly dug mine in the salt bed, holes would be drilled in the wall just big enough to accommodate one of the steel cylinders containing waste. It has been calculated that the cylinders could be inserted into the salt 10 years after the waste comes out of the reactor. The residual heat in the waste, five kilowatts from one cylinder, is then low enough for the salt not to crack. (The high heat conductivity of salt helps here.) If the calculation is confirmed by experiment in the actual mines, the wastes could go directly from the chemical-processing plant into permanent disposal and interim storage would be unnecessary. Otherwise the wastes would be placed for some years in the interim repository and then be shifted from there to permanent storage underground.

It seems to me virtually certain that a suitable permanent storage site will be found. It is regrettable that ERDA is so slow making a decision and announcing it, but after the difficulties with the Kansas site it is understandable that ERDA wants to make absolutely sure the next time.

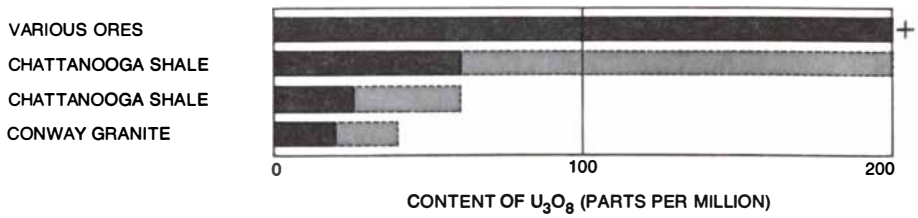
Most of the fission products have short half-lives, from a few seconds to a few years. The longest-lived of the common products are cesium 137 and strontium 90, with half-lives of about 30 years. The problem is that the wastes also contain actinides: elements heavier than uranium. In the present chemical process .5 percent of plutonium and most of the other actinides go with the wastes. Plutonium 239 has a half-life of nearly 25,000 years, and 10 half-lives are required to cut the radioactivity by a factor of 1,000. Thus the buried wastes must be kept out of the biosphere for 250,000 years.

Scientists at the Oak Ridge National Laboratory have studied the possible natural events that might disturb radioactive-waste deposits and have found none that are likely. Similarly, it is almost impossible that man-made interference, either deliberate or inadvertent, could bring any sizable amount of radioactivity back to the surface.

The remaining worry is the possibility that the wastes could diffuse back to the surface. The rate of diffusion of solids in solids is notoriously slow, and experiments at Oak Ridge have shown that the rate holds also for the diffusion of most fission products in salt. Ultimately this observation will have to be confirmed in the permanent storage site by implanting small quantities of fission products and observing their migration.

In the meantime one can draw further confidence from a beautiful "experiment" conducted by the earth itself. It has been discovered that in the part of Africa now called the Gabon Republic there existed some 1.8 billion years ago a natural nuclear reactor. A metal ore in that area is extremely rich in uranium, ranging from 10 to 60 percent. Whereas the present concentration of uranium 235 in natural uranium is .72 percent, the concentration 1.8 billion years ago was about the same as it is in present-day light-water reactors (3 percent). The ore also contained about 15 percent water. Therefore conditions were similar to those in a light-water reactor (except for the cooling mechanism). In the natural nuclear reactor plutonium 239 was formed, which subsequently decayed by emitting alpha radiation to form uranium 235. The interesting point is that the plutonium did not move as much as a millimeter during its 25,000-year lifetime. Moreover, the fission products, except the volatile ones, have stayed close to the uranium, even after nearly two billion years.

Assuming that plutonium is made in appreciable amounts, it must be kept from anyone who might put it to destructive use. Contrary to a widespread fear, however, there is little danger that plutonium could be stolen from a working nuclear reactor. The reactor fuel is extremely radioactive, and even if an unauthorized person were to succeed in unloading some fuel elements (a difficult and lengthy operation), he could not carry them away without dying in the attempt. The same is true of the used fuel cooling in storage tanks. The places



URANIUM RESOURCES OF THE U.S. are listed in this bar chart on the basis of ERDA estimates. The top row of bars refers to high-grade uranium ores. The prices throughout

from which plutonium might in principle be stolen are the chemical reprocessing plant (after the radioactive fission products have been removed), the fuel-fabrication plant or the transportation system between the plants and the reactor where the refabricated fuel elements are to be installed.

Transportation seems to be the most vulnerable link. Therefore it is probably desirable to establish the chemical plant and the fuel-fabrication plant close together, leaving only the problem of transportation from there to the reactor. Actually the problem of secure and safe transportation is essentially solved, at least in the U.S. The sophisticated safeguards now in force for nuclear weapons can be easily adapted for the transportation of nuclear materials. The protection of plants against theft is also being worked on and does not appear to present insuperable problems. For example, people leaving a plant (including employees) can be checked for possession of plutonium, even in small amounts, by means of automatic detectors, without requiring a body search. These direct measures for safeguarding plutonium are necessary and cannot be replaced by simple inventory-accounting procedures, which would be far too inaccurate. By ensuring that no plutonium (or fissionable uranium) has been diverted from U.S. plants one can be reasonably confident that no terrorists in this country can make an atomic bomb (which, by the way, is not as easy as some books and television programs have pictured it).

It has been asserted that the proposed measures for safeguarding plutonium and similar measures for protecting nuclear power plants from sabotage will interfere with everyone's civil liberties. I do not see why this should be so. The workers in the nuclear plants, the guards, the drivers of trucks transporting nuclear material and a few others will be subject to security clearance (just as people working on nuclear weapons are now). I estimate their number at less than 20,000, or less than 1 percent of our present armed forces. The remaining 200 million Americans need suffer no abridgement of their civil liberties.

Plutonium has been called the most toxic substance known. The term toxicity can be misleading in this context, because it implies that the danger lies in some chemical action of plutonium. Experiments with animals have shown that it is the level of radioactivity of the plutonium that counts, not the quantity inhaled, as is the case with a chemical poison. Nonetheless, the radioactive hazard is indeed great once plutonium is actually absorbed in the body: .6 microgram of plutonium 239 has been established by medical authorities as the maximum permissible dose over a lifetime, and an amount approximately 500 times greater is believed to lead to lethal cancer.

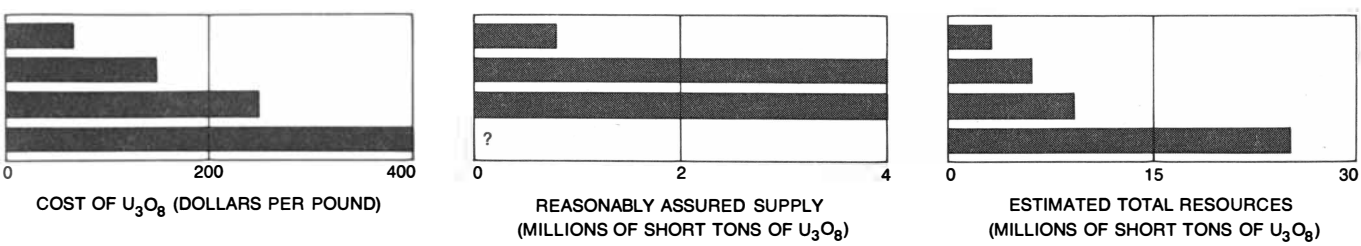
Plutonium can be effectively absorbed in the body if microscopic particles of it are inhaled. About 15 percent of the particles are likely to be retained in the lung, where they may cause cancer. Fortunately there is little danger if plutonium is ingested in food or drink; in that case it passes unchanged through the digestive tract, and only about one part in 30,000 enters the bloodstream. Therefore effective plutonium poisoning of the water supply or agricultural land is virtually impossible.

Some opponents of nuclear power have maintained that because of the very low maximum permissible dose even small amounts of plutonium in the hands of terrorists could cause great damage. This point has been put in perspective by Bernard Cohen, who has investigated in theory the effect of a deliberate air dispersal of plutonium oxide over a city. He finds that on the average there would be one cancer death for every 15 grams of plutonium dispersed, because only a small fraction of the oxide would find its way into people's lungs. Other, soluble compounds of plutonium would be even less effective than an insoluble oxide. A terrorist who manages to steal six kilograms or more of plutonium could probably do more damage by fashioning a crude bomb from it than by dispersing it in the air of a city.

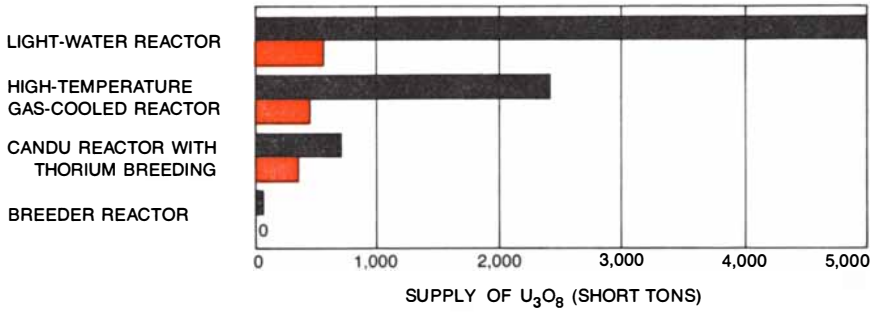
Will the spread of nuclear reactors encourage the proliferation of nuclear weapons? That in my opinion is the only really serious objection to nuclear power. The availability of fissionable material is obviously a prerequisite for making nuclear weapons. Even after the material is available, however, the manufacture of a nuclear bomb is still a massive undertaking: in each of the six countries that have so far conducted nuclear explosions, thousands of scientists and technicians have worked on the development of the weapon. Nonetheless, a number of additional countries would be capable of this effort if they wanted to make it, and if they had the material.

Many countries in need of nuclear power will soon be in the market for the purchase of nuclear power plants from any country willing to sell them. Nuclear power plants sold in international trade are usually put under the inspection system of the International Atomic Energy Agency (IAEA) in order to ensure that no fissionable material is diverted for military purposes. The IAEA needs strengthening and more money for its force of inspectors. An important additional safeguard would be to prevent the proliferation of nuclear chemical-processing plants, since it is from those plants rather than from the reactors that fissionable material could be diverted. A good proposal is that the chemical processing be centralized in plants for an entire region rather than dispersed among plants for each nation. Another approach would be to have the country supplying the reactor lease the fuel to the customer country with the requirement that the used fuel be returned.

The original fuel for a light-water reactor is mostly uranium 238 enriched with about 3 percent of readily fissionable uranium 235. If an explosive were to be made from this fuel, the two isotopes would have to be separated, a procedure that requires a high level of technology. The used fuel contains in



are in 1975 dollars and include rehabilitation of the land in the case of low-grade materials (*bottom three rows of bars*). Other estimates of total resources, measured in millions of short tons of uranium oxide (U_3O_8), range up to three times those given here.



URANIUM REQUIRED for the initial fueling (*colored bars*) and 40-year operation (*gray bars*) of a 1,000-megawatt nuclear power plant is indicated (again in millions of short tons of U_3O_8) for the four principal types of reactor system. For the breeder reactor the plutonium required for initial fueling is expected to be available as a by-product of previously operational light-water reactors. Candu is the Canadian name for their heavy-water natural-uranium reactor, which can be modified to convert thorium into the isotope uranium 233.

addition some plutonium, which can be separated from the uranium by chemical procedure, a less difficult task. The resulting plutonium has a high concentration of plutonium 240 (with respect to plutonium 239), which could be used to make rather crude bombs by a country just beginning in nuclear-weapons technology. Breeder reactors contain more plutonium per unit of power, with a smaller percentage of plutonium 240. I personally would therefore recommend that breeder reactors not be sold in international trade.

Proliferation would not be prevented if the U.S. were to stop building nuclear reactors for domestic use or if it were to stop selling them abroad. Western Europe and Japan not only need nuclear power even more than the U.S. does but also have the technology to acquire it. Moreover, they need foreign currency to pay for their oil imports and so they will want to sell their reactors abroad. The participation of the U.S. in the reactor trade may enable us to set standards on safeguards, such as frequent IAEA inspection, that would be more difficult if we left the trade entirely to others.

It has been alleged that nuclear power is unreliable. The best measure of reliability is the percentage of the time a plant is available for power production when the power is demanded. This "availability factor" is regularly reported for nuclear plants and runs on the average about 70 percent. There are fewer good data on the availability of large coal-fired plants, but where the numbers exist they are about the same as those for nuclear plants.

The "capacity factor" is the ratio of the amount of power actually produced to the amount that could have been produced if the plant had run constantly at full power. That percentage is usually lower than the availability factor for two reasons: (1) some nuclear power plants are required for reasons of safety to operate below their full capacity, and (2) demand fluctuates during each 24-hour period. The second factor is mitigated by the operation of nuclear reactors as base-load plants, that is, plants that are called on to operate as much of the time as possible, because the investment cost is high and the fuel cost is low. A reasonable average capacity factor for nuclear power plants is 60 percent. One utility has estimated that at a capacity factor of 40 percent nuclear and coal-fired plants generating the same amount of electricity would cost about the same; operation at 60 percent therefore gives the nuclear plant a substantial edge.

But are not nuclear power plants expensive to build? An examination of the construction cost of such plants planned between 1967 and 1974, and expected to become operational between 1972 and 1983, shows that the cost of a 1,000-megawatt power plant of the light-water-reactor type has risen from $135 million to $730 million in this period [*see illustration on page 25*]. Closer inspection reveals, however, that a large fraction of the cost increase is due to inflation and to a rise in interest rates during construction; without those factors the 1974 cost is $385 per kilowatt of generating capacity. This figure represents a cost increase of about 300 percent, which is more than the general inflation from 1967 to 1974. The main cause must be looked for in the steep rise of certain construction costs, particularly labor costs, which rose about 15 percent per year, or 270 percent in seven years.

The cost of building coal-fired plants has risen at a comparable rate. A major factor here has been the requirement of "scrubbers" to remove most of the sulfur oxides that normally result from the burning of coal. Coal plants equipped with scrubbers may still be about 15 percent cheaper to build than nuclear plants. Any massive increase in coal production would, however, call for substantial investment not only in the opening and equipping of new mines but also in the provision of additional railroad cars and possibly tracks, particularly in the case of Western mines. If this "hidden" investment is included, the capital cost of coal-burning power plants is not very different from that for nuclear plants. Even disregarding this factor the overall cost of generating electricity from nuclear fuel is already much less than it is for generating electricity from fossil fuel, and recent studies indicate that nuclear power will continue to be cheaper by a wide margin [*see illustration on page 27*].

There is some truth in the charge that "nuclear power does not pay its own way," since the Government has spent several billion dollars on research on nuclear power and several more billions will undoubtedly have to be spent in the future. On the other hand, the Government is also spending about $1 billion a year as compensation to coal miners who have contracted black-lung disease.

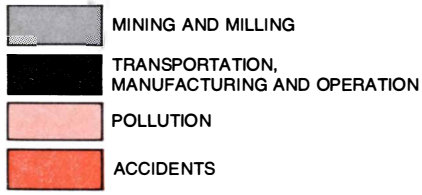
It has also been said that uranium will run out soon. It is true that the proved reserves of high-grade uranium ore are not very large, and the existing light-water reactors do require a lot of uranium. If all reactors were of this type, and if the U.S. were to set aside all the uranium needed for 40 years of reactor operation, then the total uranium-ore resources of the U.S. would only be enough to start up 600 reactors, a number that might be reached by the year 2000. Beyond that date it will be important to install reactors that consume uranium more efficiently. The most satisfactory alternative to emerge so far is the breeder reactor, which may be ready for industrial operation by 1990. The breeder in effect extracts the energy not only from the rare isotope uranium 235 but also from other isotopes of uranium, thereby increasing the supply of uranium about sixtyfold. Even more important, with the breeder the mining of low-grade uranium ore can be justified both economically and environmentally. With these added resources there is enough uranium in the U.S. to supply 1,000 reactors for 40,000 years [*see bottom illustration on preceding two pages*].

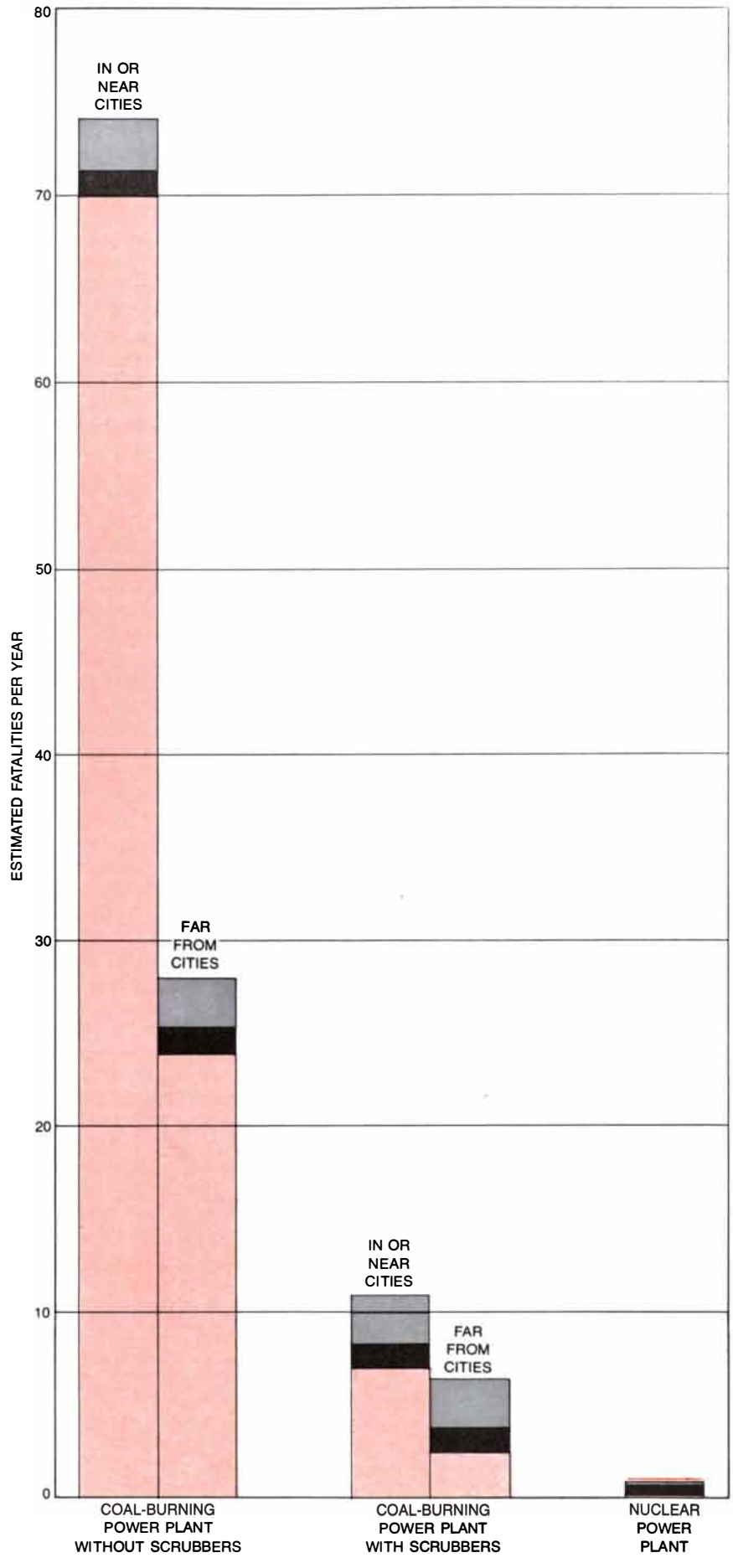
As interim alternatives two other types of reactor are attractive: the high-temperature, gas-cooled, graphite-moderated reactor and the Canadian natural-uranium reacter ("Candu"), which is moderated and cooled by heavy water.

The Candu reactor can be modified to convert thorium by neutron capture into the fissionable isotope uranium 233.

In weighing the overall health hazard presented by nuclear reactors it is appropriate to compare nuclear plants with coal-burning power plants. Recent findings indicate that even if scrubbers or some other technology could reduce the estimated health effects from coal burning by a factor of 10 (which hardly seems attainable at present), the hazard from coal would still exceed that from nuclear fuel by an order of magnitude [*see illustration at right*]. This comparison is not meant as an argument against coal. The U.S. clearly needs to burn more coal in its power plants, and even with coal the hazard is not great. The comparison does point up, however, the relative safety of nuclear reactors.

In sum, nuclear power does involve certain risks, notably the risk of a reactor accident and the risk of facilitating the proliferation of nuclear weapons. Over the latter problem the U.S. has only limited control. The remaining risks of nuclear power are statistically small compared with other risks that our society accepts. It is important not to consider nuclear power in isolation. Objections can be raised to any attainable source of power. This country needs power to keep its economy going. Too little power means unemployment and recession, if not worse.





OVERALL HEALTH HAZARDS presented by a nuclear power plant and a coal-burning power plant, both capable of generating 1,000 megawatts of electricity, are compared here in terms of the estimated number of deaths resulting from one year of operation. (Injuries and other health effects have been translated into equivalent deaths by a suitable formula.) The data were gathered primarily by C. L. Comar of the Electric Power Research Institute and L. A. Sagan of Stanford University. The coal-mining figure refers to underground mining; surface mining is much less dangerous. The figure for pollution from coal includes sulfur dioxide pollution only. As the second set of bars shows, even if sulfur dioxide "scrubbers" were to succeed in reducing the estimated hazard from coal-burning by a factor of 10, adverse health effects from coal power would still be greater than those from nuclear power.

ALBA PATERA is probably the largest single volcanic structure on the planet Mars. A partially filled caldera lies at the center of a ring of fractures that cuts across a fine pattern radiating from the caldera. The pattern is caused largely by flow features visible only in high-resolution photographs made from the spacecraft *Mariner 9*. The fractures, which are part of a set that extends outward from a bulge in the crust of Mars in the Tharsis region (*see illustration on pages 36 and 37*), developed after the main volcanic structure of Alba Patera had already formed, and thus they were deflected around it. Fracture ring is about 600 kilometers in diameter. Picture is a rendering that combines information from many photographs taken by *Mariner 9*, some of which were of uneven quality.

THE VOLCANOES OF MARS

They apparently formed over the entire span of the planet's history. Their remarkable size may be related to the fact that the crust of Mars is not divided into moving plates

by Michael H. Carr

Among the most startling of the many startling pictures sent back by the spacecraft *Mariner* 9 when it went into orbit around Mars in 1971 were those showing huge volcanoes in the planet's northern hemisphere. Earlier spacecraft, which had simply flown past Mars, had had only a view of those areas that lacked distinctive volcanic features. The existence of the enormous volcanoes located elsewhere was therefore quite unsuspected until they were revealed by the high-resolution photographs made from *Mariner* 9.

When the Martian volcanoes were first discovered, many observers were impressed not only by their size but also by their apparent newness and the apparent lack of older volcanoes. It seemed that Mars had just turned on volcanically. Closer examination of the available photographs has since revealed that there are quite a number of old volcanoes, although they are much degraded and difficult to recognize. Additional old features will quite likely be discovered. It now appears that Mars has had a long and variegated history of volcanic activity. If that is the case, then the planet's atmosphere did not form late in its evolution but is either the remnant of a primordial atmosphere or the product of continual volcanic release of gas from within the planet. Theories about the erosional history of Mars must be consistent with such an origin of the atmosphere. Similarly, models of the thermal evolution of the interior of Mars must take into account the fact that volcanism began early and was sustained for billions of years.

Mars can be roughly divided into two distinctively different hemispheres. The more southerly hemisphere is the more densely cratered and its landscape looks somewhat like the highlands of the earth's moon. The more northerly hemisphere consists mostly of sparsely cratered plains. The reason for the difference between the hemispheres is not known. The boundary between them appears to be erosional, as though the densely cratered terrain was being slowly destroyed. The erosional nature of the boundary is probably only a superficial expression of a more fundamental difference between the two hemispheres, however, because they are distinct from each other not only in their surface morphology but also in their volcanic history. Nearly all the young volcanic features are found in the sparsely cratered northern hemisphere. Volcanoes have been recognized in the southern hemisphere as well, but they are older and do not include any of the fresh-appearing large volcanoes. The distinctions must reflect major differences in the thermal history of the hemispheres, for example differences that might result from deep-seated variations in composition or in the thickness of the crust.

The most spectacular volcanic features of Mars are in the general region of Tharsis, where three of the four large "shield" volcanoes form a line running from southwest to northeast [*see illustration on pages 36 and 37*]. The fourth and largest shield volcano, Olympus Mons (first named Nix Olympica), lies some 1,000 kilometers to the northwest. The term shield is applied to volcanoes that are built up by successive eruptions of low-viscosity lava. The best-known shield volcanoes on the earth are in Hawaii. The low-viscosity lava, being able to flow a considerable distance, builds up a volcano with the shape of a broad dome. Such structures, which look like a shield in profile, generally have sides with a slope of less than six degrees.

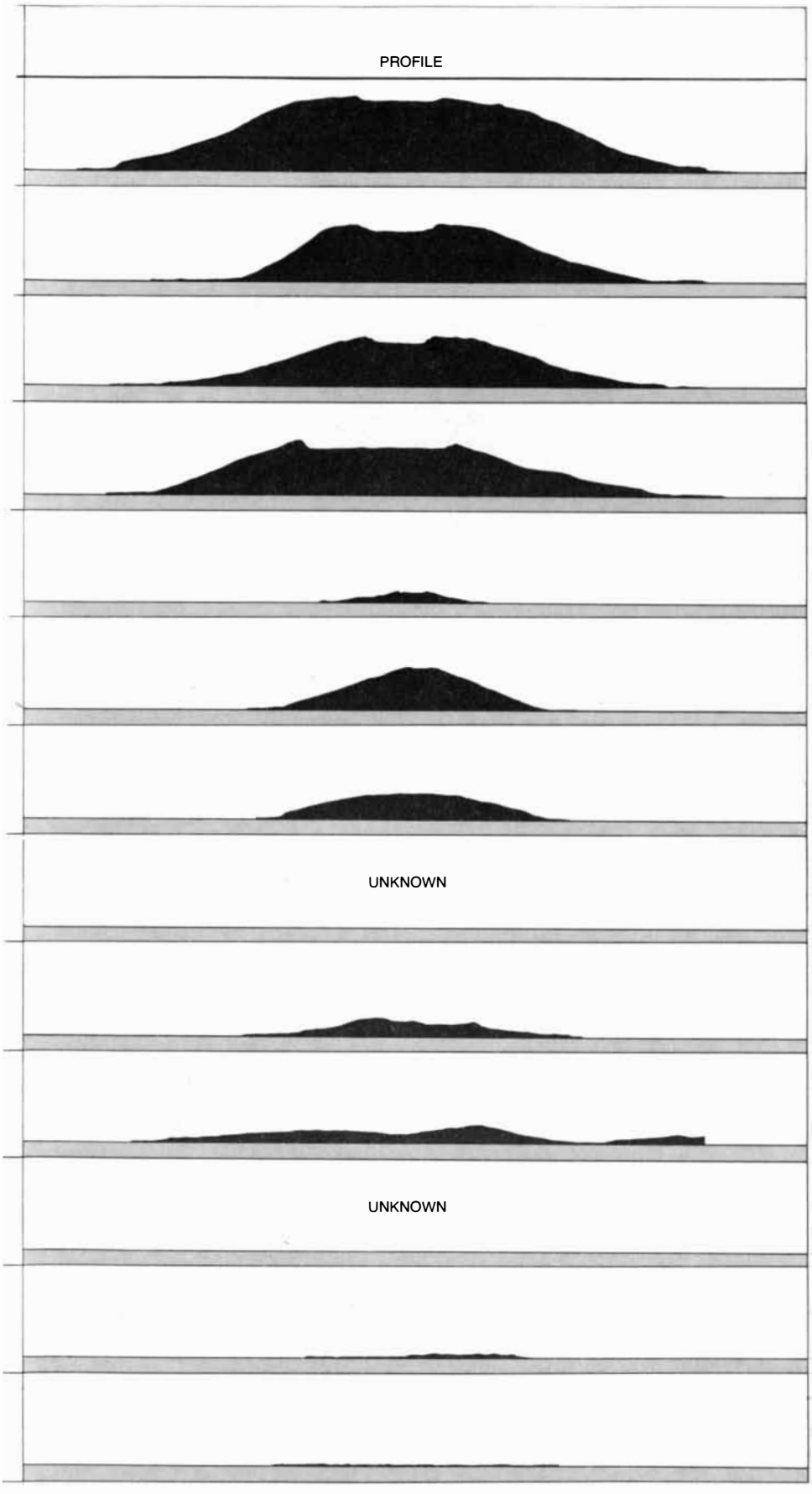
Olympus Mons is enormous by terrestrial standards. Its shield is between 500 and 600 kilometers across, some five times larger than the largest shield on the earth. At the summit, which is 23 kilometers (75,500 feet) above the surrounding plain, there is a complex caldera, or large volcanic crater. The flanks of the shield have a radial texture created by narrow channels, long fingerlike flows of lava, leveelike structures and flow lines. All these features are also found on the flanks of terrestrial shields, suggesting that the characteristics of the lava that made them are similar to those of the terrestrial lava. Cutting across the radial texture are breaks in the slope that divide the flanks of the volcano into rounded, intersecting terraces. Although the origin of the terraces is uncertain, they could have formed when lava flowing out of fissures in the flanks preferentially built up the volcano below the fissures. The complexity of the caldera implies that portions of it collapsed when lava flowed back into the mantle under the volcano after each eruptive episode. The entire shield drops to the surrounding plain in a cliff that ranges up to four kilometers (13,000 feet) in height.

The other large Tharsis shields resemble Olympus Mons in their general shape but differ from it in detail. Each is some 400 kilometers across and stands about 19 kilometers above the surrounding plain. Except around bays in their northeast and southwest margins, the three shields are not bounded by a cliff. Indeed, the plains generally overlap the edges of the shields, indicating that the plains formed after the volcanoes.

One feature of Olympus Mons that is not seen in the other shields is an aureole of distinctively textured terrain. All around the main shield and extending as far as 500 kilometers from the base of the cliff are discrete areas of curvilinear ridges and troughs, generally with their concave side facing the volcano. Each

SHIELD VOLCANO	REGION	HEIGHT (KILOMETERS)	DIAMETER (KILOMETERS)	APPROXIMATE AGE (MILLIONS OF YEARS)	DESCRIPTION
OLYMPUS MONS	THARSIS	26 ± 3	500 TO 600	200	Shield has gently sloping terraced flanks radially textured by channels, leveelike structures and flow lines. It is bounded by a cliff up to four kilometers high.
ASCRAEUS MONS	THARSIS	19 ± 3	400	400	Resembles Olympus Mons, but flanks also have numerous small pits. Northeast and southwest edges of shield are cut by intersecting cracks extending into surrounding plain.
PAVONIS MONS	THARSIS	19 ± 3	400	400	Same as Ascraeus Mons. Shield is also surrounded by numerous concentric grabens.
ARSIA MONS	THARSIS	19 ± 3	400	800	Large central caldera is surrounded by well-developed series of concentric grabens. Outer edge of shield has coarse radial texture of elongated lobes and short depressions resembling sinuous rilles on the moon.
CERAUNIUS THOLUS	THARSIS	2.5 ± 1.25	150	500 TO 1,000	Dome-shaped feature has numerous fine channels running from summit crater to base. One channel two kilometers wide connects central caldera to pool on neighboring plain.
ELYSIUM MONS	ELYSIUM	15 ± 2	200 TO 300	1,000 TO 2,000	Resembles Tharsis shields. Edge of shield is not distinctly defined but grades gently into surrounding plains.
HECATES THOLUS	ELYSIUM	7 ± 1	200	1,000 TO 2,000	Fissures, lines of vents and striae radiate from small (10-kilometer) caldera, and flanks of shield are covered with small pits. Shield meets plains in cliff at west but elsewhere plains terminate it abruptly.
ALBOR THOLUS	ELYSIUM	UNKNOWN	150	1,000 TO 2,000	Resembles Tharsis shields but seems older.
APOLLINARIS PATERA	–	4 ± 2	200	2,000 TO 3,500	Shield is considerably modified by erosion and seems older than Tharsis or Elysium shields.
ALBA PATERA	–	6 ± 2	1,600	1,000 TO 2,000	Ring of fractures 600 Kilometers across surrounds central caldera. Faint radial pattern of flow features can be traced as far as 800 kilometers from center.
TYRRHENUM PATERA	HESPERIA	UNKNOWN	500	3,500 TO 4,000	Ring of fractures 45 kilometers in diameter surrounds depression. Numerous ridges, depressions and channels extend some 200 kilometers from ring to give starlike appearance.
HADRIACA PATERA	HELLAS	1.7 ± .5	600	3,500 TO 4,000	Featureless circular area 70 kilometers across is at center of volcano, surrounded by subparallel radial ridges and valleys up to 300 kilometers long with impact craters superposed on them.
AMPHITRITES PATERA	HELLAS	<1	700	3,500 TO 4,000	Central featureless area is surrounded by faint radial texture extending 350 kilometers from center.

MAJOR SHIELD VOLCANOES ON MARS are described and their profiles are shown. All are in the sparsely cratered northern hemisphere of the planet except for the last three, which are ancient features in the southern hemisphere. Each volcano has such a large diameter compared with its height that the vertical scale of the profile is exaggerated four times. Because of Alba Patera's



great diameter only half of it is shown. Grabens are elongated depressions between two faults. The ages given are based on the number of impact craters superposed on each feature and are uncertain because the rate of impact of meteorites on Mars is not well known.

block of this lineated terrain appears to be tilted toward the volcano and is separated from adjacent blocks by smooth plain.

Although the origin of the aureole is not clear, several possibilities have been suggested. One is that the lineated terrain is the product of lava flows whose flow features have been etched by the wind; the flows were then faulted and tilted and partially covered by later lava flows. Another suggestion is that Olympus Mons was once far larger than it is today and that the cliff surrounding the volcano has encroached on the main shield, leaving the aureole behind as an erosional remnant. A third suggestion is that the aureole consists of successive layers of compacted volcanic ash. Whatever the origin of the aureole is, it is clearly related to the main volcanic center of Olympus Mons, and it indicates that the volcanic activity in the area is more widespread than is implied by the main shield alone.

The origin of the cliff around Olympus Mons is also puzzling. On the earth cliffs generally form by erosion, and it is clear that in many places on Mars cliffs formed the same way. It is difficult, however, to accept the proposition that the cliff of Olympus Mons is solely the result of erosion. All the features of the volcano suggest that the shield consists of a kind of lava that, when it solidifies, is generally resistant to erosion. Moreover, the erosion would have to be extremely selective: efficient at the face of the cliff and inefficient on the upper surface of the shield, where fine features are preserved. The difficulties with the erosion hypothesis have led some investigators to suggest that the cliff is a fault scarp that has been slightly eroded, and that the whole of Olympus Mons has moved upward with respect to its immediate surroundings. Some support is given this hypothesis by features near the edge of the cliff that resemble terrestrial horsts: blocks of land pushed upward between two faults. On the other hand, the erosion hypothesis is supported by the incipient cliff formation on the northeast and southwest edges of the other Tharsis shields.

There are numerous volcanoes in the Tharsis region in addition to the four large shields. Several dome-shaped features are present, ranging up to 150 kilometers in diameter and surmounted by a central caldera. One dome, Ceraunius Tholus, is particularly distinctive in that there are many fine channels running from the crater at the summit to the base. One large channel two kilometers

wide connects the central caldera to a walled depression in the surrounding terrain, suggesting that the lava flowed out of the caldera down the channel and formed a pool with levees around it. A similar phenomenon was observed during the eruption of Mauna Ulu in Hawaii in 1973. The channels running down the flanks of Ceraunius Tholus again indicate that the lava must be highly fluid, since channels can be created only by a liquid with a low viscosity. Other small volcanic features in the Tharsis region have the form of inverted saucers, with the central caldera being about half the diameter of the entire structure.

The second most prominent volcanic area on Mars is the Elysium region. The shield volcanoes in the region resemble those in the Tharsis region in having a central caldera and a fine radial texture. They are generally smaller, however, and seem to be older. The largest is Elysium Mons, which appears to be some 15 kilometers high and between 200 and 300 kilometers across. It has a simple central caldera from which numerous channels radiate. The edge of the shield is not distinctly defined but grades out into the surrounding plain.

There are several large volcanic features outside the two main regions of Tharsis and Elysium. Alba Patera, north of the Tharsis region, may be the largest volcanic feature on the entire planet. It consists of a ring of fractures some 600 kilometers across, and at its center is a partly filled caldera. Within the fracture ring the terrain is smooth and slopes gently inward. The terrain surrounding the fracture ring slopes gently away from the ring, which stands between four and six kilometers above the surrounding plain.

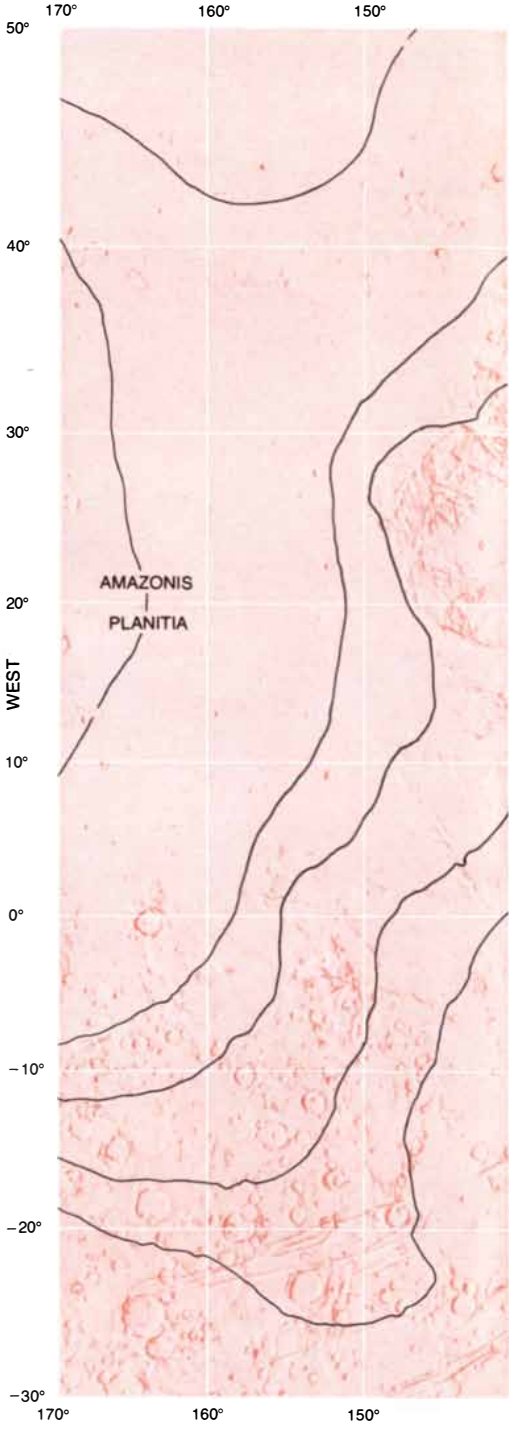
All the high-resolution pictures of Alba Patera made from *Mariner* 9 show numerous flow features outside the ring and on the flanks of the structure. Although the flow features are not visible in the *Mariner* 9 low-resolution pictures, a faint radial pattern that is probably caused by these features can be traced as far as 800 kilometers from the center. The entire object is thus 1,600 kilometers in diameter, which makes it substantially larger than Olympus Mons. In fact, Alba Patera is so large that it is questionable whether it can properly be called a volcano, a name that conjures up an image of a distinct conical structure. From the ground on Mars, Alba Patera would not even be discernible because its horizontal dimensions are so large and the slope of its flanks is less than a quarter of a degree. Nevertheless, it is undoubtedly a volcanic structure from which huge volumes of lava have erupted.

All the features I have described so far are in the northern hemisphere of Mars. The volcanic features in the densely cratered terrain of the southern hemisphere tend to be much older. Moreover, they have been extensively modified by meteorite impacts, wind erosion and partial burial. Those influences, together with the fact that the terrain is heavily cratered, make the southern hemisphere volcanoes difficult to recognize. The difficulties are compounded by the fact that the photographic coverage of the southern hemisphere was poor because *Mariner* 9 made most of its pictures of the hemisphere before a dust storm that engulfed the planet during the early weeks of the mission had fully abated.

Nevertheless, several unambiguously volcanic features have been identified in the southern hemisphere. One, Tyrrhenum Patera, consists of a central caldera surrounded by a fracture ring; numerous radial ridges, depressions and channels give the feature a starlike appearance. The entire structure is some 400 kilometers across. Two other structures, Hadriaca Patera and Amphitrites Patera, are volcanoes on the rim of Hellas, a large impact basin. Amphitrites Patera may be larger in diameter than Olympus Mons, but it is less than a kilometer high. Hadriaca Patera is approximately four kilometers high. The lava seems to have erupted mostly from the center of each feature, creating a strong radial pattern as it flowed from the center to the outlying areas. The low vertical dimension and the broad horizontal dimensions of these features once more imply that the lava was not very viscous.

Hadriaca Patera and Amphitrites Patera are of particular interest in that they indicate volcanism started quite early in the history of Mars. The number of impact craters superposed on these features suggests that they came into being near the end of the time when the planet was being heavily bombarded by debris left over from the formation of the solar system. On the earth's moon the bombardment began to subside some 3.9 billion years ago, and there are reasons to believe it subsided on Mars at the same time.

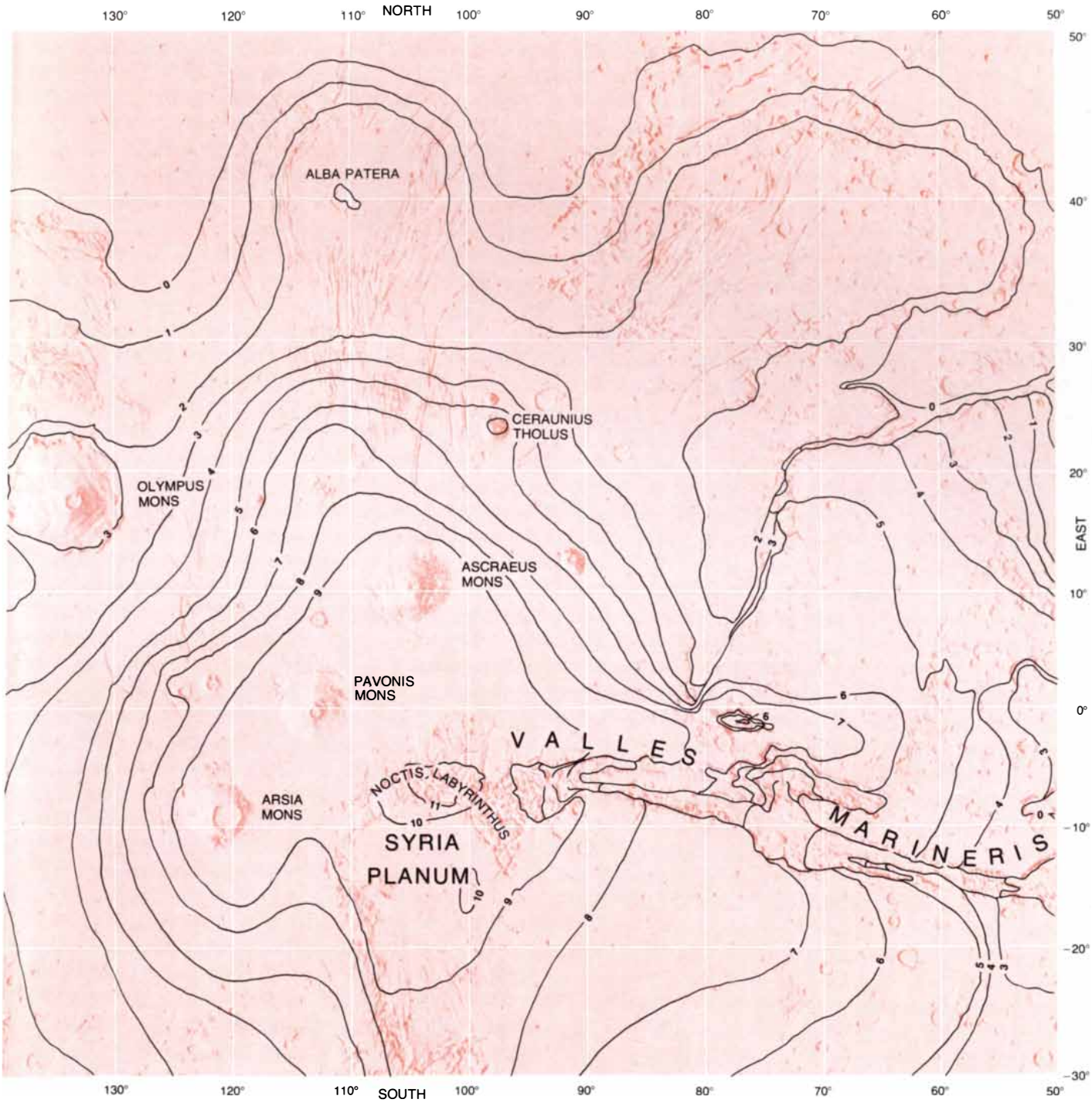
Exactly how does the volcanic activity on Mars compare with such activity on the earth? The theory of plate tectonics holds that the rigid crust of the earth is divided into huge plates that move laterally with respect to one another. The plates diverge at the mid-oceanic ridges where new crust forms, and they converge and ride over or under one another at subduction zones, which are generally recognizable as chains of islands and young mountains. The type of volcanic activity on the earth varies according to its location with respect to the plate boundaries, and all types are affected by the motion of the plates. First, at mid-oceanic ridges new crust is



THARSIS REGION ON MARS is the most prominent volcanic region on the planet. It is shown here on part of a topographical Mars chart produced by the U.S. Geological Survey. The contour lines, which are omitted on the volcanoes themselves, mark ele-

continually being formed by the quiet upwelling of magma that fills the void between the diverging plates. Second, where the plates converge one plate rides under the other and is partially melted. Third, volcanoes can arise in areas in the middle of a plate when the plate drifts over a "hot spot" in the mantle. A last possible major type of volcanism is represented by the flood basalts of the Columbia River Basin in the U.S. and the Deccan Trap in India. How the flood basalts fit into the theory of plate tectonics is uncertain, but they are particularly important for planetary studies because of the widespread volcanic floodplains visible on the moon, on Mars and possibly on Mercury.

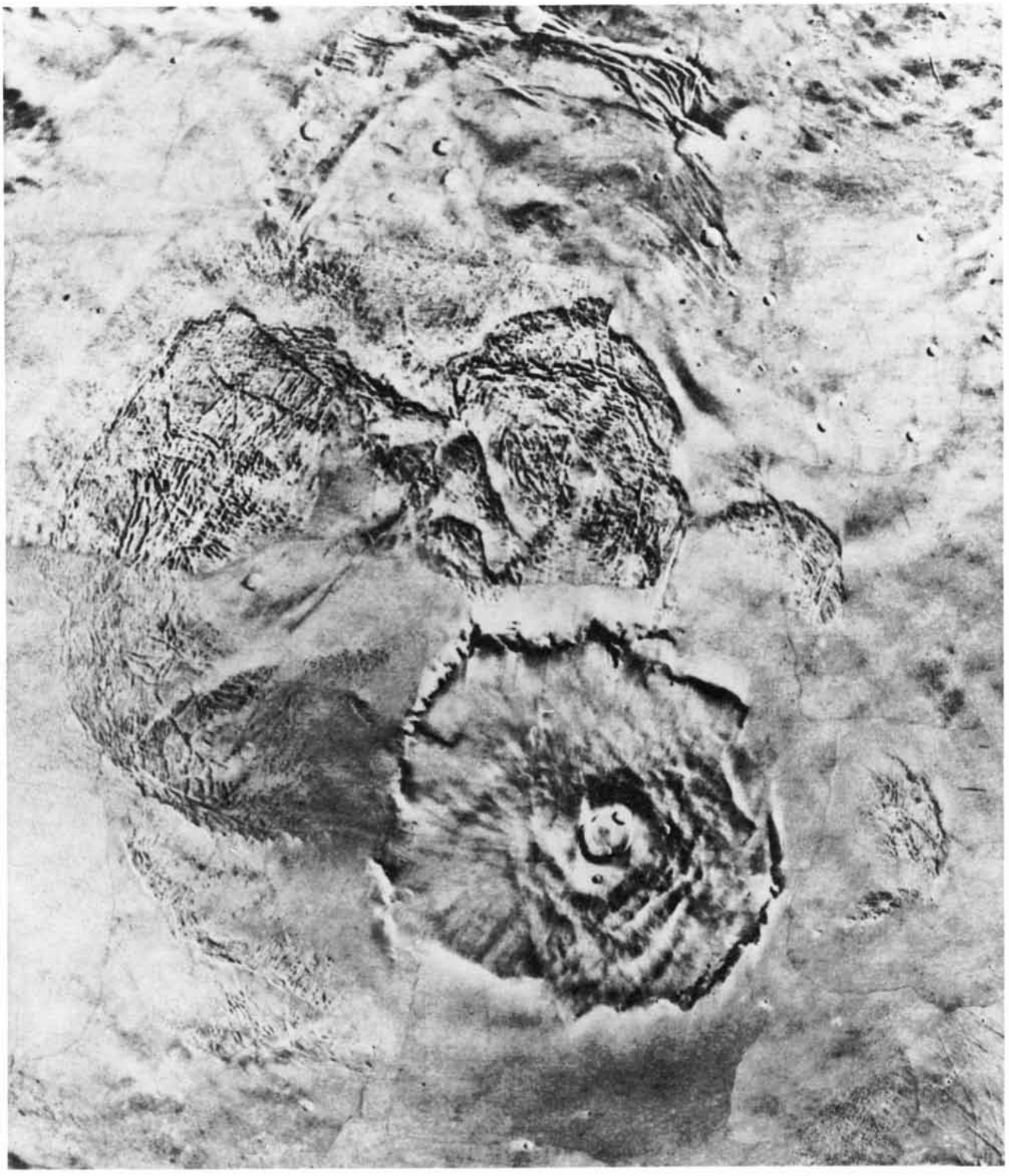
The crust of Mars appears not to be divided into plates that move with respect to one another. The equivalent of the volcanism along the mid-oceanic ridges on the earth is not observed on Mars. On the other hand, since that type of volcanism does not result in a morphologically distinctive volcanic feature, its absence is not obvious. Still, the



vations above a reference level where the atmospheric pressure on Mars is 6.1 millibars (.0061 the atmospheric pressure at sea level on the earth). The contours outline the broad bulge, informally called the Syria Rise because it is highest in Syria Planum, 5,000 kilometers across and seven kilometers high in the Martian crust on which the three shield volcanoes Ascraeus Mons, Pavonis Mons and Arsia Mons are located. Fractures radiate outward for several thousand kilometers from the center of the bulge and were apparently created at the same time as the bulge. The great canyon of Mars, Valles Marineris, is part of the fracture system and is also aligned along the radial direction. The plain Chryse Planitia mentioned in text is just off map to the right and north of the equator.

curved chains of folded mountains that typify the junctions of converging plates on the earth are conspicuously absent on Mars; indeed, Martian compressional features of any kind are rare. If the crust of Mars is not divided into moving plates, however, it is not surprising that the volcanic features that on the earth characteristically form at the junction of plates and result directly from their motion are rare on Mars, if indeed they exist at all.

Volcanism at the subduction zones on the earth tends to be violent and pyroclastic because the lava is highly viscous and contains many volatile substances. The volcanoes of the "belt of fire" that traces the subduction zones around the Pacific are characterized by explosive activity and by the large volume of ash they eject. The type of vol-



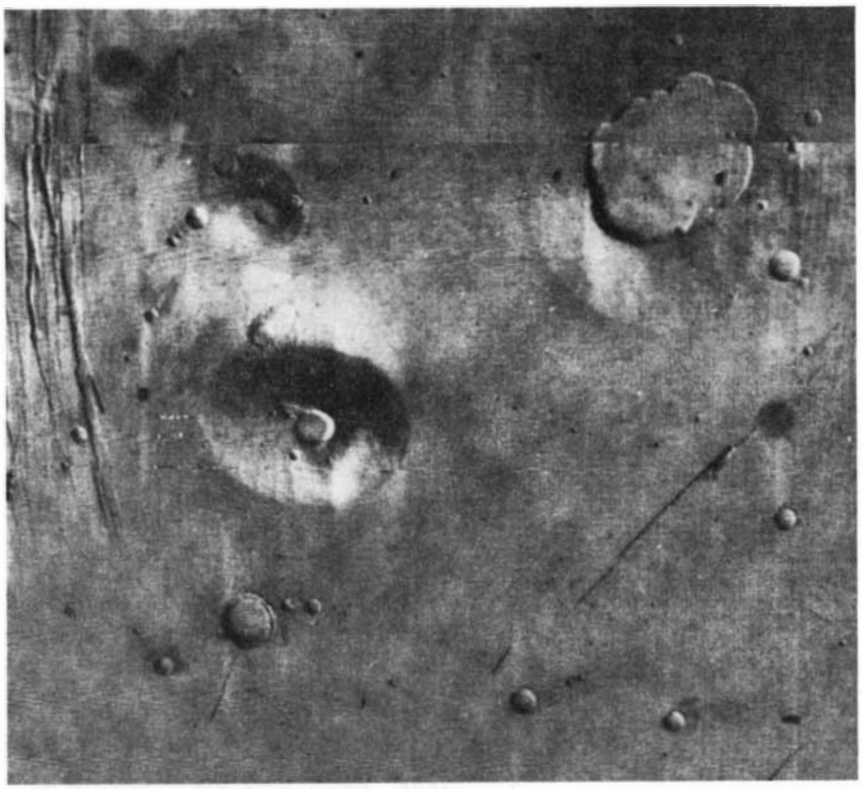
AUREOLE OF HIGHLY TEXTURED TERRAIN around the large shield volcano Olympus Mons is clearly related to the central volcano, as is shown in this mosaic of photographs. How it was formed is not yet clear. The main shield of the volcano is between 500 and 600 kilometers in diameter and ends abruptly in a cliff that in some places stands four kilometers above the surrounding plains.

cano that is created at the subduction zones is a conical stratovolcano built up of layers of ash and lava. The stratovolcano is rare on Mars, although several features that appear to be small volcanic cones have been recognized. The stratovolcanoes are much smaller than the shield volcanoes, however, and it should be emphasized that the small scale of any such feature on Mars compared with the scale of the shields would make it difficult to unambiguously recognize and identify. Perhaps there are many such features on Mars, but if there are, they are not recognized. In any case lines of stratovolcanoes corresponding to island chains on the earth have not been found.

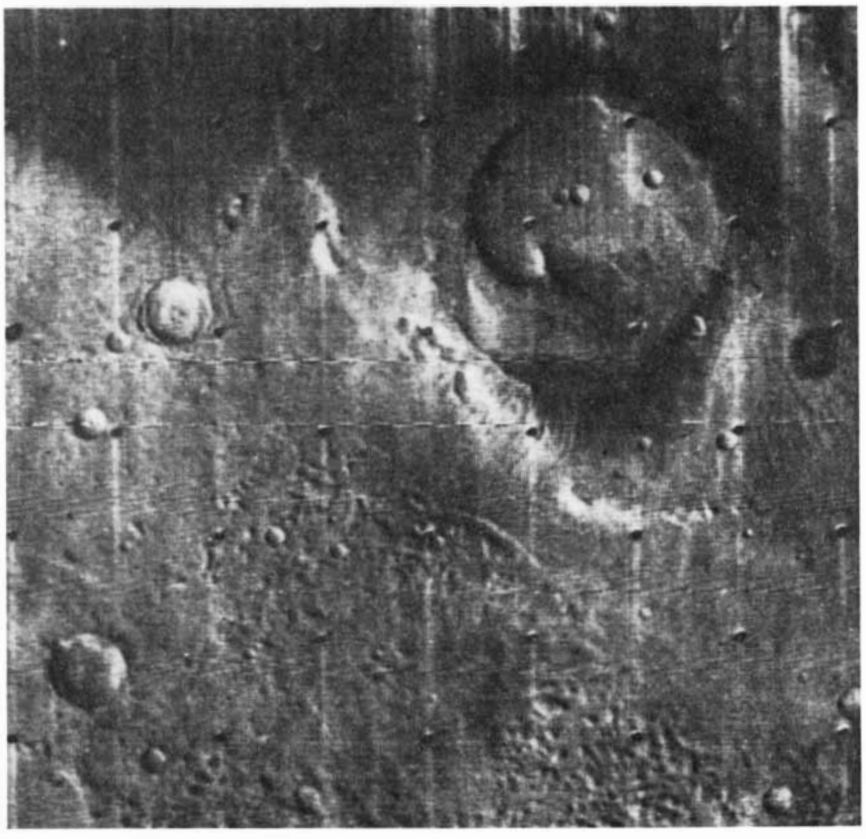
The two remaining major types of terrestrial volcanoes, the volcanoes that develop over hot spots and the flood eruptions, are well represented on Mars. The largest and best-documented terrestrial examples of the features that have developed over hot spots are the Hawaiian shield volcanoes. The active shields lie at the southeastern end of the Hawaiian-Emperor chain, a line of extinct volcanoes that stretches several thousand kilometers across the Pacific. The volcanoes progressively increase in age with distance from the present focus of volcanic activity on the southeastern side of the main island of Hawaii.

The chain is believed to have come about as a result of movement of the lithosphere (the crust and part of the upper mantle) over a hot spot in the upper mantle. As the plate moves to the northwest new volcanoes erupt to the southeast and old ones die out, leaving a trail of extinct volcanoes that traces the path of the plate over the hot spot. Each volcano in the chain appears to go through a similar evolutionary cycle. Initially the shield is built rapidly by the quiet effusion of lava; that stage is followed by a stage that is explosive and less productive of lava; then the volcano becomes extinct.

In many ways the large shield volcanoes of Mars resemble the Hawaiian shields in the early building stage. As we have seen, the Martian volcanoes' slope and surface texture indicate that they were built by low-viscosity lava. The central calderas have crisp curved walls and flat floors similar to the calderas of Mauna Loa and Kilauea in Hawaii, both of which are in the building stage. The Martian volcanoes evidently do not, however, pass through the final pyroclastic stage. In the Hawaiian volcanoes cinder cones tend to fill the summit caldera in the last building stage, which completely transforms



CERAUNIUS THOLUS is the dome-shaped shield volcano that appears at lower left center in this photograph. A channel two kilometers wide runs from the caldera at its summit down to a pool of lava on plain. Other volcanic structures appear to left and right of center.



APOLLINARIS PATERA, shown near the top right corner, is an old and partly eroded volcano about 2,000 kilometers to the southeast of the Elysium region (*see map on next page*). It appears to be older than any of the volcanoes in either Tharsis region or Elysium.

the overall appearance of the volcano.

No cinder cones can be seen within the caldera of any Martian shield volcano. This may merely indicate that the wind has blown away all the pyroclastic products or that the cones are too small or too low in profile to be seen. On the other hand, in view of the total lack of evidence of cones in the calderas of the chain of Tharsis shields, it seems more likely there never was a late pyroclastic stage. If that is so, it suggests that the pyroclastic stage in the evolution of the Hawaiian shields is the result of changes near the surface connected with the motion of the plates and not the result of some change deep within the mantle.

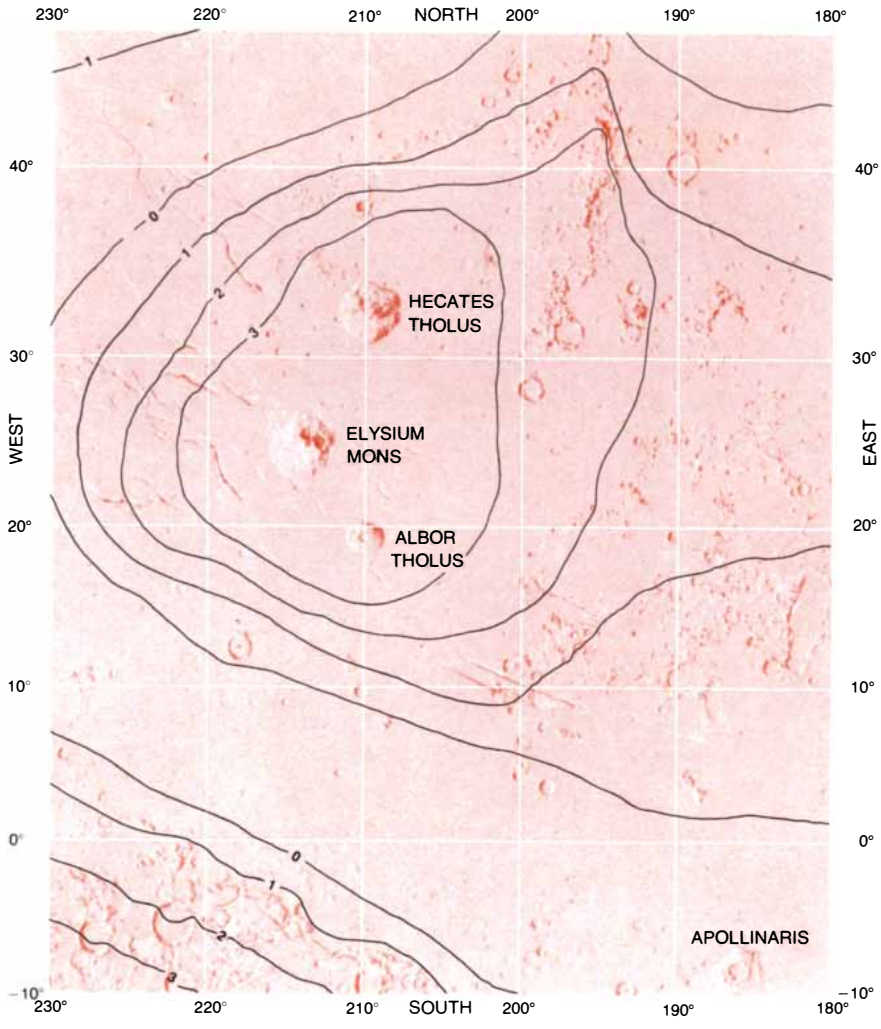
The most obvious difference between the Hawaiian shields and the Martian shields is their size. That difference is also understandable when it is viewed in terms of the motion of plates. The Hawaiian shields are relatively short-lived. They grow only as long as they remain over the magma source within the stationary mantle, and they die as the motion of the plate carries them away. The Martian shields, on the other hand, are fixed over a magma source, and so they continue to grow for a long period of time, constrained only by the supply of magma and the requirements for a hydrostatic head (internal pressure) large enough to force the magma to the surface. As a result a huge volcanic pile accumulates.

Data recently gathered on the age and volume of the volcanoes of the Hawaiian-Emperor chain indicate that the average rate of eruption over the past 50 million years from the hot spot currently under Hawaii is .1 cubic kilometer of lava per year. At that rate the main edifice of Olympus Mons would have formed in 200 million years, a relatively short span in the history of Mars. The large size of the Martian shields therefore probably reflects the stability of the planet's crust and does not indicate a high rate of lava eruption. Indeed, if the features are as old as we believe they are, it is likely that the eruption rates on Mars are very small and that the planet as a whole is far less active volcanically than the earth. This conclusion is supported by the survival of so much of what appears to be primitive cratered terrain.

Tharsis, the planet's most prominent volcanic region, lies on a broad bulge in the Martian crust near the center of an extensive array of radial fractures. The bulge is some 5,000 kilometers across and seven kilometers high at its highest point. The center of the bulge is to the southeast of the Tharsis region in Syria Planum, close to the intricately fractured Noctis Labyrinthis at the western edge of the great canyon Valles Marineris. The bulge, informally called the Syria Rise, is asymmetrical: it is more than twice as steep on its northwestern flank, where the Tharsis volcanoes are, as it is on its opposite flank. There appears to be a similar bulge in Elysium, the second most prominent volcanic region on the planet, although the bulge is smaller, only 2,000 kilometers long and two or three kilometers high, and the fractures are absent.

Although in general the fractures radiate from the center of the Syria Rise, they deviate locally from the radial direction and tend to be concentrated in specific areas. The most intensely fractured area is to the northeast of Tharsis, where a fanlike array of fractures converges on the line of Tharsis volcanoes. An opposing fan to the southwest of the volcanoes is less well developed. Another intensely fractured zone runs south-southeast from Pavonis Mons, and the development of the canyon Valles Marineris to the east was possibly controlled by the fractures. It is almost certain that intensely fractured rocks underlie most of the young plains in the Tharsis region, because wherever the older terrain is exposed it is quite broken and rugged. So widespread are the fractures around the Syria Rise that they extend over almost a complete hemisphere of the planet. Clearly their creation was a major event in the history of Mars. That event was probably connected in some way with the formation of the bulge and with the focus of volcanic activity in the region.

Any theory about the origin of the Syria Rise must account for one peculiar feature: the lack of what is called isostatic compensation. On the earth mountains and other highlands generally consist of light material with deep roots extending into the heavy mantle; lowlands



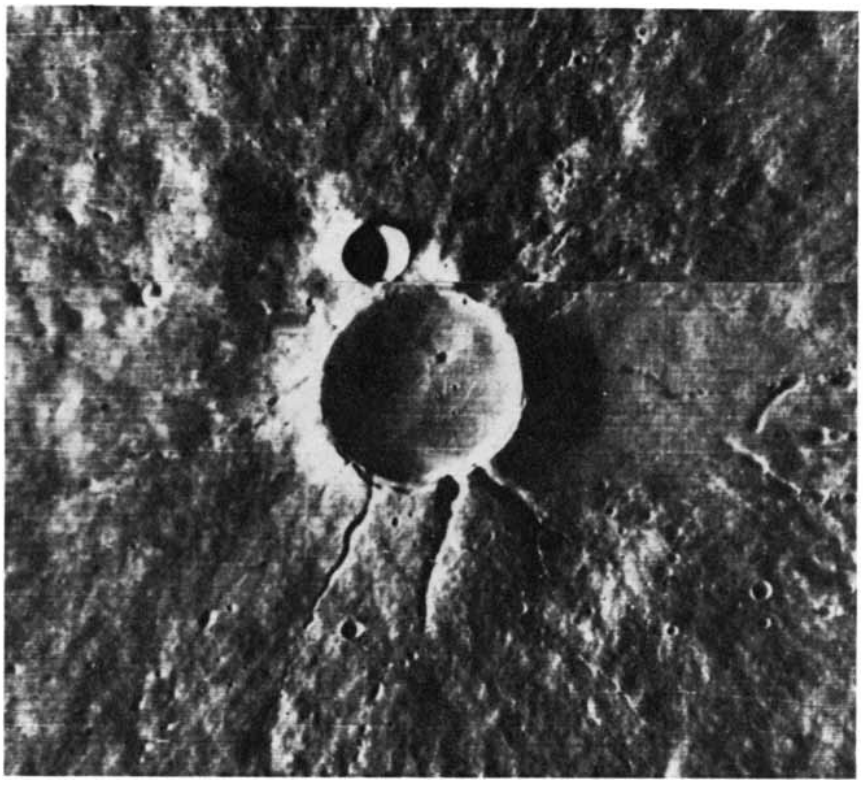
ELYSIUM REGION ON MARS is the second most prominent volcanic region on the planet. It too is characterized by a broad bulge in the Martian crust that is somewhat smaller than the bulge in Tharsis region: 2,000 kilometers across and two or three kilometers high.

are commonly areas over thinner regions of the crust, so that the mantle is closer to the surface. The result is that topography and gravity are not correlated as one might expect: there is usually no positive gravitational anomaly (increase in gravity) over mountains and no negative gravitational anomaly (decrease in gravity) over plains.

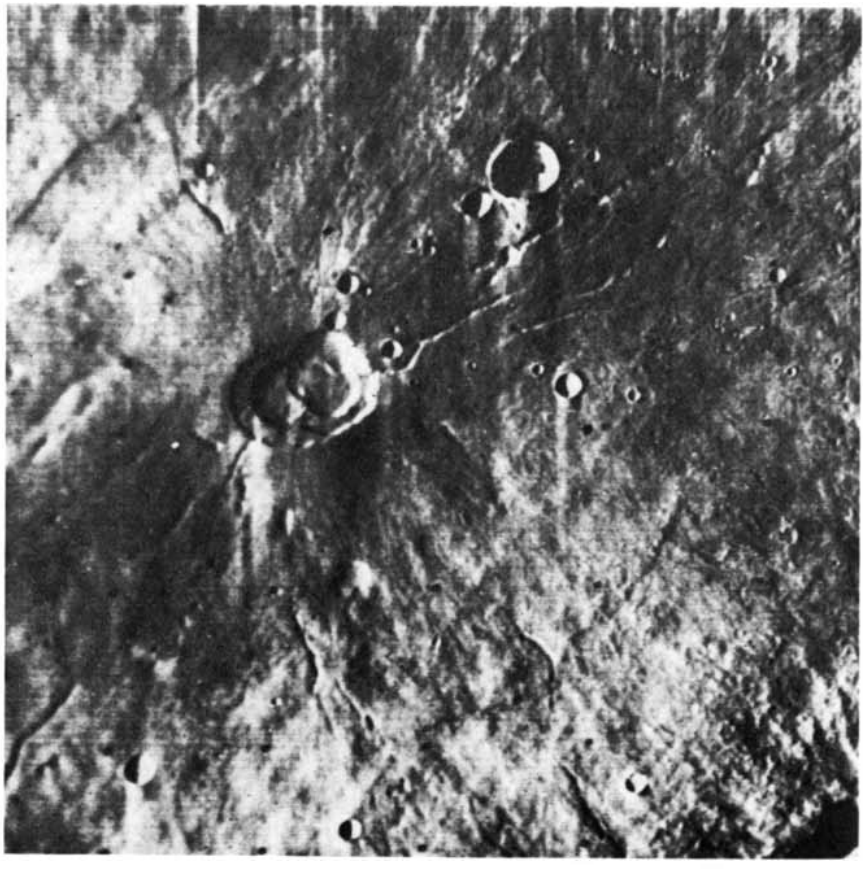
The same is true for Mars—except for the region of the Syria Rise and its immediate surroundings. Detailed analysis of the orbital path of *Mariner 9* as it was tracked from the Jet Propulsion Laboratory has shown that there is a positive gravitational anomaly over the Syria Rise and a corresponding negative gravitational anomaly in the adjacent basins of the Amazonis Planitia and Chryse Planitia. The data appear to imply that somehow material was transferred internally from the low areas to the high ones during the formation of the Syria Rise and that the planet failed to compensate for the redistribution afterward. If the bulge formed a billion years ago, as seems likely, the lack of isostatic compensation places a lower limit on the internal viscosity of the Martian mantle.

What created the Syria Rise is not known. Two possibilities suggest themselves: convection in the mantle or the presence of a hot spot. Convection in the mantle has long been invoked to explain why certain types of volcanic activity are localized on the earth. Its most recent variant is the plume hypothesis, which proposes that certain kinds of volcanism are found above regions where there is a plume, or concentrated upwelling, of material from deep in the mantle. The mechanical effects of the plume, and the thermal expansion and increased partial melting of the crust over the plume, cause the overlying lithosphere to be fractured and ultimately magnetized. The concept has been extended to suggest that stresses at the base of the lithosphere may become so great that the lithosphere breaks up into new plates moving away from the point of upwelling. If such a plume caused the Syria Rise, it was either much larger than the plumes proposed to explain the localization of terrestrial volcanoes, or the effects of the plume were amplified by their sustained action on a stable crust that is much thicker than the crust of the earth.

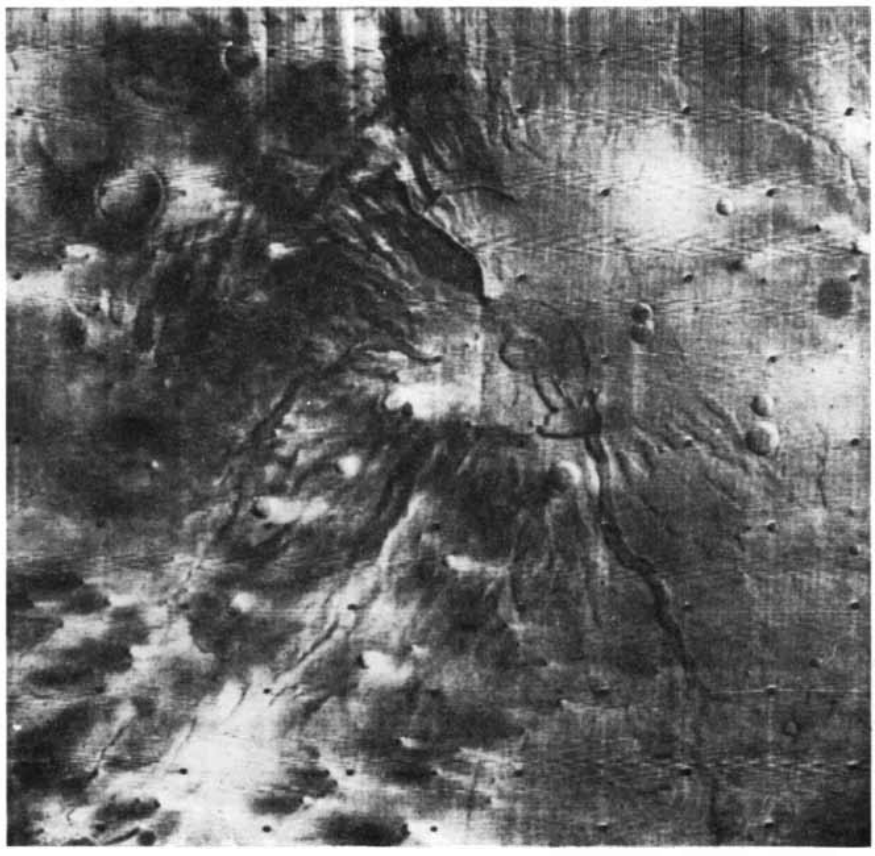
The closest equivalents on the earth to the Syria Rise on Mars are broad upwarps ranging between one kilometer and 10 kilometers in height and up to several thousand kilometers across. Because of the motion of the plates, how-



SUMMIT OF ELYSIUM MONS is shown in this high-resolution photograph. Channels, lines of depressions suggestive of collapsed lava tubes and other fine features radiate from a central caldera (*larger crater*) 13 kilometers across. The smaller pit is an impact crater.



SUMMIT OF HECATES THOLUS shows a complex caldera 12 kilometers in diameter with a number of fissures, lines of vents and channels extending from it down flanks of shield.



TYRRHENUM PATERA (*top right*) is one of the oldest volcanic features on Mars. It is located in the southern hemisphere in the region of Hesperia, where relatively old volcanic plains appear to cover most of the large craters in the region. The center of Tyrrhenum Patera is a complex depression that is surrounded by a fracture ring about 45 kilometers in diameter. Channels, low ridges and depressions extend some 200 kilometers from the ring.

ever, such upwarps tend to be short-lived compared with the Syria Rise. It has been suggested that in some cases as the upwarp is formed rifts develop and the lithosphere is divided into new plates. On Mars the crust may be too thick for it to break up. Its thickness can be estimated from the height of the volcanoes. The hydrostatic head needed to pump magma to the top of the Tharsis volcanoes requires that the magma be melted at a depth of at least 200 kilometers below the surface, about four times greater than the depth required for the volcanoes of Hawaii. Thus the Martian lithosphere at the time the Tharsis volcanoes formed was considerably thicker than the lithosphere of the earth. It may have been subjected to stresses similar to those that caused the earth's lithosphere to break up, but it failed to do so because of its great thickness. Hence within the Martian crust the effects of the kind of stresses that caused the breakup of the earth's crust may still be preserved, whereas at the same time the effects of plate motion are absent. Therefore on Mars one is able to examine the consequences of deep-seated processes more directly than one can study them on the earth.

With the data from *Mariner 9* in hand investigators have devoted considerable attention to the problem of determining the absolute ages of the surface features on Mars. The subject is controversial, and differences between the various interpretations have by no means been resolved. The main difficulties lie in assessing the rate at which impact craters form on the Martian surface and the rate at which they are destroyed by erosion and burial. We have good data on the impact rates on the moon, at least over the past four billion years. They have been obtained by comparing the number of craters superposed on a geological unit with the absolute age of the unit determined from samples brought back from the moon.

On the moon the impact rate appears to have been high about 3.9 billion years ago; then it declined rapidly to a much lower rate and has remained low ever since. That history has resulted in the well-known dichotomy of the lunar surface: its separation into densely cratered uplands and sparsely cratered plains. The densely cratered areas formed before the rapid decline in the impact rate; the sparsely cratered areas were created after the decline. Surfaces with intermediate crater densities are rare because they could form only during the short interval at the end of the decline.

Until we have absolute dates from soil samples on Mars, the impact rates there must be inferred indirectly. There are two basic theoretical models. In one the impact-flux history of Mars and of the moon are quite similar; in the other the recent impact rates on Mars are considerably higher than those on the moon. Clearly the ages of sparsely cratered areas given by the second model are younger than the ages given by the first.

Several models have also been proposed for the history of the obliteration of the craters. On the moon the dominant mechanism for destroying craters is other impacts. Additional factors, such as slumping of the soil caused by seismic shocks, may be important locally, but the obliteration history can in most cases be described satisfactorily on the basis of impact alone. On Mars the wind, in addition to impacts and volcanism, is an effective agent for destroying craters; in some places water may be too. The effectiveness of such aeolian and fluvial processes can range widely both in place and in time, so that rates cannot be definitively modeled.

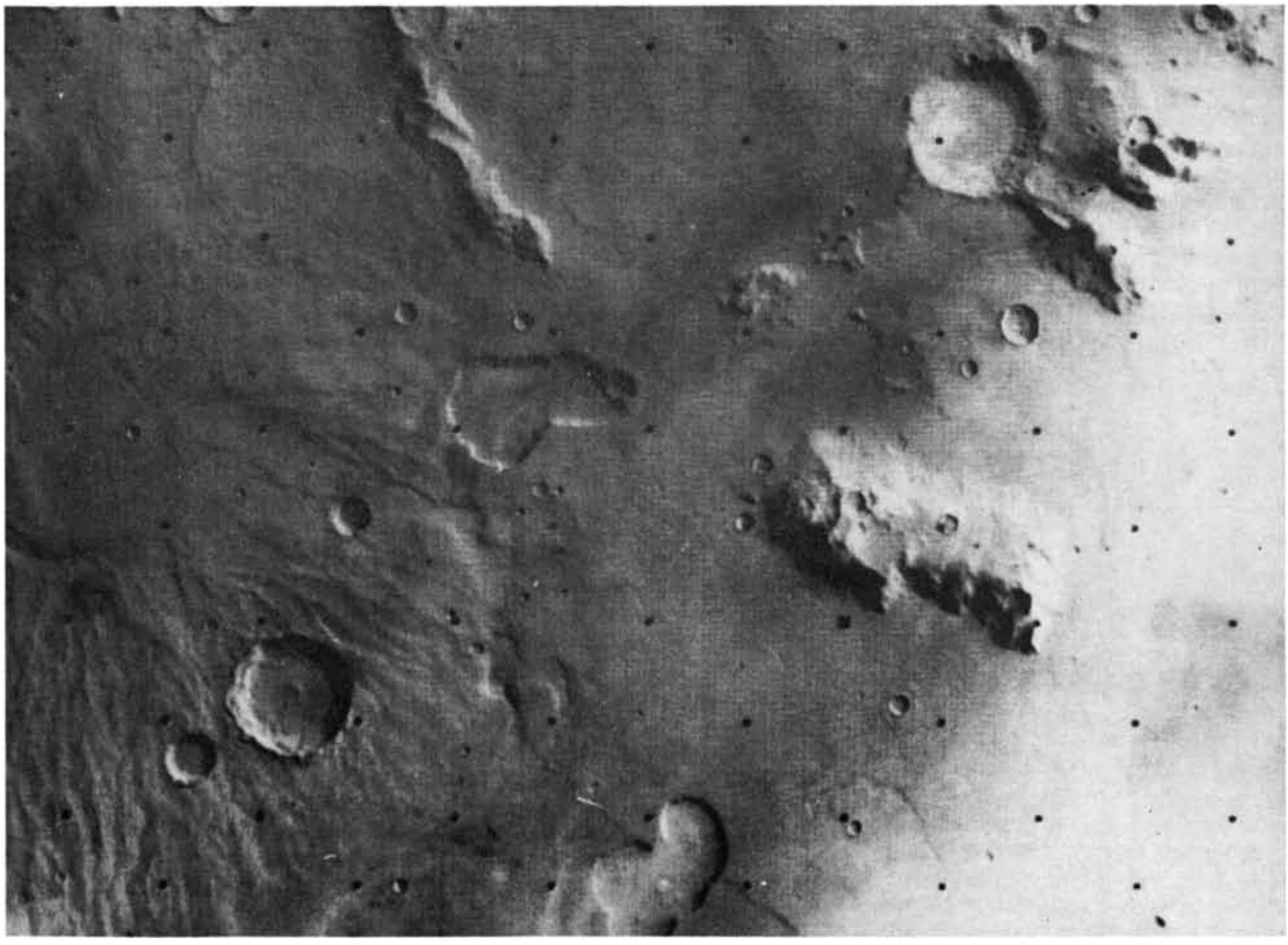
All investigators of the problem agree that major episodes of obliteration are required to explain the paucity of small craters in the densely cratered terrain and the contrast in shape between the small craters and the large ones. Where the investigators differ is in the number and timing of the obliteration episodes. For example, Laurence A. Soderblom of the U.S. Geological Survey infers that the impact-flux history of Mars has been similar to that of the moon and that at least in the equatorial region the high-obliteration period coincided with the early high-flux period and ended some 3.8 billion years ago. Other workers infer higher impact rates for Mars than for the moon and say that there were major obliteration events as recently as between 200 and 450 million years ago. The effect of the Soderblom model is to space all the distinguishable volcanic events out over the past four billion years of Mars's history; the other models tend to compress all distinguishable events into the recent past. A detailed comparison of the various models lies outside the scope of this discussion. Here I shall assume that the Soderblom model is correct; at the same

time I realize that several aspects of obliteration history have yet to be reconciled with this model.

The plains provide the most reliable crater statistics. The plains around the shields in the Tharsis region yield an average age of some 200 million years. Those plains show a wide range in crater density, and parts of them may be either considerably younger or considerably older than the average. The plains to the north of the Tharsis region near the fracture ring of Alba Patera, the plains to the southeast of the Tharsis region in Syria Planum and Sinai Planum, and the plains of Elysium all seem to be between one and two billion years old. Some older plains, such as those of Lunae Planum to the east of the Tharsis region and those of Hesperia Planum around Tyrrhenum Patera, yield ages as great as between 3.5 and four billion years.

The volcanoes can also be dated, but the figures are much less reliable because of poor statistics and the difficulty of distinguishing between volcanic craters and impact craters. The volcanic craters, however, generally yield ages close to the age of the surrounding plain. Olympus Mons appears to be 200 million years old, the two northernmost volcanoes of the Tharsis ridge 400 million years old and the southernmost Tharsis volcano (Arsia Mons) 800 million years old. The smaller shield volcanoes in the Tharsis region all appear to be between one and two billion years old, although here again the statistics are poor. The fracture ring Alba Patera and the Elysium volcanoes also seem to be between one and two billion years old. Ages of 3.5 or four billion years are derived for Tyrrhenum Patera in Hesperia and the two volcanoes, Hadriaca Patera and Amphitrites Patera, at the edge of Hellas. Thus according to the Soderblom model the formation of shield volcanoes and lava plains has spanned the entire history of Mars. It should be emphasized, however, that the ages are extremely tentative and quite dependent on the model.

Is Mars volcanically active today? Several arguments suggest that it is. First, on the Soderblom model the volcanoes seem to have been active at least until 200 million years ago, that is, until 96 percent of the planet's history had passed. If that is the case, the volcanic activity of Mars has spanned billions of years, and it is improbable that we have sampled the planet just as it has turned off volcanically. Second, large variations in the crater densities on the lava plains in the Tharsis region suggest that parts of the plains are younger than the average age of 200 million years. Third, the fact that we do not recognize surfaces younger than 200 million years old does not mean that they do not exist. If the Soderblom chronology is correct, then the rate of volcanic activity is low. Accordingly we should not expect to find that extensive volcanic deposits have accumulated just in the small fraction of the planet's history represented by the past 200 million years. Moreover, any of the chronologies that are alternative to the Soderblom model give even younger ages and thus increase the likelihood of present-day volcanic activity on Mars.



HADRIACA PATERA, at the far left, is another ancient volcanic feature, located on the northeastern rim of the impact basin Hellas in the southern hemisphere. At its center is an ill-defined smooth, circular area some 70 kilometers in diameter from which low ridges extend outward as far as 300 kilometers. In several places the pattern is interrupted by large, clearly defined impact craters.

The Search for New Families of Elementary Particles

At least one particle that fits none of the established families has emerged from experiments with high-energy neutrinos. The new particle must exhibit some hitherto unobserved property of matter

by David B. Cline, Alfred K. Mann and Carlo Rubbia

In the continual give and take between experimental discovery and theoretical understanding in modern physics, the present moment is one of happy coincidence. A number of new machines for creating elementary particles have recently begun operating, and new devices and techniques for detecting the particles have been invented. At the same time theoretical innovations have provided a basis for the expectation that there are unknown states of matter to be discovered with the new machines. This convergence of theory and experiment has inspired a diligent search for new families of elementary particles. One result of that search was the discovery in 1974 by investigators at the Brookhaven National Laboratory and at the Stanford Linear Accelerator Center of the particles called *J* or psi. At least one other particle with unusual properties has been detected in experiments at the Fermi National Accelerator Laboratory (Fermilab). It is not yet clear whether or not the two new kinds of particle are closely related.

The particle observed at Fermilab has a mass and a lifetime that place it in a region of the elementary-particle spectrum that remains almost unexplored. The particle was created at Fermilab through the interactions of an intense beam of high-energy neutrinos, and its existence was deduced indirectly from the presence of neutrinos among the products of its decay. The experiment depended critically on the singular properties of neutrinos, and it could not be performed until the apparatus and techniques for dealing with those elusive particles became available.

The new particle is not merely an addition to one of the numerous known families of elementary particles; its significance is that it reveals an entirely new family. The existence of the new family in turn implies that there is some hitherto unobserved property of matter that distinguishes the particles in the new family from all the more familiar ones. So far the nature of that property is obscure, and in the interpretation of the Fermilab experiments there is still more confusion than certainty. There is also, however, more excitement in particle physics today than at any other time in the past few decades.

Physicists recognize four basic interactions of matter or, expressed another way, four fundamental forces. They are called the strong, the electromagnetic, the weak and the gravitational forces. If the strong force is set equal to 1, then the strength of the electromagnetic force is 1/137, or very roughly 10^{-2}, that of the weak force is 10^{-13} and that of gravitation is 10^{-39}. When two particles interact through the agency of one of these forces, the force is thought to be transmitted by the exchange of an intermediate particle peculiar to that force. The strength of the interaction is related to the time required for the exchange and to the distance over which the exchange can be accomplished. The time and distance are in turn related to the properties of the exchanged particle.

Strong interactions ordinarily take place in about 10^{-23} second, and the strong force is regarded as being strong precisely because it can bring about a change in such a brief interval. The electromagnetic interaction takes about 137 times longer, or roughly 10^{-21} second. The time required for weak processes varies considerably with the energy of the interaction; it can be as little as 10^{-18} second or as much as 15 minutes.

The range of the fundamental interactions is thought to be inversely proportional to the masses of the exchanged particles. For the electromagnetic interaction this relation has been confirmed: the electromagnetic force is transmitted by photons, which have zero mass, and the range of the force is infinite. The particle thought to propagate the gravitational force, the graviton, should also be massless, in order to account for the infinite range of gravitation, but gravitons have not yet been detected. The strong force is transmitted by the particles called mesons, which do have mass, and as a result the strong force has a finite range. The least massive meson, the pi meson, or pion, has a mass of about 140 MeV, or million electron volts. (In modern high-energy physics it is customary to refer to the mass of a particle in terms of its energy equivalent.) The range of the strong interaction is about 10^{-13} centimeter, which corresponds approximately to the size of the particles that make up the atomic nucleus.

The weak force has a very short range; when it was first described by Enrico Fermi in 1934, its range was thought to be zero, which would have required an intermediate particle of infinite mass. More recent theories suggest that the weak force is extended in space, although its range is exceedingly small: less than 10^{-15} centimeter, or a hundredth that of the strong force. The particles exchanged in weak interactions are therefore expected to be very massive; measurements at Fermilab suggest a minimum mass of 20 GeV, or billion electron volts. These particles have not been detected, but they have been

named; for many years they have been called *W* particles.

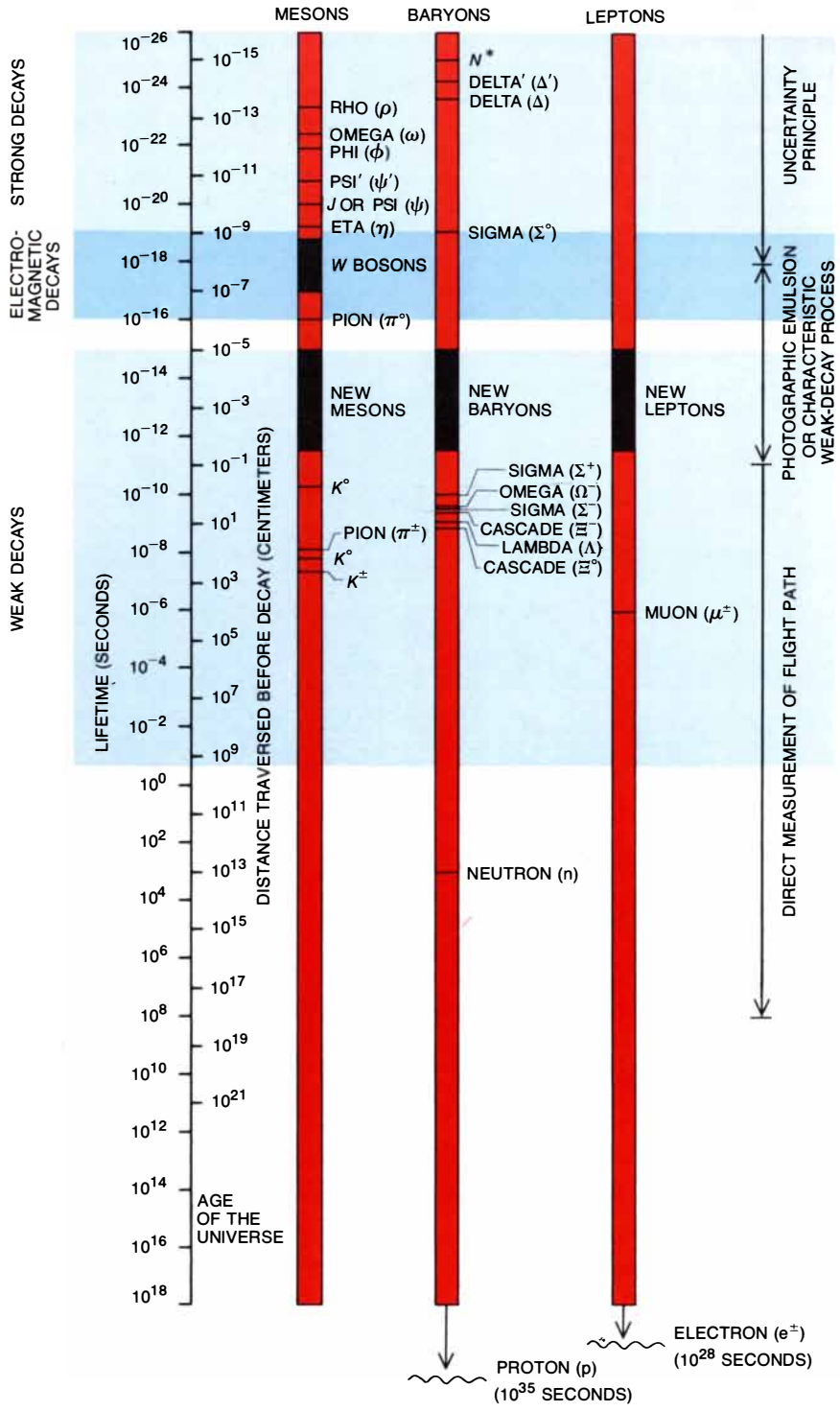
Of the four basic interactions only gravitation is universal. The weak force affects every particle but one, the exception being the photon. The electromagnetic force is confined mainly to those particles that bear an electric charge. The strong force is the most selective of the forces; it serves as the basis for dividing the elementary particles into two broad groups, the hadrons, which "feel" the strong force, and the leptons, which do not.

Within those groups particles are classified by an elaborate system of quantum numbers, assigned on the basis of experimental measurements. Quantum numbers describe the properties of a particle in the same way that a list of facial features might describe a person. The complete list of quantum numbers defines the particle uniquely; together with the mass, such a list represents all one can possibly know about the particle.

Every quantum number corresponds to a property that is conserved during the interactions of particles, that is, the sum of the quantum numbers for a given property does not change during the interaction. Some quantum numbers, such as spin angular momentum and electric charge, are invariably conserved, and so is energy, or mass. In the strong interaction all quantum numbers are absolutely conserved, and in the electromagnetic interaction all but one (called isotopic spin) are conserved. In the weak interaction, however, a few quantum numbers are sometimes changed. The quantum numbers whose conservation is violated are all peculiar to hadrons, and they include the properties called strangeness, intrinsic parity and charge conjugation.

For the classification of particles the most important quantum number is spin angular momentum. When it is measured in the natural units of Planck's constant divided by 2π, its value is invariably an integer or half an integer. Particles with integral values of spin angular momentum are called bosons, because they obey what are called Bose-Einstein statistics. Particles with half-integral spin obey Fermi-Dirac statistics and are called fermions.

All the known leptons are fermions; indeed, all of them have the same spin angular momentum of 1/2. Among the hadrons, on the other hand, there are both bosons and fermions. The hadronic bosons are the mesons (the particles that transmit the strong force), and they have values of spin such as 0, 1, 2 or 3. The fermions form a separate group, called



LIFETIME OF A PARTICLE is determined primarily by its mode of decay, and the lifetime in turn determines how the particle can be detected. Many particles that can decay only through the weak interaction are relatively long-lived and travel a substantial distance before they decay. Their lifetimes can be determined by a direct measurement of their flight path in a particle detector. Particles that decay through the strong or electromagnetic interaction survive only briefly, and their flight paths are much too short to be measured. Their lifetimes can be determined, however, through an application of Werner Heisenberg's uncertainty principle, which states that the lifetime is inversely proportional to the uncertainty in the particle's energy, or mass. For particles with intermediate lifetimes, extending from roughly 10^{-18} second to 10^{-11} second, neither of these methods is practical. Some short particle tracks might be detected in photographic emulsions, but most particles with lifetimes in this region must be detected indirectly by the products of their decay. Several groups of particles have been predicted to lie in this region *(gray bars)*. They are expected to decay through the weak interaction, but faster than other weakly decaying particles.

baryons, with spin quantum numbers of 1/2, 3/2, 5/2 and so forth. Within each of the larger groups of hadrons the particles are assembled in families distinguished by a single value of spin angular momentum.

The photon and the W particles are neither hadrons nor leptons. They have a spin of 1 and are thus bosons; they are called vector bosons because the quantum-mechanical equation that describes particles with a spin of 1 takes the form of a four-dimensional vector. Gravitons, with a spin of 2, are also bosons, and so of course are the mesons that transmit the strong force. Thus all the fundamental forces are propagated by bosons.

The baryons with the lowest mass are those with a spin of 1/2. There are eight of them: two nucleons (N) with an average mass of 939 MeV, a lambda particle (Λ) with a mass of 1,116 MeV, three sigma particles (Σ) with an average mass of 1,193 MeV and two cascade particles (Ξ) with an average mass of 1,317 MeV. How are these differences in mass to be explained? They are usually attributed to a hypothetical fifth interaction, sometimes called the semistrong force. The semistrong force has neither the theoretical stature nor the experimental support of the four fundamental forces but is introduced solely to explain the "mass splittings" of hadrons. If the semistrong interaction could somehow be turned off, all the spin-1/2 baryons would have the same mass; they would degenerate into a single state. The semistrong force splits the degenerate state into particles of different mass, and the splitting is accompanied by the introduction of two new quantum numbers, isotopic spin and strangeness. Both are "internal" quantum numbers: their effects cannot be directly observed in the behavior of an individual particle but derive from the characteristics of the families of hadrons.

Through the semistrong interaction, particles with different values of strangeness or isotopic spin acquire different masses [*see top illustration on opposite page*]. The lambda and sigma particles, for example, both have a strangenenss of -1, but because they have different values of isotopic spin (0 and 1 respectively) they are split apart in mass by about 180 MeV. The cascade and the nucleon have the same isotopic spin but differ by two units of strangeness; as a consequence the cascade is about 375 MeV more massive than the nucleon.

By a similar mechanism the electromagnetic interaction splits each hadron state into particles of different mass and electric charge (and thereby violates the conservation of isotopic spin). If the electromagnetic force could be turned off, the two nucleons, for example, would be identical in mass and all other properties. Because the force cannot be turned off, the nucleon is split into two states, the proton with an electric charge of $+1$ and the neutron with a charge of 0. Because the electromagnetic force is less powerful than the semistrong force the mass splitting between the proton and the neutron is only 1.3 MeV, about 136 times less than that between the lambda and the nucleon.

The same mechanisms generate comparable mass splittings among the mesons [*see bottom illustration on opposite page*]. The lowest-mass mesons are those with a spin angular momentum of 0; they include three pions (π), four kaons, or K mesons (K), and the eta (η) and eta′ (η') mesons. The pion and K states differ in both isotopic spin and strangeness and are split in mass by about 350 MeV. The eta and eta′ are a special case: they have the same values of isotopic spin and strangeness but differ in yet

			STRONG INTERACTION	ELECTRO-MAGNETIC INTERACTION	WEAK INTERACTION
		INTERMEDIATE PARTICLE	MESONS	PHOTON (γ)	W^+ W^- W°
		MASS OF INTERMEDIATE PARTICLE	140 MeV AND GREATER	0	GREATER THAN 20 GeV
		RANGE	10^{-13} CENTIMETER	INFINITE	10^{-15} CENTIMETER
		TIME SCALE	10^{-23} SECOND	10^{-21} SECOND	10^{-10} SECOND
		ENERGY	CONSERVED	CONSERVED	CONSERVED
		MOMENTUM	CONSERVED	CONSERVED	CONSERVED
		ELECTRIC CHARGE	CONSERVED	CONSERVED	CONSERVED
		LEPTON NUMBER (L_e, L_μ)	UNDEFINED	CONSERVED	CONSERVED
		STRANGENESS	CONSERVED	CONSERVED	NOT CONSERVED
		CHARM	CONSERVED	CONSERVED	NOT CONSERVED
HADRONS	BARYONS	PROTON (p)			
		NEUTRON (n)			
		LAMBDA (Λ)			
	MESONS	PION ($\pi^\pm$)			
		PION (π°)			
		KAON ($K^\pm$)			
		KAON (K°)			
LEPTONS		ELECTRON ($e^\pm$)			
		MUON ($\mu^\pm$)			
		NEUTRINO (ν_e, ν_μ)			

FUNDAMENTAL FORCES, or interactions of particles, are thought to be transmitted by intermediate particles. The time scale of the interaction is a measure of its strength, and the range of the interaction is inversely related to the mass of the intermediate particle. Certain properties, such as energy, momentum and electric charge, are conserved in all interactions; certain others, however, defined by quantum numbers such as strangeness, are not invariably conserved in weak interactions. Charm is a hypothetical quantum number analogous to strangeness. The strong force affects only that class of particles called hadrons, which is subdivided into baryons and mesons. Electromagnetism affects all particles that bear an electric charge. The weak force is "felt" by all particles except the photon.

another quantum number—they are separated in mass by about 410 MeV. The pions are further split by the electromagnetic interaction into states with electric charges of 0, +1 and −1 and with a mass difference of about 4 MeV. The electromagnetic mass splittings of the kaons are more complicated because the kaons are strange particles. In addition to positive and negative states there are two neutral kaons, with opposite values of strangeness.

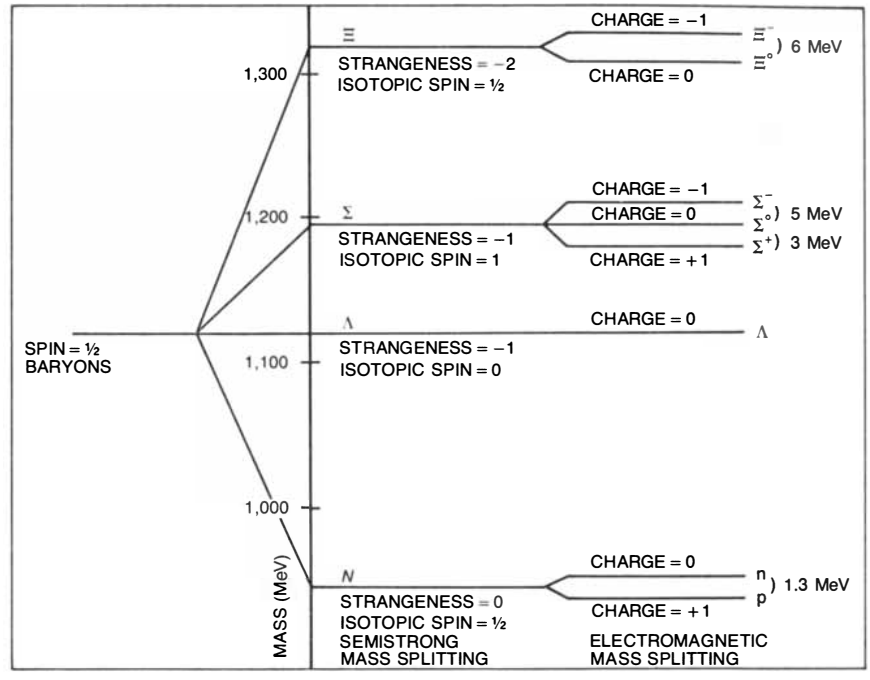
Mass splittings are also observed among the leptons, but as yet there is no satisfactory explanation of them. The entire known family of leptons consists of just four particles: the electron (e), the muon (μ) and two kinds of neutrino (ν_e and ν_μ).

The electron is exceedingly light, with a mass of about .51 MeV; the muon is about 200 times heavier, at 105.6 MeV. No mechanism is known to explain this mass splitting, and indeed the very existence of the muon is one of the most persistent mysteries of particle physics.

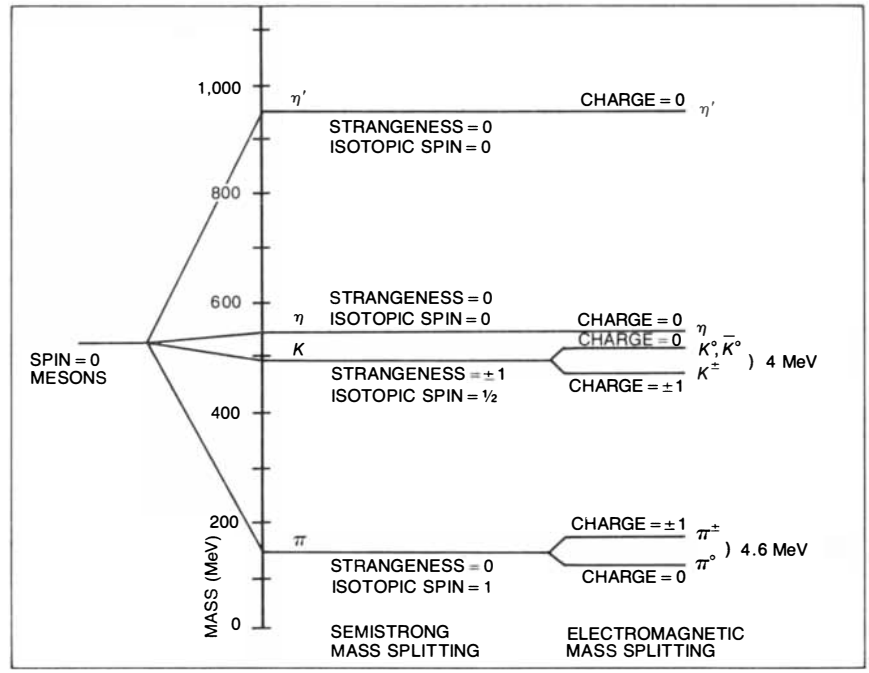
The electron and the muon each bear a distinctive lepton quantum number, designated L_e and L_μ. The quantum numbers are conserved in all interactions and are carried by the two kinds of neutrino. The neutrinos have no electric charge and therefore do not feel the electromagnetic force; as leptons they are also excluded from feeling the strong force. They can interact only by the weak force and as a result they scarcely interact with matter at all. They are thought to be massless, but that fact has not been established experimentally.

The largest of all mass splittings are those predicted for the vector bosons. As we have seen, the mass difference between the photon (with zero mass) and the W particles must be at least 20 GeV, and it could be much greater. The electrically charged W^+ and W^- bosons were introduced by Hideki Yukawa in 1935; the neutral W^0 boson was added to the family in the 1960's and led to the prediction of neutral weak currents, a type of interaction in which particles interact through the weak force without transferring electric charge. Neutral weak currents have recently been observed, although that cannot be considered conclusive evidence that the W^0 exists [see "The Detection of Neutral Weak Currents," by David B. Cline, Alfred K. Mann and Carlo Rubbia; SCIENTIFIC AMERICAN, December, 1974].

Recent predictions of the mass of the W^+ and W^- bosons range as high as 54 GeV, and the mass of the W^0 could well be as large as 80 GeV. These enormous mass splittings are postulated by ambi-



MASSES OF THE BARYONS are explained by postulating a single, hypothetical state that is "split" into particles of different mass. The baryons with a spin angular momentum of 1/2 include four kinds of particle: the nucleon (N), the lambda (Λ), the sigma (Σ) and the cascade (Ξ). They are split apart in mass by a hypothetical semistrong interaction, according to their values of the quantum numbers strangeness and isotopic spin. Each of these particles is then split by the electromagnetic interaction into states of different electric charge. The nucleon, for example, is split into the neutron and the proton. The electromagnetic mass splittings, which are smaller than the semistrong splittings, are not shown to scale.



MASS SPLITTINGS OF MESONS are larger than those of baryons, but they are generated by the same mechanisms. The mesons with a spin angular momentum of 0 are split by the hypothetical semistrong interaction into four states: the pion (π), the kaon (K), the eta (η) and the eta′ (η'). The pion and kaon are further split by the electromagnetic interaction into charged and neutral states. For particles that decay through the weak interaction the magnitude of the mass splitting between states influences lifetime of the particles; larger mass splittings promote faster decays. New families of particles are expected to have larger mass splittings, and hence faster weak decays, than the known mesons and baryons.

tious theories that attempt to show that the weak force and the electromagnetic force, in spite of their apparent dissimilarity, are in fact merely two manifestations of a single phenomenon [see "Unified Theories of Elementary-Particle Interaction," by Steven Weinberg; SCIENTIFIC AMERICAN, July, 1974]. The theories explain the mass splittings by assuming that in some imaginary initial state all four vector bosons were massless; it then follows as a consequence of the theories that three of the particles acquire huge mass while the fourth retains its zero mass.

The unified theories of the weak and electromagnetic interactions also suggest that there may be additional leptons, which would be much more massive than the known ones. Both the massive W bosons and the additional, heavy leptons play the same role in the theories: they eliminate certain infinite mathematical terms that would otherwise destroy the meaning and predictive value of the theories.

Among the scores of identified particles only a handful are stable and serve as components of ordinary matter. The photon is known to be stable; the most fundamental principles of quantum mechanics argue that it cannot decay. The proton and the electron are presumed to be stable, and that presumption is supported by experimental evidence indicating that their lifetimes must be many orders of magnitude greater than the conjectured age of the universe. Finally, the neutrinos are also thought to be stable, but that has not been unequivocally demonstrated.

All other particles, in seeking the state of lowest energy, decay more or less rapidly and eventually reach a final state consisting of stable particles. Some hadrons decay through the strong interaction. The eta′ meson, for example, can decay to yield an eta meson and two oppositely charged pions, a reaction written in symbols: $\eta' \rightarrow \eta + \pi^+ + \pi^-$. It is easy to verify that in this interaction mass is conserved: the eta′ has a mass of about 958 MeV, and the sum of the masses of the decay products is about 829 MeV, leaving an excess of 129 MeV that appears as the kinetic energy of the products. Electric charge is also conserved: the eta′ is neutral, and the charges of the decay products are 0, $+1$ and -1, for a net charge of 0. Similarly, it can be shown that spin angular momentum, strangeness and isotopic spin are all conserved. The decay takes place in the characteristic time span of the strong interaction, about 10^{-23} second.

Many particles cannot decay by the strong interaction because there is no combination of particles that has lower mass and that conserves all quantum numbers. A few hadrons, such as the neutral pion, decay electromagnetically. The neutral pion has the lowest mass of any hadron, and it cannot decay by the strong force simply because there are no strongly interacting particles of lower mass into which it could be transformed. Instead it almost always decays into two photons. This process too conserves mass, electric charge and all other quantum numbers. It requires about 10^{-16} second.

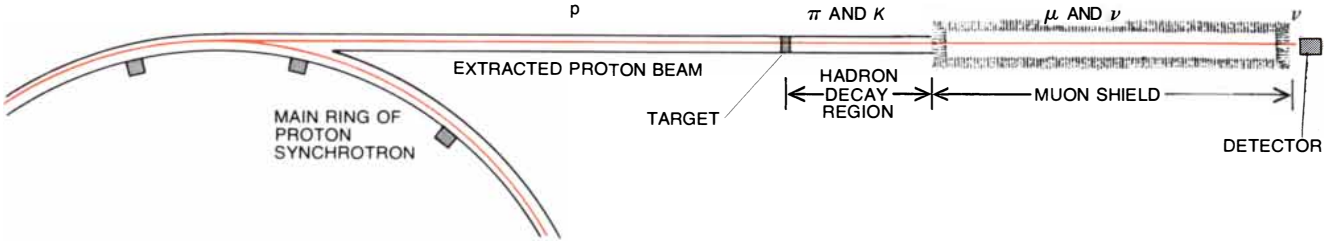
For many particles neither of these processes is possible; they must decay by means of the weak interaction. For example, consider the decay of the lambda particle, which was shown above to be a member of the family of baryons with a spin angular momentum of 1/2. The lambda most often decays into a proton and a negative pion or a neutron and a neutral pion; the two reactions are written $\Lambda \rightarrow p + \pi^-$ and $\Lambda \rightarrow n + \pi^0$. Since all the particles involved are hadrons, it might seem that the decay could proceed through the strong interaction. Mass, electric charge and spin angular momentum are all conserved. A strong-interaction decay is forbidden, however, by the requirement that the strong interaction conserve strangeness. The lambda has a strangeness of -1; the proton, neutron and pion all have zero strangeness. In both modes of decay strangeness is not conserved but changes by one unit. The decay must therefore take place through the weak interaction, and the time required is much greater than that needed for other decays: about 10^{-10} second.

There are many other hadrons that must decay by the weak force because some conservation law is violated in their disintegration. The charged kaons, for example, which have a strangeness of ± 1, can decay into a charged pion and a neutral pion or into a muon and a neutrino. Because strangeness is not conserved the decay can proceed only through the weak interaction, and it requires about 10^{-8} second.

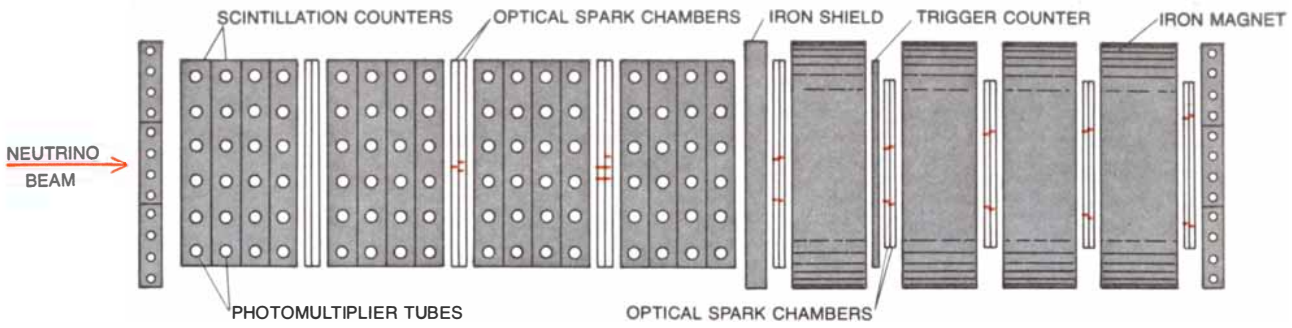
In a few cases a weak decay is mandatory even though all quantum numbers are conserved. Consider the decay of the neutron. It cannot yield a proton and a pion because the combined mass of the proton and pion is greater than that of the neutron. Similarly, it cannot decay electromagnetically into a proton and a photon because that reaction would violate the conservation of electric charge. The only allowed transition proceeds through the weak interaction and yields a proton, an electron and an antineutrino. It is written $n \rightarrow p + e^- + \bar{\nu}_e$. It is the process known as beta decay, a familiar form of radioactivity and an essential part of the nuclear chemistry of stars.

In the decay of the neutron no conservation laws are violated. Energy, charge and spin angular momentum, as well as strangeness, are all the same after the decay as they were before. Nevertheless, the time required for the decay is extraordinarily long: 10^3 seconds, or roughly 15 minutes. A very similar decay of the lambda baryon, $\Lambda \rightarrow p + e^- + \bar{\nu}_e$, is many times faster even though it violates the conservation of strangeness. Why is the neutron's decay inhibited?

The explanation embodies an important principle applicable to many other



BEAMS OF NEUTRINOS are generated indirectly from protons accelerated in the 400-GeV proton synchrotron at the Fermi National Accelerator Laboratory (Fermilab). The protons collide with a metal target, generating pions and kaons. These mesons then decay, mainly into muons, neutrinos and antineutrinos. The muons and any surviving hadrons are filtered out by passing the beam through a mound of earth about a kilometer long; when the beam leaves the mound, only neutrinos and antineutrinos remain.



DETECTOR employed at Fermilab in the search for new families of particles is sensitive to events with two muons in the final state (dimuon events). Energetic neutrinos interact with nucleons in tanks containing about 60 tons of a liquid that scintillates, or emits a flash of light, when an electrically charged particle passes through it; the scintillations are recorded by photomultiplier tubes. Beyond the scintillation counters is an iron shield that filters out hadrons and provides an additional target for neutrino interactions. The optical spark chambers record the paths of the particles, and the magnets deflect them so that their momentum can be measured. Particles observed in the final spark chambers are assumed to be muons because muons are the only charged particles that can pass through the iron shield and the iron magnets. The presence of two muons indicates that an unseen neutrino has also been created; at least one of the muons and the neutrino could be the products of some new particle that decays through the weak interaction.

weak decays. The mass splitting between the lambda and the nucleon states (generated by the semistrong interaction) is about 177 MeV; this represents the energy available to drive the decay process. On the other hand, the mass splitting between the neutron and the proton (generated by the electromagnetic interaction) is only about 1.3 MeV. It is the lack of energy that slows the decay of the neutron; indeed, it slows it far more than the failure to conserve strangeness slows the decay of the lambda particle.

This principle shows that the mass splittings within a family of particles are intimately related to the weak-decay rates of the particles. If the particle states are close together in mass, then there is little energy available for their decays and they can be expected to have relatively long lifetimes.

Suppose there is a new group of hadrons, in which the mass splittings are larger than those in the well-known families, such as the spin-1/2 baryons. The particles in such a group would necessarily bear a new quantum number; in fact, it is the new quantum number that would define the new group and distinguish its members from the familiar particles. The quantum number would be another internal one, perhaps closely analogous to strangeness.

Within the new group one particle might decay into another by the strong force, just as one strange particle can be transformed into another (of lower mass) by the strong force. Transitions to other states, however, such as decays to known baryons and mesons, could take place only through the weak force, because such transitions would fail to conserve the new quantum number.

Because of the larger mass splittings among the hypothetical new hadrons the weak transitions would not be inhibited by lack of energy. One can therefore predict that a distinctive characteristic of such new hadrons would be weak-interaction decay times considerably shorter than those of the better-known particles. The lifetimes would still be longer, however, than the 10^{-23} second characteristic of particles decaying by the strong interactions.

Decays of leptons necessarily proceed by the weak interaction. The electron and the neutrinos do not decay at all; decays of the muon, the only lepton that is capable of transitions to a lower-energy state, must conserve not only energy and electric charge but also the lepton quantum numbers, L_e and L_μ. As a consequence the muon invariably decays into an electron, an antineutrino of the electron type and a neutrino of the muon type. The decay process is written $\mu^- \rightarrow e^- + \bar{\nu}_e + \bar{\nu}_\mu$. The initial state has an L_μ quantum number of $+1$, and so does the muon neutrino in the final state; since the L_e quantum numbers of the electron and the antineutrino cancel, the lepton quantum numbers are conserved. The mass splitting between the muon and the electron is about 105 MeV, and the time required for the decay is about 10^{-6} second.

If heavier leptons exist, they will probably decay in a manner analogous to that of the muon, that is, by emitting a neutrino and an antineutrino. Once again, however, the larger mass splittings will accelerate the decay; the heavy leptons are expected to have much shorter lifetimes than the muon has.

The massive intermediate vector bosons should also decay by the weak interaction. A probable decay mode for the charged W particles is one in which they give rise to an electron or a muon and an antineutrino. For the negative vector boson, for example, the reactions would be $W^- \rightarrow e^- + \bar{\nu}_e$ and $W^- \rightarrow \mu^- + \bar{\nu}_\mu$. In both processes electric charge and the lepton quantum numbers L_e and L_μ are conserved. The mass splittings that separate the W particles from the leptons are enormous, and the weak decays should therefore be correspondingly rapid. The lifetime of the W particles is estimated to be about 10^{-18} second.

We have presented arguments for the existence of three kinds of undiscovered particle: a new group of hadrons, heavy leptons and massive intermediate vector bosons. All three are of theoretical interest. The discovery of a new group of hadrons, because it would introduce a new quantum number, might help to distinguish between various theories of the structure of hadrons. The discovery and theoretical explication of strange particles 25 years ago led to such a new synthesis. Because the leptons are a small family any addition would seem exciting. Moreover, there is evidence to suggest that the leptons are simple, pointlike particles, without internal structure, and at present there is no theoretical explanation of how mass splittings could develop among such particles, or of why two quantum numbers (L_e and L_μ) should be needed to describe just four of them. Knowledge of additional lepton states and their properties might provide an approach to answering these questions. Finally, the existence of W particles or heavy leptons or both is essential to the success of the unified theories of the weak and electromagnetic interactions. If the W particles can be

found, that will confirm the satisfying conjecture that each of the fundamental forces is transmitted by an intermediate boson whose mass is inversely proportional to the range of the force.

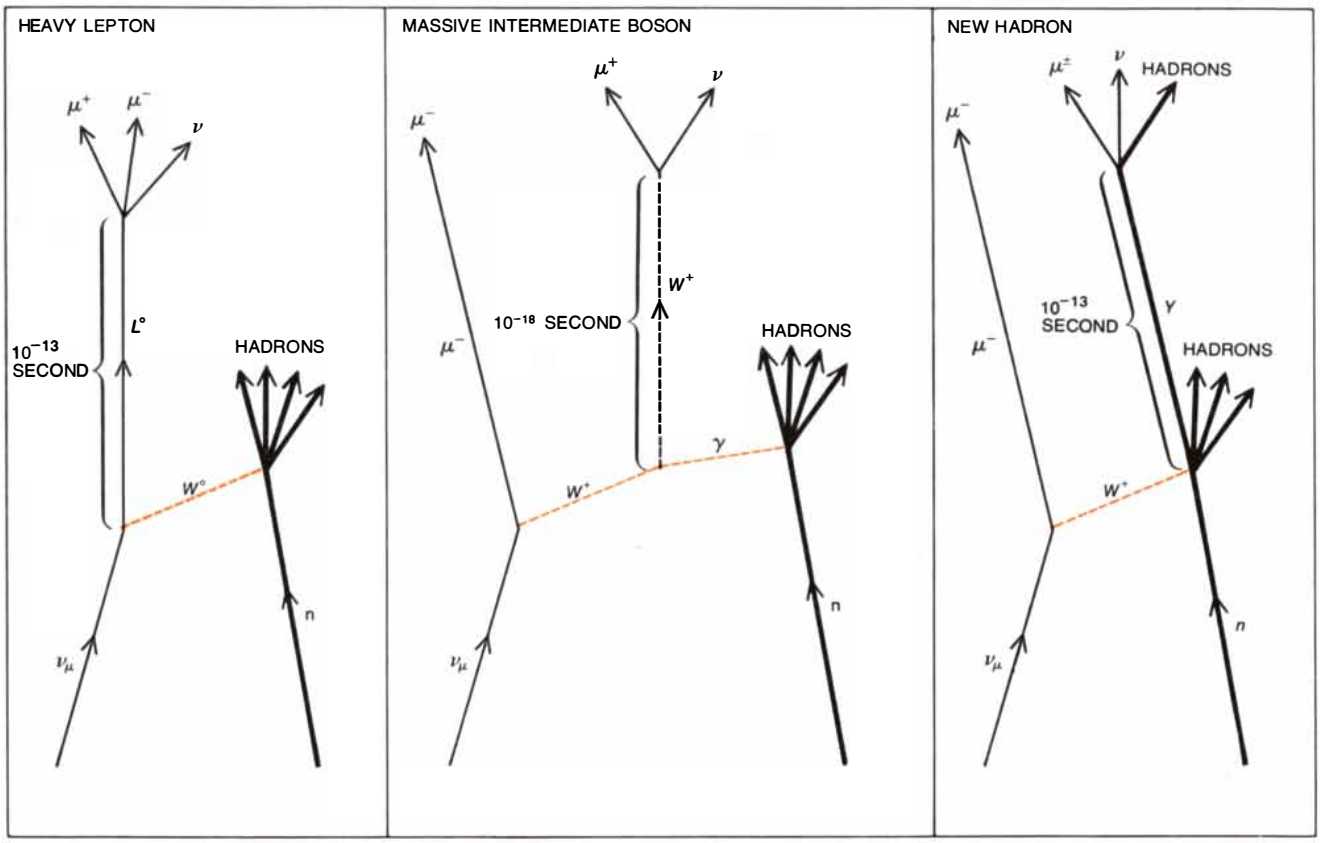
If it is assumed for the purpose of argument that all three new states exist in nature, why have they not been observed experimentally? One possibility is that they are simply too massive and the existing accelerators do not have enough energy to manufacture them. If the W particles are as massive as some theories predict, then they may in fact be beyond the reach of present machines, but it is much less likely that new hadrons and heavy leptons are that massive. High energy, however, is not the only limitation of experimental apparatus. It is necessary not only to create a particle but also to detect its presence, and lifetime is an important factor in detection.

A particle that is stable or relatively long-lived can readily be identified. If it is electrically charged, it will leave measurable tracks in a bubble chamber or a similar detector. Its velocity, momentum and energy can all be determined. If it is electrically neutral, it is more difficult to detect, but in general the products of its collisions will identify it.

These methods of detection are useful only for long-lived particles because they require the particle to traverse a macroscopic distance within the detector. Particles at high energy generally travel with velocities approaching the speed of light; if such a particle has a lifetime of 10^{-13} second, it will move less than a tenth of a millimeter before it decays. That distance is too small to be discerned in most detectors, which are therefore limited to particles with a lifetime longer than 10^{-13} second.

Particles that are very short-lived can also be detected, albeit by methods that are less direct and that provide less information about the particle. The method of detection employed is an application of Werner Heisenberg's uncertainty principle and it consists in searching for an enhancement at a particular energy in the probability of interaction between known particles. Such an enhancement is called a resonance. One expression of the uncertainty principle relates the uncertainty in the energy at which the resonance is found to the lifetime of the particle being created: the greater the uncertainty, the longer the lifetime. A great many hadrons that decay strongly in about 10^{-23} second have been detected by this method. The J or psi particles were also found as resonances. At the Stanford Linear Accelerator Center the least massive of them appeared as an enhancement at an energy of 3.1 GeV in the probability that electrons and positrons would annihilate one another. The uncertainty in the determination of that energy was several thousand electron volts, which led to a calculation of a lifetime of about 10^{-20} second [see "Electron-Positron Annihilation and the New Particles," by Sidney D. Drell; SCIENTIFIC AMERICAN, June, 1975]. The lifetime of the eta meson, about 10^{-19} second, was measured in a similar way.

In principle any particle could be detected as a resonance, but for those with lifetimes longer than about 10^{-18} second the method is impractical. For long-lived particles the theoretical uncertainty in



DIMUON EVENTS might result from three possible interactions of a muon-type neutrino with a neutron; in each the final state includes two muons and a neutrino. The interaction might create a neutral heavy lepton (L^0), a massive intermediate boson (W^+), one of the carriers of the weak force, or a hadron (Y) that bears a new quantum number. The characteristics of the dimuons detected at Fermilab suggest that a new hadron is the most likely interpretation. In these diagrams leptons are represented by light lines, hadrons by heavy lines and intermediate bosons by broken lines. Virtual particles, which can never be detected, are shown in color.

the resonance energy is very small, and it is overwhelmed by uncertainties introduced by the experimental apparatus.

There is a large gap in the spectrum of particle lifetimes that can be readily measured: it extends from about 10^{-11} second to about 10^{-18} second [*see illustration on page 45*]. The region is not completely barren; as we have seen, it contains the neutral pion, with a lifetime of about 10^{-16} second. With that exception, however, the region is unexplored. If it is to be searched systematically for new particles, other methods of detection will be needed.

New hadrons, heavy leptons and W particles are all expected to have lifetimes between 10^{-12} and 10^{-18} second. That prediction is based on the mass splittings between the particle states, which are expected to be larger than those of known particles. Conventional methods of detection are therefore not likely to be effective in searching for the new states.

The most promising approach to detecting the new particles is to examine the products of appropriate interactions for some distinctive mode of decay. As we have shown, new hadrons, heavy leptons and W particles must all decay by the weak interaction. The hadrons should make transitions to lighter, ordinary hadrons, with the emission of a muon or an electron and a neutrino. Both heavy leptons and the charged W particles can be expected to have a muon and a muon-type neutrino (with opposite L_μ quantum numbers) among their decay products. Thus it seems that an identifying signature of the decays of those particles is the presence of at least one neutrino in the final state.

The neutrinos cannot be detected directly; they would pass through any apparatus one might devise without interacting in any way. They nevertheless reveal their presence because of the rigorous conservation of the L_e and L_μ quantum numbers. The outgoing neutrinos escape detection, but as a result the remaining combination of particles exhibits an apparent violation of lepton conservation. In the confident belief that such a violation is impossible, the presence of the neutrinos is inferred.

Suppose a weak interaction is initiated by the collision of a muon-type neutrino with a neutron. If the reaction proceeds through a charged current (which is much more probable than the rare neutral weak current), conservation laws require that the final state include a muon and at least one baryon, such as a proton. The reaction would read: $\nu_\mu + n \rightarrow p + \mu^-$. Suppose, however, a W^+ particle were created in the interaction; electric charge could then be conserved without transforming the neutron into a proton and the reaction would be described as $\nu_\mu + n \rightarrow n + W^+ + \mu^-$. The W^+ could not be detected directly because it would disintegrate quickly (in about 10^{-18} second), but the products of its decay, a muon and an antineutrino, would make up part of the final state. Thus the overall reaction would appear to be $\nu_\mu + n \rightarrow n + \mu^+ + \mu^- + \nu_\mu$, where the muon-type neutrino and the positive muon in the final state are the decay products of the W^+.

The outgoing neutrino in the above reaction would inevitably escape observation, and it would also be difficult to detect the outgoing neutron; the final state would therefore appear to consist of a positive muon and a negative muon, both of which could readily be detected. A simple calculation shows that the lepton number L_μ of the particles detected in the final state is $-1 + 1$, or 0, whereas the L_μ of the initial state was $+1$. The apparent violation of L_μ conservation reveals that a neutrino was also present in the interaction, and it serves as a signal that some new particle has been created.

If the new particle had been a neutral heavy lepton instead of a W^+, the final state would have been quite similar. If it had been a hadron belonging to some new group, the final state would also have included two muons, although not necessarily muons with opposite electric charge. In each case the existence of a missing neutrino would be indicated by an apparent violation of lepton conservation. In order to distinguish between the three hypotheses it would be necessary to examine more closely the characteristics of the two muons.

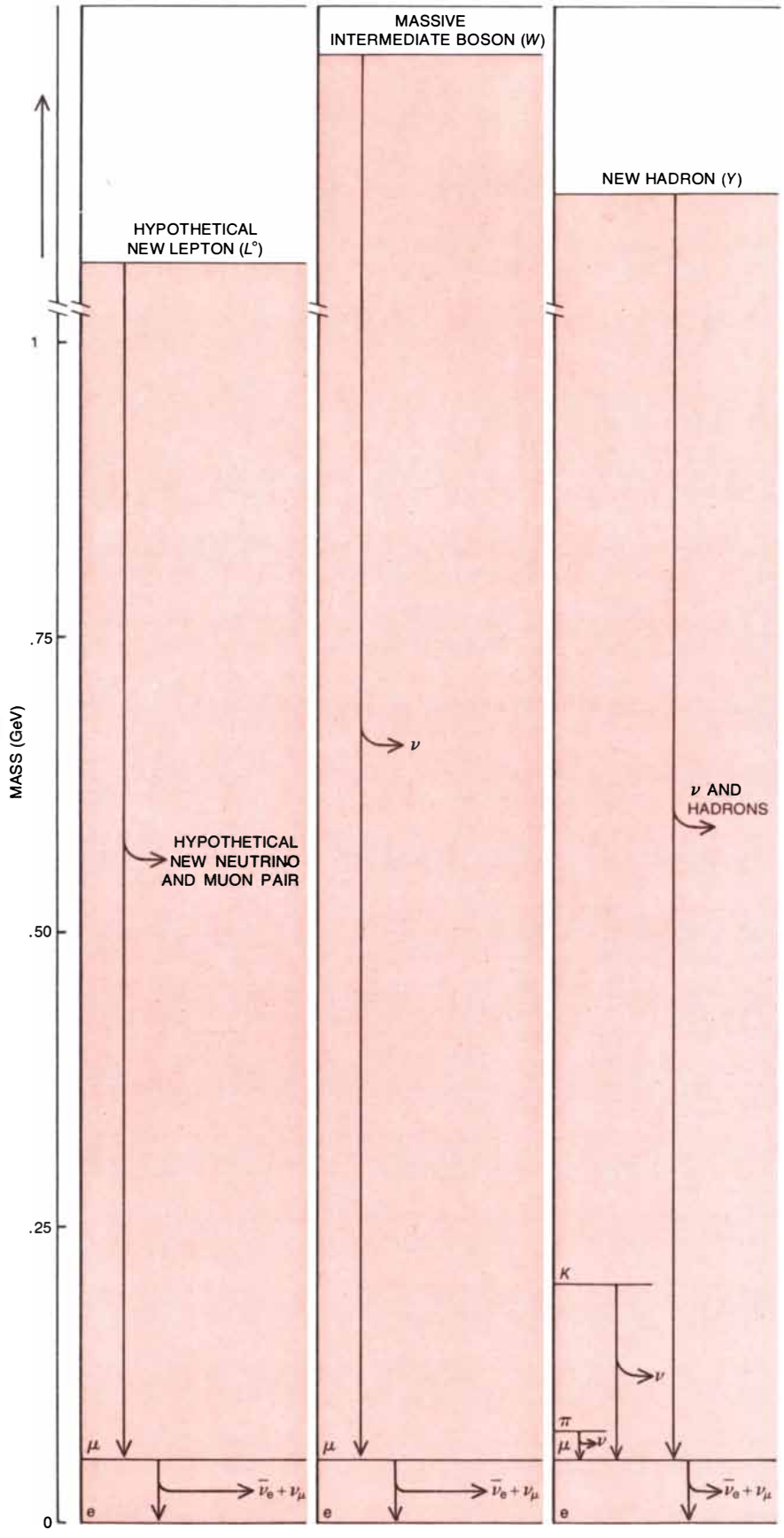
Such interactions have come to be known as dimuon events. At Fermilab we have conducted a careful search for them in collaboration with 10 other physicists from Harvard University, the University of Pennsylvania, the University of Wisconsin and Fermilab. In the past two years we have succeeded in observing events with the expected characteristics, and the observation of dimuon events has subsequently been confirmed by another group of physicists from the California Institute of Technology and Fermilab. The first dimuon event was observed in April, 1973, shortly after the Fermilab accelerator began operating.

In principle W particles or heavy leptons could be generated in many kinds of interactions, including those of one hadron with another. Members of new hadron families could also be made in collisions of ordinary hadrons, although only as pairs of particles and antiparticles because of the strict conservation laws that govern strong interactions. High-energy hadron collisions are complex events, however, in which many extraneous processes would compete with those we wanted to observe. We therefore adopted a different technique: that of bombarding nucleons with high-energy neutrinos. In such events only weak interactions are possible, and the results are comparatively easy to interpret. Moreover, new hadrons can be produced singly, since the weak interactions can alter the internal quantum numbers of hadrons.

The first requirement of this technique is a beam of high-energy neutrinos. At Fermilab they are manufactured indirectly by protons extracted from the laboratory's main accelerator, a 400-GeV proton synchrotron. The protons collide with a metal target, generating a beam of secondary hadrons made up primarily of pions and kaons. These mesons decay as they pass through an evacuated pipe some 300 meters long. As we have seen, both kinds of meson decay predominantly into muons, neutrinos and antineutrinos. At the end of the decay region the muons and any remaining hadrons are filtered out by a mound of earth about a kilometer long. The beam emerging from the mound consists almost exclusively of neutrinos and antineutrinos.

The neutrinos must have high energy in order to create particles of large mass. It is equally important that the beam have high intensity, that is, that it consist of a great many particles per second. High intensity is essential because the probability is quite small that any given neutrino will interact with a nucleon. The beam at Fermilab has an intensity of about 10^9 neutrinos per second; even so only about one neutrino per minute interacts. All the rest pass through the apparatus unaltered.

The detector employed in the dimuon experiments consisted of several components arranged in sequence along the axis of the incident neutrino beam. The first section was made up of four chambers containing about 60 tons of a liquid that scintillates, or emits a flash of light, when an electrically charged particle passes through it. The large mass of liquid provided an effective target of nucleons for the neutrinos, and it also



DECAYS OF NEW PARTICLES would all lead to the emission of a muon and a neutrino. For the *Y* hadrons this mode of decay would be analogous to the decays of the pion and the kaon, but the *Y* particles would decay much faster. The mass of the lightest *Y* particle is estimated to be equivalent to an energy between 2 and 4 GeV (billion electron volts); *W* bosons must have a mass of at least 20 GeV; masses of new leptons are uncertain.

absorbed most of the energy of the interactions. In each tank the scintillations were detected by photomultiplier tubes; the total quantity of light emitted was a measure of the energy deposited in the liquid. Between the containers of liquid were optical spark chambers that provided a visual display of the cascade of particles created by each interaction.

Immediately beyond the liquid scintillator was a large mass of iron (also about 60 tons). The iron served as an additional target for neutrino interactions, and furthermore it filtered out some of the extraneous particles created in the scintillation counter. Beyond the iron shield were four cylindrical iron electromagnets, each 12 feet in diameter and four feet long. The magnets deflected charged particles passing through them so that the momentum of the particles could be measured. The paths of the particles were recorded by additional spark chambers between the magnets.

The detector was activated on the arrival of a charged particle at a triggering counter immediately beyond the first of the electromagnets. Such a particle would necessarily be a muon; no electron or hadron could pass through the iron target and the iron magnet. On activation of the detector the electrical output of all the scintillation counters was recorded and a high voltage was applied to the spark chambers. The resulting tracks in the spark chambers, which followed the ionized paths where charged particles had passed, were recorded photographically. We searched for dimuon production simply by checking the photographs of each recorded event for the presence of two charged particles that had penetrated the iron barriers.

In several series of experiments, employing beams with various mixtures of neutrinos and antineutrinos, about 80,000 activations of the detector were recorded. A total of 82 activations met the criteria we had established for identifying authentic dimuon events. For an event to qualify, the first requirement, of course, was that it have two muons in the final state. It was also necessary that the muons be generated at the same moment and at the same point in space, and that the momentum of both of them be measurable.

Several explanations that involve only conventional processes and particles might be invoked to account for the dimuon events. One of the most obvious is the hypothesis that the second muon is created by the decay of a charged

pion or kaon produced in the initial neutrino interaction. Because of the different densities of the two targets, however, muons from meson decays would be expected to originate in the liquid scintillator about four times more often than in the iron barrier. Actually the same number of dimuons was observed in the iron as in the liquid, and that ratio excludes the pion and kaon mechanisms.

It is also possible that the apparent lack of conservation of the lepton quantum number L_μ is due not to a missing neutrino but to a third muon that escapes detection. The spark-chamber photographs were therefore searched for events with three muons in the final state. No such events have been found.

We believe that all the conventional explanations can be refuted or assigned a very low probability. As a result there remain three plausible interpretations of the dimuon events. They must represent the production and decay of one of the particles discussed above: a new hadron, a neutral heavy lepton or a charged intermediate vector boson.

In order to discriminate between the three possibilities it is necessary to consider the properties of the muon pairs. Two properties in particular are of great importance. First, our measurements have shown that the two muons generally do not have comparable values of momentum. Instead, for events induced by neutrinos the negative muon usually has larger momentum than the positive one. Second, the muons do not always have opposite electric charge; events in which the charges are the same make up about 10 percent of the total.

Neither of these observations can be reconciled with the hypothesis that the dimuons are generated by a heavy lepton or a W particle. The production of muons with like charges in particular excludes those mechanisms from consideration. The overall interaction in the detector can be imagined to take place in two stages. In the first the neutrino is converted into other particles; in the second the recoiling nucleon decays into other hadrons. The observation of two muons with the same charge cannot then be explained by the decay of any particle produced in the first stage; hadrons in the second stage must also be involved. This conclusion has been reinforced by comparing the experimental data with theoretical calculations made by Abraham Pais of Rockefeller University and Sam B. Treiman of Princeton University.

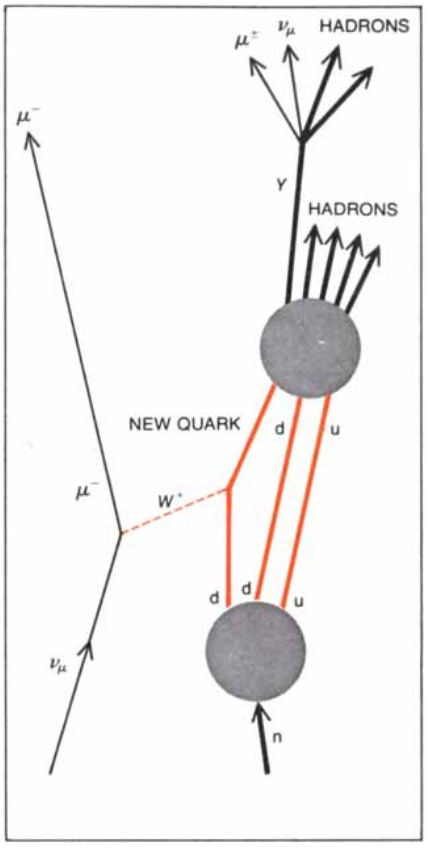
From these arguments we have concluded that the most likely explanation of the dimuon events observed at Fermilab is the creation and decay of at least one member of a new family of hadrons. We have designated the members of this new family Y particles. The Y particles must necessarily possess a new quantum number conserved in the strong and electromagnetic interactions but not in the weak interaction.

From other properties of the dimuons we estimate the mass of the Y particles to be between 2 and 4 GeV. Their lifetime is appreciably less than 10^{-8} second, and it is probably less than 10^{-10} second. It appears that their decay products include leptons only from 10 to 20 percent of the time; the remainder of their decays yield hadrons only. This suggests that the Y particles are produced much more copiously in high-energy neutrino events than our results might at first seem to indicate.

Further experimental investigations will be needed before the nature and significance of the Y particles are clear. We have assumed that there are both Y mesons and Y baryons, but that has not been definitely established. The masses and lifetimes of the particles must be measured accurately, and their modes of decay must be described. In spite of these uncertainties the Y particles represent the most direct and substantial evidence we now have for the existence of hadrons with a new quantum number. The Y particles apparently exhibit this quantum number directly, whereas the J or psi particles reflect the existence of a new quantum number only indirectly. It is not yet known whether the same quantum number distinguishes both states.

Both the Y and the J or psi particles can be interpreted in terms of the sophisticated theoretical model of hadrons introduced in 1963 by Murray Gell-Mann and George Zweig of the California Institute of Technology. It is in this model that all hadrons are regarded as being composite structures made up of simpler constituents called quarks. Baryons consist of three quarks, and mesons consist of a quark and an antiquark. In the original formulation of the model there were three kinds of quark, labeled u, d and s. The s quark carries one unit of strangeness, and all strange particles incorporate at least one s quark or $\bar{s}$ antiquark.

In the quark model the mass splittings among the hadrons are merely a reflection of mass splittings among the quarks. It is because the s quark is considerably heavier than the other quarks that the strange hadrons are heavier than their nonstrange counterparts.



QUARK HYPOTHESIS might explain the creation of new hadrons in neutrino interactions. In the quark theory all baryons are made up of three quarks; the neutron consists of quarks labeled *d*, *d* and *u*. Through a virtual W^+ boson the neutrino converts one of the *d* quarks into a quark of a new kind, which bears a new quantum number. The quarks then interact to generate a spectrum of hadrons, including at least one (*Y*) that bears the new quantum number. Some theorists identify the quantum number with the hypothetical property designated charm.

More recently it has been suggested that there may be a fourth quark, carrying a new quantum number conserved in the strong and electromagnetic interactions. Sheldon Lee Glashow of Harvard and James D. Bjorken of the Stanford Linear Accelerator Center proposed that the new quantum number be called charm, and the fourth quark is now labeled the c, or charmed, quark [see "Quarks with Color and Flavor," by Sheldon Lee Glashow; SCIENTIFIC AMERICAN, October, 1975]. If the charmed quark exists, it implies the existence of a new class of hadrons, which could decay to ordinary hadrons only through the agency of the weak interaction. Moreover, because the charmed quark would be more massive even than the s quark, the mass splittings of the

new hadrons would be larger than those observed among the familiar ones.

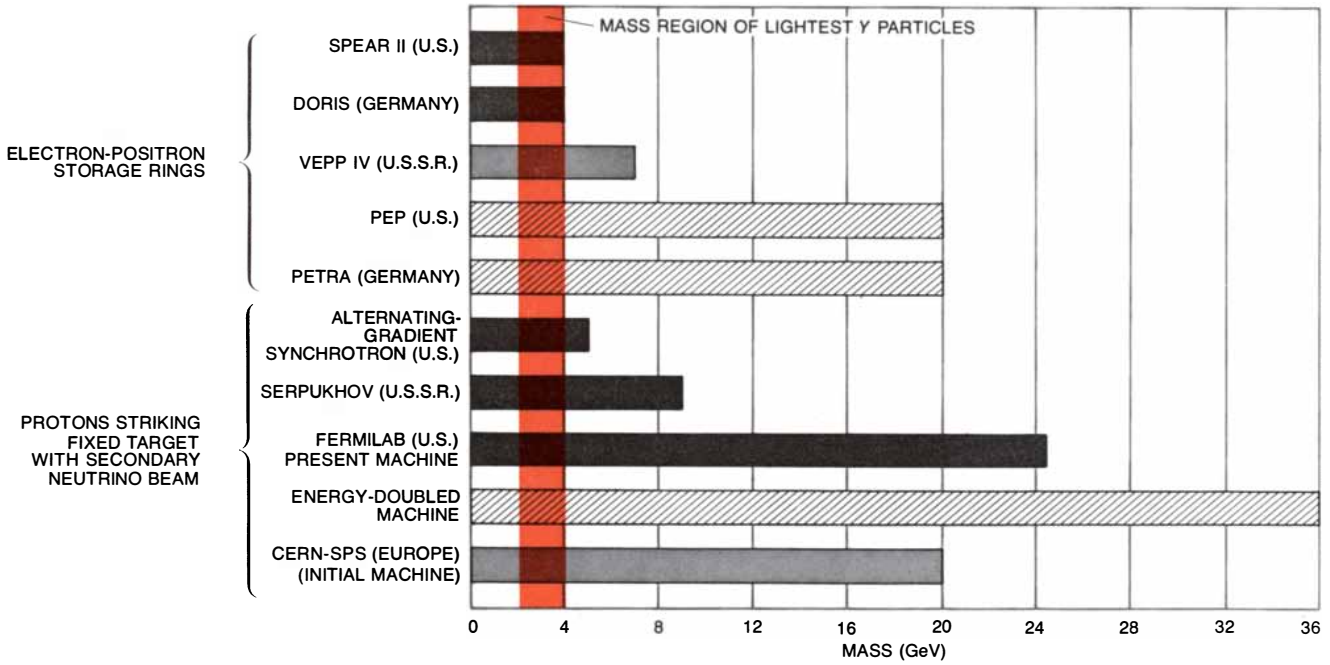
If the Y particles, with their new quantum number, are to be incorporated into the quark model, they require at least one additional quark. It seems reasonable to identify that fourth quark with the charmed one proposed by Glashow and Bjorken; this identification can be proved, however, only through further experiments. The charm hypothesis is supported by strong theoretical arguments independent of the recent discoveries; moreover, it can explain the J or psi states as well as the Y. The Y states are assumed to be composed of c quarks in combination with other kinds, whereas the J or psi is a bound state of a c quark and a $\bar{c}$ antiquark. It thus appears that theory and experiment may again be converging, as they did 25 years ago, on a new family of hadrons.

It seems ironic that the observation of dimuon events induced by neutrinos and antineutrinos should lead to the discovery of a new quantum number of hadrons. For many years dimuon events were sought as a signal that massive intermediate vector bosons had been produced. Now that that signal has been detected it has a meaning completely different from the one we had expected.

What are the prospects for the discovery of other new particles? One conclusion that now seems justified is that the mass of the W particles is much larger than that of the new hadrons; we have established experimentally that the mass of the charged W bosons must be greater than 20 GeV. The mass and other properties of heavy leptons are more difficult to predict. They might be of various types, some of which would be more difficult to create and detect than others. For example, there might be heavy leptons with a mass less than that of the Y hadrons. The Y particles would then decay through these leptons to ordinary leptons and a new kind of neutrino; the ultimate result would be observed as a dimuon event. Thus both new particles would have been created in a single experiment. At present the data are too meager to exclude this possibility.

One naïve inference that could be made from the existence of Y mesons with a mass between 2 and 4 GeV is that the meson states of lowest mass in each major group of particles have a simple distribution: their masses may be spaced more or less regularly. There seems to be a simple relation between the mass of the pi meson, that of the K meson and that of the Y meson of lowest mass; if the relation is confirmed, it suggests that there may be another new meson, bearing yet another quantum number, in the mass region between 8 and 15 GeV. The search for such a meson is likely to become a preoccupation of experimental physicists, particularly those who experiment with high-energy neutrinos, in the coming years.

The endeavor to find additional meson states and other particles will be facilitated by several new accelerators and other machines, including some that have recently begun operating and others that are still under construction [*see illustration below*]. At Fermilab the main proton synchrotron is being modified by the addition of superconducting magnets that will double its energy. With new machines of this kind it will be possible to explore regions of still higher mass, and with indirect methods of detecting particles it should be possible to determine the membership and the properties of the new families. In spite of the perplexing multiplication of particles and quantum numbers there is hope that the picture of nature emerging from these studies will be a simple and elegant one.



TEN PARTICLE ACCELERATORS capable of exploring the mass region of the Y particles will be operating in the next decade. Five of them are in operation now (*black*); two more are under construction (*gray*); the remaining three are being planned (*hatched bars*). In electron-positron storage rings particles collide with equal and opposite velocity; in the remaining accelerators neutrinos collide with nucleons in a fixed target. The masses given represent not the energy to which a particle can be accelerated but rather the maximum mass of a particle that can be created through the interactions of the accelerated particles. At Fermilab the energy of the proton synchrotron will be doubled through the installation of superconducting magnets, increasing the energy of neutrino interactions 50 percent. Some accelerators should be capable of making heavy leptons and W bosons as well as new families of hadrons.

Plain wrapping available if purchaser prefers not to advertise intentions.

Typical contents: KODAK EKTAPRINT R-500 Processing Kit; 10-sheet package of KODAK EKTACHROME RC Paper, Type 1993; Kodak Publication No. Z-205; KODEWORD Darkroom Audio Cassette TC-3.

Typical outlay for said items: About the same as for a pair of shoes.

Typical objective: Weekend in which to experience the pleasure of moving from black-and-white printing into making 8″ x 10″ color prints from a few favorite color slides with one of those tube processors now on the market and step-by-step, real-time guidance by a friendly voice, aural signals, and music that only secondarily brings cheer to the darkroom but primarily keeps check on proper timing.*

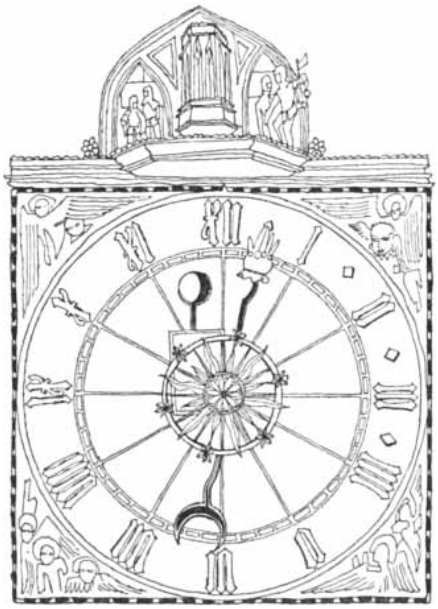
(Possible professional objective: An upcoming "poster session" requiring attractive and pertinent visuals to stop interested parties and stimulate discussion. See "The new way to 'give a paper'," *Science* for Nov. 7, 1975, p. 502, or *Scientific American* for November '75, p. 55.)

Alternate selections: Other paper and chemicals in trial-size quantities, other literature and KODEWORD™ Cassettes—all to serve a resurgence of interest in the darkroom among the lively-minded. Convenience, economy, and confidence offered to those wishing to process their own color slides, process their own color negatives, process their own color prints from color negatives, process color prints from slides by means other than a tube processor—or even just do good black-and-white work. Photo dealers have full details.

Kodak

*See p. 10, *Scientific American,* Sept. 1975, on Bach fugues in the darkroom and how they helped make color photography practical.

SCIENCE AND THE CITIZEN



Fire in the Ashes

Over the next 20 years U.S. industry will be expected to invest some $60 billion in uranium-enrichment facilities and roughly a tenth of that amount, $6 billion, in facilities for recovering plutonium and unburned uranium 235 from the ashes of nuclear power stations. The first process, although technically difficult, seems by far the more attractive investment, partly because it presents virtually no special safety problems and yields a product that offers no temptation to a potential atomic-bomb maker. The second process, in contrast, would seem to offer the fewest commercial incentives imaginable. Spent fuels typically contain 1 percent or less of plutonium and the same amount of uranium 235; to separate them in ultrahigh purity from scores of fission products calls for the most complex chemistry ever attempted on an industrial scale. Moreover, the spent fuel is so intensely radioactive that all processing must be conducted by remote control behind thick walls. Even the equipment must be designed for remote maintenance and remote replacement. One of the two principal products, purified plutonium, must be handled with extreme caution not only because it is a potential bomb material but also because its intense radioactivity makes the accidental inhalation of particles of it extremely hazardous. Finally, the ultimate wastes must be encapsulated in such a way that they can safely be stored for thousands of years.

It is hardly surprising that private companies have not been clamoring to enter the fuel-reprocessing business. A plant of economic size, one capable of handling perhaps 1,500 metric tons of wastes a year, costs upward of $500 million. In the years since the original plutonium-separation facilities at Hanford, Wash., were shut down, the Government has obtained its military plutonium from the large Savannah River plant in Georgia. The feedstock is natural uranium that has been irradiated by neutrons in nonpower reactors. A much smaller Government plant at Arco, Idaho, reprocesses spent fuel from nuclear submarines. The only privately financed reprocessing plant that has ever been operated in the U.S. is one set up by Nuclear Fuel Services, Inc., in West Valley, N.Y. Designed to process 300 metric tons of spent fuel per year, the plant reprocessed 630 metric tons of fuel from nuclear power stations between 1966 and 1972. Since then it has been shut down pending permission from the U.S. Nuclear Regulatory Commission to expand its capacity to 750 tons per year. Originally financed by W. R. Grace & Co. and AMF Incorporated, the West Valley plant is now a wholly owned subsidiary of the Getty Oil Co.

A reprocessing plant with a capacity of 300 metric tons of fuel per year built by the General Electric Company at Morris, Ill., proved seriously defective during trial runs with cold (nonradioactive) feed material. As a result GE has withdrawn its request for an operating license and is considering whether or not to rebuild the plant. In Barnwell, S.C., Allied General Nuclear Services is nearing completion of a $500-million reprocessing plant with a capacity of 1,500 tons per year. The company is a subsidiary of General Atomic and Allied Chemical Corporation. Exxon Corporation is contemplating a plant with a capacity of 1,500 tons per year but has not yet applied for a construction license. If Exxon goes ahead with its plans and all licenses are granted, the U.S. may have four private reprocessing plants with a total capacity of 4,000 metric tons per year by 1985.

Meanwhile spent fuel from the nation's 56 nuclear power stations has been piling up in repository sites. The total backlog is now about 2,000 tons (containing enough plutonium for about 3,000 Nagasaki-size bombs), and it will grow by another 500 to 1,000 tons before any of the plants is ready for operation. By 1985, when 200 or more nuclear power plants are expected to be operating in the U.S., the annual discharge of spent fuel will be about 3,500 tons per year and will be increasing at the rate of about 20 percent annually.

The basic decision on whether or not to allow the use of nuclear fuel containing a mixture of uranium 235 and recycled plutonium is still to be made by the Nuclear Regulatory Commission. The commission has recently solicited statements from groups and organizations that may be opposed to the widespread use of plutonium because it is a health hazard or for other reasons. If the commission should be persuaded to prohibit the use of mixed fuel, which seems highly unlikely, the plutonium in spent fuel (far exceeding any conceivable military need) would have no commercial market, thus destroying the economic base of a still unborn fuel-reprocessing industry and significantly increasing the cost of nuclear power.

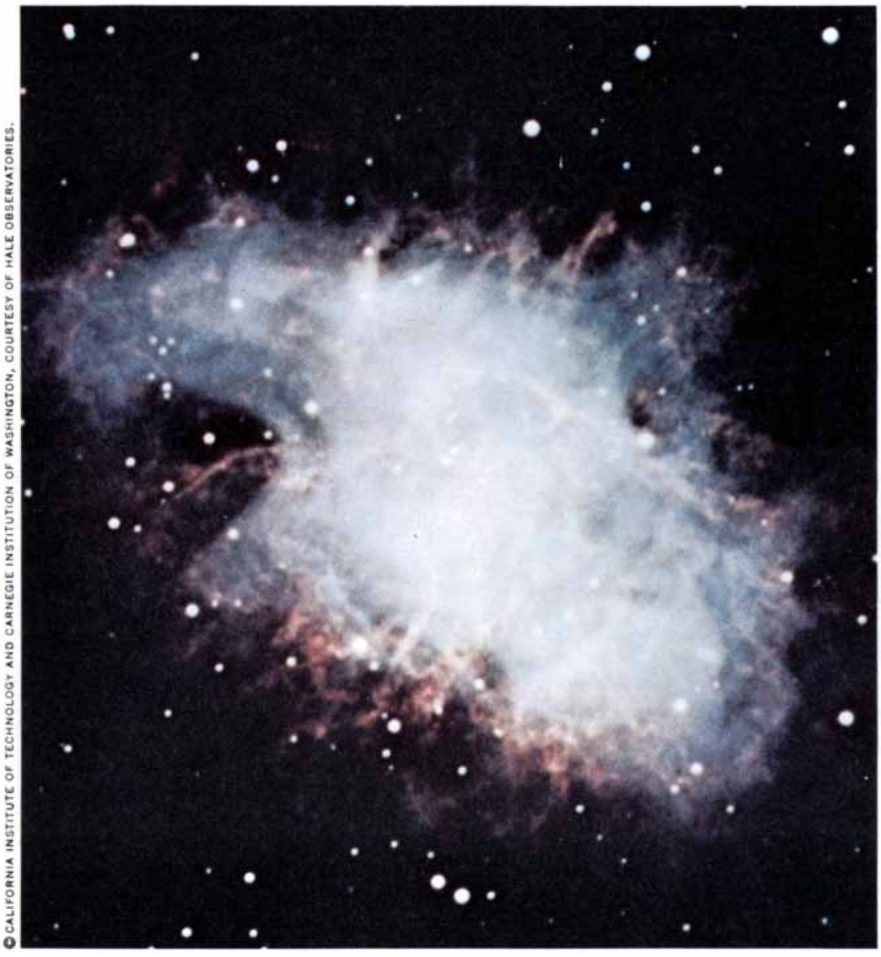
Inhumane and Indiscriminate

The effort to restrict or prohibit the use of incendiary weapons by international agreement—an aspect of arms control and disarmament largely neglected since the 1930's—is currently being taken up with a renewed sense of urgency, spurred in part by recent advances in the technology of fighting with fire. Over the past few years the General Assembly of the United Nations has approved by overwhelming majorities a series of resolutions condemning the use of napalm and other incendiary weapons in specific military situations. Last May a group of 18 countries placed a draft of a comprehensive ban on incendiary weapons before the International Committee of the Red Cross, which has been assigned the task of "modernizing" the Geneva Conventions of 1949. In pursuance of its mandate the Red Cross has convened several conferences of experts "to examine in more detail the question of particular conventional weapons which may cause unnecessary suffering or have indiscriminate effects." Incendiary weapons are expected to be high on the agenda of the next such conference, scheduled to begin soon at Lugano in Switzerland.

Until now the debate over what are

DP Science Dialogue

Notes and observations from IBM that may prove of interest to the scientific and engineering communities



The Crab Nebula is the gaseous remnant of a supernova first sighted by Chinese astronomers in 1054. Scientists believe it is shining today as a result of a pulsar into which the supernova evolved.

A Supernova is Born

One of the most spectacular stellar events occurs during the first 40 days after a star explodes. Called supernovae, these giant dying stars can attain a luminosity greater than the combined brightness of all the stars in an entire galaxy. This incredibly intense light eventually evolves into a "tail" of weaker light, lasting for a year or longer.

Although there is general agreement on many aspects of the life cycle of certain kinds of stars, explanations of supernovae and their causes are still the subject of intense scientific debate. Over the past two years, a group of IBM scientists led by Dr. Gordon Lasher has developed a computer model that describes how Type I supernovae behave during the first 40 days following the explosion. (Supernovae other than Type I do not follow predictable patterns.)

"The model gives us a start towards a quantitative understanding of what a star is like at the end of its life," says Dr. Lasher. "It is one more piece of the puzzle of how the universe evolves."

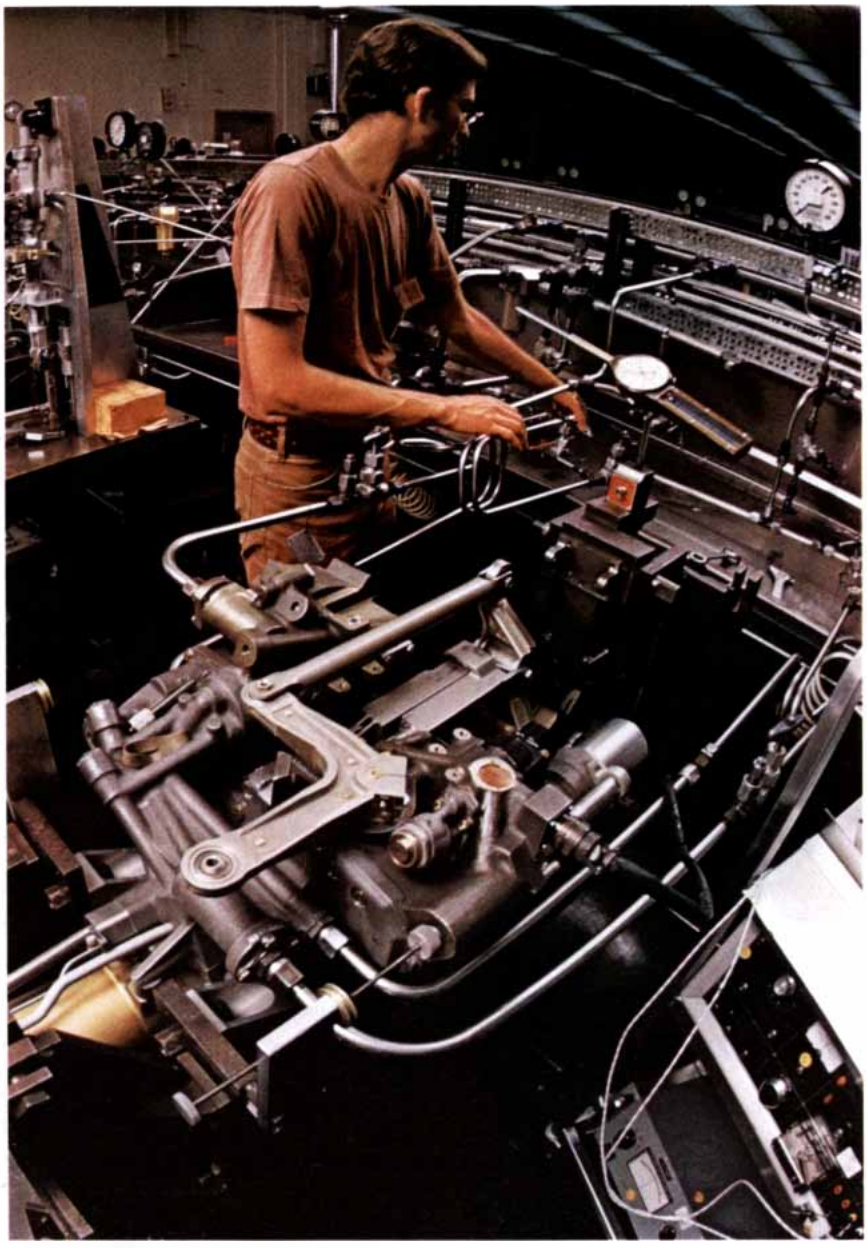
Developed at IBM's Thomas J. Watson Research Center in Yorktown Heights, New York, the model is based on the assumption that the rapid rise in radiant energy emitted from supernovae is the direct result of the collapse of the star's core. A small fraction of the energy released in that process is believed to cause an explosion and subsequent expansion of the low-density helium envelope surrounding the super-dense core. The rapidly thinning envelope eventually permits radiant energy to emerge in the form of intense visible light.

All stars pass through several observed stages—from their initial formation by the contraction of interstellar gas and dust to their "main sequence" phase, during which gases inside the star's core reach a temperature of several million degrees Kelvin—hot enough to drive a nuclear reaction in which hydrogen is converted to helium. Our sun is estimated to be four billion years into its "main sequence" lifetime, and is expected to last another five billion years.

In stars with masses larger than our sun, the core heat continues to intensify until all the helium in the core is converted to increasingly heavier elements, eventually becoming iron. A gaseous helium envelope surrounds the core and the "main sequence" is much shorter—maybe 10 million years.

"In these extremely heavy stars, the nuclear fuel is totally depleted and the core mass is so dense that the electrons

(Continued on 4th page)



At Bertea Corporation, an inboard elevator actuator package, designed with computer help, undergoes performance tests.

Designing Safety into Hydraulic Flight Control Systems

The hydraulic flight control systems produced by the Bertea Corporation of Irvine, California, can literally determine whether an aircraft ever gets off the ground—and how safely it will fly once it does. As a result, every component must be tested and re-tested to insure its ability to withstand the extreme stresses, pressures, voltages and temperatures that may occur during flight.

"We really can't afford to wait until a key component, like a jet stabilizer servo control package, is already in production to begin measuring its performance," says Bill Henry, director of engineering. "We must be able to do simulations of those same tests while the system is still in the design phase."

At Bertea, those early simulations are now conducted at computer terminals. The resulting graphs enable engineers to determine whether any design modifications are needed.

The stabilizer control arrangement is just one of many complex servo packages designed and manufactured by the company. The servos control the movements of all the ailerons, rudders and elevators of commercial and military aircraft.

"All of the control packages we design function in a dynamic environment. That's why we rely so heavily on interactive computing," adds Henry.

Interactive computing enables engineers to work directly with the company's System/370 Model 135 computer via four IBM terminals. The terminals run under the IBM Conversational Monitoring System (CMS), a time-sharing system created specifically for intermediate-size computers. With CMS, engineers can solve complex stress problems, simulate real flight conditions, or access a library of completed designs.

"Interactive computing has helped us solve one of our most difficult design problems—determining servo stability," says Bruce Borden, supervisor of analytical services. "For example, the flight control surfaces operate under conditions that vary tremendously, such as loads, pressures and temperatures."

Each of these conditions is described by a series of complex nonlinear equations that are analyzed interactively using an IBM program—Continuous System Modeling Program (CSMP).

"CSMP, along with other programs we have written, allows us to simulate exactly how the unit will react at every second over time. In effect, it enables our digital computer to behave as an analog model of the servo would," explains Borden. "With it, we can generate a graphic plot at the terminal indicating whether the device performed within prescribed performance parameters. We can vary any factor, or combinations of factors, and get a new descriptive plot printed at the terminal within minutes."

Interactive computing is also used to analyze what Bill Henry calls the "hole-distance" problem. As he explains, hydraulic servos are hollow metal manifolds with hundreds of holes drilled through them to allow the passage of hydraulic fluid. The pressure and direction of the fluid trigger different responses in the servo mechanism. "We use interactive programs to check the spatial relationships of all the holes, taking into account drill drift and critical intersection points," says Henry.

A strong advocate of interactive computing, Dr. Robert Langford, Bertea's president, comments: "We regard the computer as an indispensable tool that helps us explore all our options—material availability, costs and manufacturing time—long before we ever go into production. For these reasons, we expect to rely even more heavily on the computer as a planning aid in the future."

A New Way to Solve the Double Curve Problem

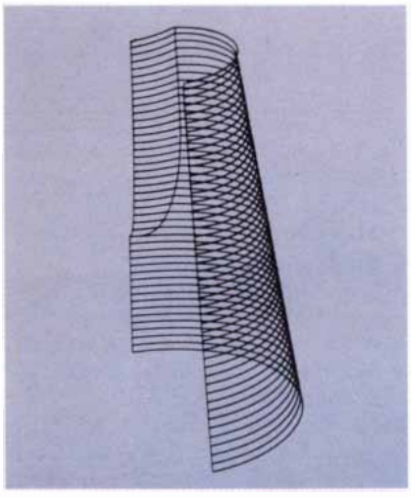
"Defining the geometry of curved surfaces is one of the toughest, most time-consuming problems we face," says Jim Mayfield of Teledyne Ryan Aeronautical. But Mayfield's team of engineers and mathematicians, working closely with the IBM Scientific Center in Los Angeles, has developed a way to solve that problem in minutes instead of weeks.

And Frank Oldfield, director of aerospace engineering at the San Diego-based manufacturer of pilotless aircraft, talks of another major benefit: "With the faster response time the system provides, we can test a variety of designs and select the best one."

Doubly curved surfaces, like the wings or air ducts of an aircraft, are particularly difficult to design because they cannot be easily described mathematically. A straight edge or even a simple cylindrical shape can be defined by several pairs of equations. But it takes hundreds of calculations to determine the final shape of a wing, which tapers gradually from root to tip. And even after the wing design is complete, figuring out basic data like surface area and volume requires a formidable amount of additional work.

Now, Teledyne Ryan is using the new Numerical Geometry System to aid in the creation, flow and use of geometric information. The system operates from terminals linked by IBM's Time Sharing Option to a System/370 Model 155 computer.

Data that defines design criteria is entered directly at a terminal. By keying in any one of more than 30 mathematical commands built into the system, the engineer can automatically manipulate the data to generate a wide range of basic geometric information like planes, scalers, areas, volumes and intersection points. The system also provides a permanent data base of information that can be referenced later by other departments, including the Numerical Control group.



The graphic representation of an air inlet duct was designed on the computer-based Numerical Geometry System.

The inlet duct graphically represented (below, left) is used in the Model 274 subsonic drone aircraft produced at Teledyne Ryan.

"Every aircraft is a compromise among a variety of disciplines, such as aerodynamics, thermodynamics and structural design," explains Mayfield. "Just to give one example, a wing must provide enough lift to get the plane off the ground, be strong enough to carry the necessary loads with the least structural weight, yet simultaneously be designed to minimize the amount of drag it produces."

Coordinating these often competing constraints makes designing the curved outer skin of the aircraft a highly iterative process, involving many departments. The initial blueprints, for example, originate in the advanced engineering group and must be translated from two-dimensional drawings to three-dimensional surfaces by the Loft-Lines group.

"Before we developed the Numerical Geometry System, there was little analytic definition in the design phase," says Mayfield. "The Loft-Linesman literally had to visualize the three-dimensional part, attempting to create a trial surface to match the geometric constraints. After being reviewed by the original designer and many other departments, the surface, along with desired modifications, would be passed back to the Loft-Lines group. This would occur over and over until an acceptable compromise could be reached. It was an extremely time-consuming process."

The Numerical Geometry System has not entirely replaced the Loft-Linesman's unique skill. But it provides the mathematical tools for creating a surface directly from given constraints—a capability that yields numerous advantages—like designing an inlet duct in hours instead of weeks.

And according to vice president and controller Roy Fields, "We have already realized a significant return on our investment in the system in the Loft-Lines area alone. And we feel it will yield even greater benefits in the future when it is extended to the aerodynamics, thermodynamics and stress analysis groups."

Computer Helps Farmers Fight the Blight

Figuring out the best time to spray potato crops to prevent blight used to be largely a matter of guesswork. But for the past four years, more than 100 farmers from Maine to Florida have been getting assistance in their battle against the lethal, blight-producing Phytophthora infestans fungus from an unusual source—a computer program known as "Blitecast".

Developed and administered by plant pathologists at the Pennsylvania State University in University Park, Pa., the Blitecast program predicts the appearance of the fungus seven to fourteen days in advance. Data for the program is collected with hygrothermographs and rain gauges, which monitor temperature, humidity and rainfall in each field on a daily basis. At the end of the week, growers telephone the data to researcher John Peplinski, who uses a typewriter terminal in his office to enter the information into the university's System/370 Model 168. The data is analyzed against the conditions that are known to promote fungus growth—warm days, cool nights and extended periods of humidity exceeding 90%.

Within a few seconds, the computer prints out a recommendation indicating whether it will be necessary for the farmer to spray fungicide that week. "It used to be standard procedure to spray automatically once a week," says Peplinski. "By using our program, many growers have cut their spraying schedules substantially, without sacrificing crop safety. At $4 per acre for fungicide, farmers can save considerable money—and cut pollution at the same time—if conditions don't indicate fungus growth."

The computer can also produce a longer report that summarizes blight forecasts, weather conditions and spray recommendations made to date for each participant. A typical file extends for the duration of the potato growing season—generally from May through August.

In research related to the Blitecast program, scientists at Penn State are also studying the extent to which reduced concentrations of fungicide will effectively kill blight, and the precise relationship between weather conditions and tuber maturity. "Our goal is to increase crop yield by every possible means," says Peplinski. "Using the computer is the only way we can get accurate predictions to our farmers quickly enough to be effective."

At the Penn State University test plot in Rock Springs, plant pathologists study conditions causing potato blight.

Supernova is Born...

(Continued from first page)

disassociate from their specific nuclei and form a kind of 'sea' surrounding all the nuclei," explains Dr. Lasher. "This electron sea exerts pressure that we can calculate using the laws of quantum mechanics. But there is a critical mass—generally accepted to be 1.4 solar masses—after which the electrons cannot withstand the gravitational forces of the core. With no other alternative, the core collapses in upon itself.

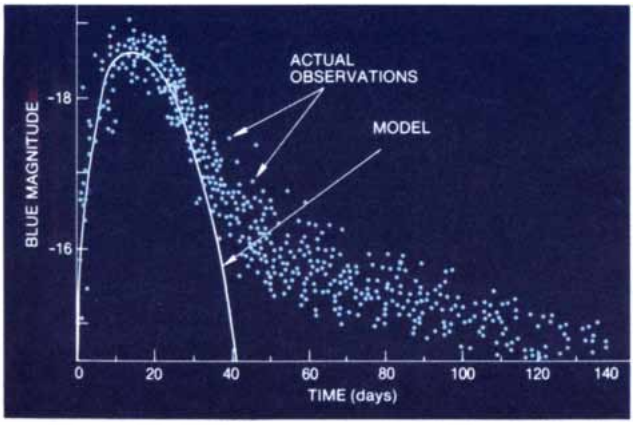
"By accepting the collapsed core theory, we were able to base our model on only three parameters: the total amount of energy released by the collapsing core, the mass of the expanding helium envelope and the initial density of that envelope," adds Dr. Lasher.

The model predicts a roughly parabolic light curve in which light intensity peaks around the 20th day. This finding corresponds closely with actual observations of 38 Type I supernovae collected and organized by a team of Italian astronomers.

Eventually, Dr. Lasher and two post-doctoral fellows at the research center, Dr. Alan Karp and Dr. Kwing Lam Chan, hope to extend their model to describe the velocity at which the gas envelope expands. They are currently using an IBM System/370 Model 168 to do the calculations. The general concept of their work has been based on that of another visitor to the research center, Dr. Charlotte Gordon.

"We still don't know exactly what happens after a supernova consumes all the energy in the core," adds Dr. Lasher. Some astronomers believe supernovae become neutron stars, also called pulsars. Others believe that extremely heavy supernovae (those with cores three times the mass of the sun) may become "black holes".

"There is still a great deal of mystery about the last stages of stellar evolution," says Dr. Lasher. "We hope our model will pave the way for further inquiries."



The supernovae model (smooth curved line) is compared with actual observations of the blue light intensity of 38 Type I supernovae.

DP Science Dialogue is concerned with topics which may prove of interest to the scientific and engineering communities. Your comments and suggestions are welcome. Just write: Editor, DP Science Dialogue, IBM Data Processing Division, 1133 Westchester Ave., White Plains, N.Y. 10604.

76-1S

IBM

Data Processing Division

held to be the inherently inhumane and indiscriminate effects of incendiary weapons has been handicapped by the lack of precise information on the severity of wounds caused by incendiary weapons compared with those caused by other weapons, and on the actual numbers of civilian casualties likely to result from the employment of incendiary weapons. In an attempt to provide the necessary information the Stockholm International Peace Research Institute (SIPRI), an independent organization funded by the Swedish parliament, has compiled a detailed report titled *Incendiary Weapons*. The report was written by Malvern Lumsden, a member of the SIPRI research staff, and has been published in the U.S. in cooperation with the MIT Press.

The SIPRI book describes the historical development of incendiary weapons, together with attempts to restrain the use of these weapons by international agreement. The closest these diplomatic efforts have ever come to success, it seems, was at the disarmament conference of the League of Nations in 1933. In spite of the ultimate failure of that attempt, the SIPRI book relates, incendiary weapons were widely regarded before 1939 "as illegal and inhumane, in the same category as weapons, such as mustard gas, which cause chemical burns. The question of illegality was, however, totally ignored during World War II, when a definite policy of incendiary bombing emerged." The mass incendiary attacks carried out in that period against cities in Germany and Japan killed about a million people, mostly civilians; nonetheless, according to postwar studies, the attacks did not have the decisive effect on the outcome of the war that was claimed at the time.

Since World War II, the SIPRI report continues, napalm bombs have been used increasingly by some 25 countries for air attacks in close support of ground troops, and for air strikes in rear areas independent of ground operations. Particularly large quantities of napalm were expended by the U.S. and its allies in Korea and Indochina. Although official figures on the use of incendiary weapons in Indochina have never been made public, the SIPRI researchers estimate (on the basis of previously unpublished data) that nearly 400,000 tons of aerial napalm bombs were dropped during the war in Indochina by the U.S. and its allies. In addition napalm was dropped in large quantities in oil drums from helicopters and fired from flamethrowers mounted on armored vehicles and river-patrol boats. Counting these weapons, together with white phosphorus, magnesium and thermite weapons (which were employed for a variety of purposes, including several attempts to cause massive forest fires), the SIPRI team concludes that the total quantity of incendiary munitions used in the Indochina war was greater than that used in any other theater of any other war. Moreover, the report asserts, in at least some periods of the Vietnam war the majority of the napalm bombs dropped were intended not for close tactical air support but rather to destroy rear-area targets, including villages.

A large part of the SIPRI book is devoted to a detailed, illustrated examination of the medical problems that arise from both the thermal and the toxic effects of particular incendiary weapons on the human body. The general conclusion is that "the extreme suffering and indiscriminate effects of incendiary weapons are impossible to justify by considerations of 'military necessity.' "

Warning that "serious attempts... to limit the military use of technology by legal and humanitarian constraints" have in the past often "been hindered by a lack of insight into emerging technologies," the SIPRI report points to a "new generation" of incendiary weapons reportedly under development. The emerging weapons include, for example, a device said to contain a new incendiary material capable of creating a "chemical fireball" that radiates enough thermal energy to inflict third-degree burns on exposed people in a few seconds. The only other weapons that produce radiant energy on this scale are nuclear bombs. The new incendiary material (which apparently consists of a highly volatile, self-igniting organometallic liquid such as triethylaluminum, slightly thickened with a long-chain polymer such as polyisobutylene) can be incorporated in a wide variety of grenades, shells and cluster bombs. The pertinent lesson to be drawn from the historical record, the SIPRI study notes, is that "to be effective, political or legal restrictions must be one step ahead of the technology they are trying to hold in check rather than one step behind."

Falling Quanta

Quantum mechanics and the general theory of relativity are surely among the grandest constructions of 20th-century physics, but the two theories seldom intersect. The gravitational effects described by general relativity predominate on a cosmic scale, but they are negligible for particles of small mass. Quantum mechanics determines the behavior of atoms and their constituent particles, but it rarely has an observable effect on macroscopic events. Nevertheless, in a recent experiment the theory of gravitation and the theory of quantum mechanics were tested simultaneously.

The experiment did not represent the first observation of the influence of gravitation on a subatomic particle. That was accomplished 25 years ago by Andrew W. McReynolds of the Brookhaven National Laboratory, who detected the deflection of a neutron falling freely in the earth's gravitational field. The path followed by the neutron is a parabola, just like that of an artillery shell. McReynolds' experiment, however, involved only those properties of the neutron that can be described in the language of classical physics; quantum mechanics was not required.

In the more recent experiment, which also employed neutrons, a quantum-mechanical interpretation is inescapable. The effect observed requires that the neutron be regarded not as a discrete particle but as a wave; through the formalism devised by Louis de Broglie that is possible in quantum mechanics but not in classical physics. The experiment was conceived by Roberto Colella and Albert W. Overhauser of Purdue University and was conducted by them with Samuel A. Werner, then of the Ford Motor Company. They employed the Ford Nuclear Reactor at the University of Michigan, and they describe their work in *Physical Review Letters*.

The experiment was performed with a neutron interferometer, which is analogous in function to the more familiar light interferometer. In the latter instrument a beam of light is split into two beams, then recombined to generate a pattern of light and dark interference "fringes," which vary according to the relative phase of the two beams. In the light interferometer the beams are reflected by mirrors or prisms; in the neutron interferometer they are reflected by the planes of atoms in a large single crystal. Colella, Overhauser and Werner employed a cylindrical crystal of silicon about three inches long and two inches in diameter.

Parts of the crystal were cut away to leave three semicircular slabs connected by the remaining portion of the cylinder. The neutron beam struck the first of the slabs obliquely and by partial reflection was split into two beams. Both of those beams were split by the second slab, and at the third slab two of the beams were recombined. The flux of neutrons in the two interfering beams and in one of the

noninterfering beams was measured with small detectors placed beyond the third slab.

The influence of gravitation on the neutrons was tested by rotating the interferometer around the axis of the incident beam. The two interfering beams formed a parallelogram, which initially was horizontal. When the apparatus was turned 90 degrees in either direction, one leg of the parallelogram was raised above the other leg by somewhat less than an inch. If gravitation had no effect on the quantum-mechanical behavior of neutrons, no change in the output of the detectors would be expected. Actually the rotation of the interferometer produced fluctuations in output representing clearly defined interference fringes. The number of fringes was in reasonable agreement with theoretical predictions. The method thus detects a difference in the earth's gravitational potential over a vertical distance of less than an inch. Indeed, the interferometer is so sensitive that it was necessary to correct for the bending of the crystal under its own weight.

A classical interpretation of the flight of a neutron in a gravitational field would show that when the interferometer is turned and one beam is elevated over the other one, the neutrons in the higher beam experience a smaller acceleration toward the center of the earth. The classical explanation is impossible, however, because the experiment involves interference, a phenomenon characteristic of waves. A quantum-mechanical interpretation is mandatory. It demonstrates that the difference in the altitude of the particles in the two beams results in a difference in the phase of the corresponding waves.

Overhauser points out that the experiment is a test of the principle of equivalence, an axiom underlying the general theory of relativity. The principle states that mass as measured by gravitational attraction and mass as measured by inertia cannot be distinguished. The principle has been repeatedly verified on a macroscopic scale; this is apparently the first demonstration that it applies in the realm that is governed by quantum mechanics.

More recently Colella, Overhauser and Werner, together with C. F. Eagen of the Ford Motor Company, have employed the neutron interferometer to demonstrate another principle of quantum mechanics. Subatomic particles have intrinsic spin angular momentum, which in the case of the neutron is measured in half-integral units. Quantum mechanics requires that when the spin axis of a particle with half-integral spin is made to precess through one complete rotation, the sign of the wave function for the particle is reversed. This prediction contradicts classical physics (and common sense), which states that an object does not change at all when it is rotated through 360 degrees. Overhauser and his colleagues caused the spin axes of neutrons to precess by passing one beam in their interferometer through a magnetic field. A change in the sign of the wave function is equivalent to a 180-degree change in phase, so that the effect of the precession was observed as a change from constructive interference to destructive interference.

Colella, Overhauser and Werner are now preparing to repeat their gravitational experiment at the University of Missouri, with which Werner is now affiliated. With improved accuracy they hope to detect the influence of the rotation of the earth on the phase of the neutron de Broglie waves. The analogous experiment with light was successfully performed 50 years ago by A. A. Michelson, H. G. Gale and F. Pearson.

Epidemic on Wheels

Deaths and injuries from motor-vehicle accidents are reaching epidemic proportions in developing countries around the world, according to the World Health Organization. Traffic accidents in the young nations of Africa amount to a "social scourge," and all too often the victims are young, educated Africans whose increased earning power has enabled them to buy a motorcycle or an automobile. Statistics from three Latin-American countries, Chile, Costa Rica and Venezuela, reveal that, as in the U.S., traffic accidents have become the leading cause of death among young adults.

About 250,000 people throughout the world are killed in traffic accidents each year, and more than seven million are injured. Although the U.S. has the highest number of people killed in traffic accidents of any country (about 50,000 per year), it has one of the lowest rates of fatalities per motor vehicle or passenger mile. For example, in the U.S. there are six fatalities per 100 million passenger miles, whereas in Kenya and Uganda there are from 55 to 65 fatalities per 100 million passenger miles. In India the fatality rate per motor vehicle is 10 to 15 times higher than it is in the U.S. In all countries the death rate from traffic accidents is higher for males than it is for females.

The majority of developing countries have a higher incidence of traffic accidents involving pedestrians than of accidents involving motor vehicles alone. Among the causes, the WHO reports, are poor roads, pedestrian ignorance of road signs, lack of instruction in the use of roads and heavy pedestrian and bicycle traffic on the roads.

To combat the growing epidemic of traffic accidents, the WHO has undertaken a worldwide epidemiological study of road traffic accidents and is encouraging the development of preventive programs. "If traffic accidents are tackled by methods similar to those used against the great killing diseases," the organization states, "the present epidemic of road deaths could be made to disappear just as plague and smallpox have now been eliminated almost everywhere in the world."

Good Return

It is sometimes alleged that modern food-production techniques depend so heavily on energy inputs—for fertilizer, farm machinery, processing and trucking—that they are quite inapplicable in underdeveloped countries and in the long run are too wasteful of energy even for developed countries such as the U.S. Fertilizers, for example, improve crop yields, but their production, transport and application call for large quantities of energy. Are fertilizers genuinely effective in terms of energy return, that is, do they enhance yields enough to balance their consumption of energy? The answer is yes, according to R. G. Hoeft and J. C. Siemens of the University of Illinois. In an article in *Crops and Soils Magazine* they argue that the application of nitrogen fertilizer to a cornfield produces more extra energy value than is spent to make, haul and spread the fertilizer, and so the energy investment is a good one.

Green plants convert solar energy into chemical energy; a bushel of corn, for example, stores 95,000 kilocalories of energy. Of the nutrients required by plants, nitrogen is most likely to be present in an amount that is inadequate for maximum yield and thus for maximum energy conversion. In the U.S. corn belt a nitrogen fertilizer such as anhydrous ammonia, urea or ammonium nitrate, if applied at the recommended economically effective rate of 200 pounds of nitrogen per acre, results in an increased corn yield of about 83 bushels per acre, or an energy return of some 7.9 million kilocalories per acre. Hoeft and Siemens subtracted from that the energy required to make and haul 200 pounds of nitrogen

in the form of anhydrous ammonia and apply it over an acre, and ended up with a net energy return of nearly 6.6 million kilocalories per acre. In terms of energy harvested per unit of fertilizer energy expended, the return is 5.95 to 1.

The Killer Bees Revisited

Much attention has been given in recent years to the fact that honeybees of an African race (*Apis mellifera adansonii*) were accidentally released in Brazil in the mid-1950's, where they have been cross-breeding with commercial honeybee populations and also have established their own wild colonies. The northward and southward expansion of the bees' range from the original point of release, coupled with the African race's quick reaction to any disturbance of the hive and its tendency to pursue the disturber with unusual persistence, have led to predictions that the "killer bees" might eventually spread a reign of terror throughout the Western Hemisphere, including the U.S. The threat has now been assessed by Roger A. Morse, professor of apiculture at Cornell University. He dismisses it as being grossly exaggerated.

Writing in *New York State Environment,* Morse points out that the African bees' reputation for aggressiveness is no worse than that of more than one other honeybee race, such as the Cyprian bee, that is regularly raised by beekeepers. Visiting Brazil in 1973, Morse found commercial beekeepers enthusiastically expanding their Africanized bee colonies because the hybrids produced some 40 percent more honey than nonhybrids. Asked about the "nastiness" of the African race, the beekeepers replied that the increased yield more than compensated for wearing an extra pair of pants and receiving a few more stings than usual while harvesting the honeycombs. Moreover, the hybrids were reported to be less aggressive than the pure African race. Their responsiveness to being disturbed fell about midway between that of the African *adansonii* and that of the least aggressive honeybee race, the Italian *ligustica.*

As for the possibility that the African bees would invade the U.S., Morse notes that *adansonii* is adapted to living under tropical and subtropical conditions. The African bees are unable to survive the cold season in temperate climates as other honeybees do by forming winter clusters, and the Africanized hybrids share this behavioral handicap. Thus the temperate zones in both South and North America are barriers to further migration by either the pure or the hybrid African bees.

Christopher Wren: Astronomer

It is well known that Christopher Wren was a leading architect of the 17th century; it is less well known that he was also a leading astronomer. J. A. Bennett of the Royal Astronomical Society has recently examined Wren's career as an astronomer and the connection between his astronomical interests and his architectural ones.

Writing in *Journal for the History of Astronomy,* Bennett describes the vigorous mathematical tradition of English natural philosophy in the 17th century. At that time mathematics was regarded less as an academic subject than as a practical one, yet pure mathematics was an integral part of the tradition. The mathematical sciences actually comprised the diverse disciplines of astronomy, music, surveying, the measurement of quantities of all kinds, perspective drawing, the design of optical instruments, fortification, machines and architecture.

It was in this tradition that Wren was educated at Oxford, was appointed Gresham Professor of Astronomy in 1657 and was one of the founders of the Royal Society of London. His work in astronomy ran the gamut from the theoretical through the observational to the practical. He theorized about the peculiar elongated appearance of Saturn (which his contemporary Christiaan Huygens later showed was due to the planet's rings), he devised a graphical method for calculating the time and extent of eclipses of the sun and of occultations of stars by the moon, and he considered the problem of how to determine longitude on the earth. He proposed a program for trying to determine whether or not the stars showed an annual parallactic motion as a result of the earth's revolution around the sun (as had been predicted by Copernicus). He designed several telescopes and a machine for grinding lenses in such a way that the lenses would not suffer from spherical aberration.

Wren's shift from astronomy to architecture was a natural one. Moreover, the transition was not sudden nor was it ever complete. He began working on architectural projects in the early 1660's, while he was still professor of astronomy. Yet even as architectural commissions came to occupy most of his time, he never lost interest in natural philosophy and never lost his position in the scientific community; indeed, in 1681 he was elected president of the Royal Society. Thus his transition from astronomy to architecture was apparently more of a professional change than an intellectual one, a shift from one of the mathematical sciences to another.

"Wren has always been something of an anomaly for architectural history," Bennett concludes. "The answer may lie in looking at his work, not so much in relation to the mainstream of European architectural development, but as a unique expression of a distinctive English tradition in the mathematical sciences."

Knucklers

As every true baseball fan knows, a knuckleball is a pitch that changes direction unpredictably as it approaches the batter. The pitcher holds the ball with his fingernails or the first joints of his fingers, so that when he delivers the ball it spins much less than it would if he were holding it as he would normally. Robert G. Watts and Eric Sawyer of Tulane University have investigated the dynamics of the knuckleball by mounting a baseball in a wind tunnel in such a way that the forces acting on it in the airstream can be measured. First of all, they confirm that the knuckleball is capable of changing direction as it approaches the batter; the forces acting on a slowly spinning ball are sufficient to make it veer more than a foot in the distance between the pitcher and the batter. Writing in *The American Journal of Physics,* they say that a knuckleball veers because the seams on the slowly rotating ball "form a roughness pattern that is nonsymmetric," so that the flow of air past the ball is uneven. "This will naturally cause a nonsymmetric lateral force distribution and result in a net force in one direction or another."

Some pitchers think a knuckleball should be thrown so that it does not spin at all. They are wrong, according to Watts and Sawyer, who found that when the ball is not spinning, it is not subject to the varying lateral force that acts on it when it is spinning slowly. It is this varying force that makes the flight of the knuckleball so unpredictable.

The ball should veer most, write Watts and Sawyer, when it is thrown so that it makes about a quarter of a revolution on its way to the plate. It should not matter much how fast the ball is thrown; since "the magnitude of the lateral force increases approximately as the square of the velocity," the total lateral deflection "is independent of the speed of the pitch."

A DNA OPERATOR-REPRESSOR SYSTEM

An operator is a segment of DNA adjacent to a gene; a repressor is a protein that binds to the operator and controls the expression of the gene. How such a system works in a virus is explored in detail

by Tom Maniatis and Mark Ptashne

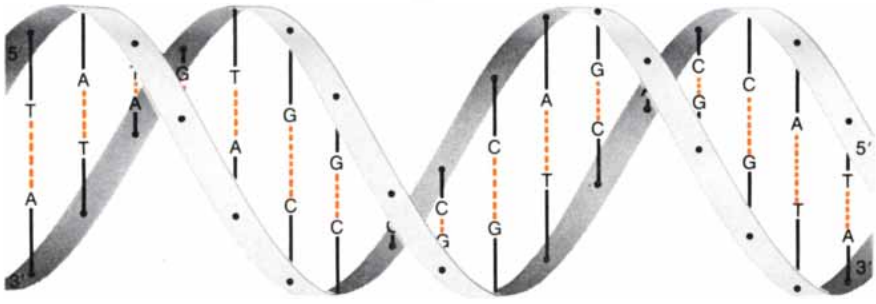
A fundamental property of living cells is their ability to turn their genes on and off in response to extracellular signals. In the human body, for example, every cell (with the exception of a few cell types such as the red blood cell) has the same set of genes, yet in the course of embryonic development cells take on different shapes and functions as their genes are selectively switched on and off. How are the genes regulated? Is there a common mechanism underlying such regulation in different organisms?

Through the study of gene regulation in bacteria and viruses it has been learned in recent years that a fundamental mechanism of gene control depends on the interaction of protein molecules with specific regions on the long-chain molecule of DNA, the material that embodies the genetic instructions of all organisms from bacteria to man. As a result of this interaction genes are switched on or off. In the best-understood instances genes are switched off by controlling molecules named repressors. The existence of repressors was first hypothesized in 1960 by François Jacob and Jacques Monod of the Pasteur Institute in Paris. Seven years later Walter Gilbert (in collaboration with Benno Müller-Hill) and one of us (Ptashne), working independently at Harvard University, succeeded in isolating repressors from bacteria [see "Genetic Repressors," by Mark Ptashne and Walter Gilbert; SCIENTIFIC AMERICAN, June, 1970]. Later it was shown that repressors could bind tightly and specifically to sites on DNA called operators and that in so doing they could prevent genes adjacent to the operators from being transcribed and translated into proteins.

Since these early discoveries we and many others have pursued the molecular details of gene repression. This article is a brief progress report on some of the things that have been learned. We now know, for example, the sequence of bases, or code units, in DNA that constitutes the operators to which a repressor binds. In the case we shall discuss here the operators have several nearly identical binding sites, each capable of being recognized by the same repressor molecule. We are only beginning to learn why several sites are provided when seemingly one would do the job.

Before we describe this recent work let us quickly review the molecular structure of the gene. In man, as in bacteria, a gene can be defined as a sequence of bases along a DNA molecule. (In certain viruses the gene consists of RNA rather than DNA.) The DNA molecule consists of two long chains of nucleotides wound in a double helix and linked to each other by hydrogen bonds. Each nucleotide consists of a deoxyribose sugar, a phosphate group and one of four nitrogenous bases: adenine, guanine, thymine or cytosine (abbreviated *A*, *G*, *T* and *C*). The sugar and phosphate groups form the backbone of each chain; the bases extend toward the central axis of the double helix and pair with the bases extending from the other chain. The sequences of bases along the chains are complementary: *A* always pairs with *T* and *G* always pairs with *C*. The information content of DNA is specified by the sequence of bases. A typical gene consists of roughly 1,000 base pairs.

The translation from gene to protein begins when the enzyme RNA polymerase copies the base sequence into a complementary sequence on the linear molecule of "messenger" RNA. The intracellular translating machines called ribosomes attach themselves to the messenger RNA and translate its base sequence into a sequence of amino acids, which are linked to form a protein molecule. Since there are only four different bases and 20 different amino acids, a sequence of three bases is needed to spec-

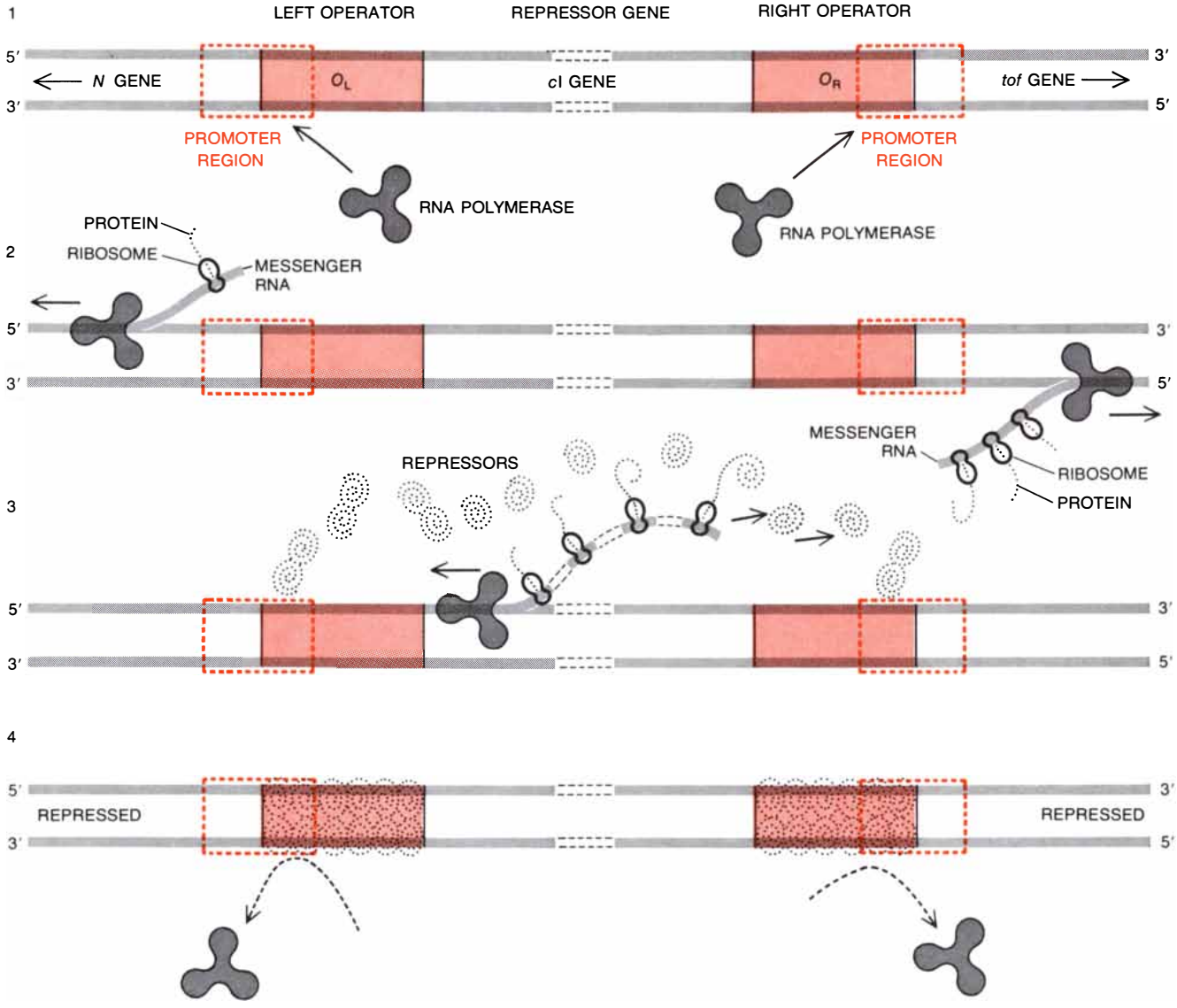


DOUBLE HELIX OF DNA encodes the genetic information of all cellular organisms and most viruses. (In some viruses the genetic material is RNA.) The genetic code is written in the particular sequences of nitrogenous bases that connect the two strands of the DNA molecule. The bases are of four kinds: adenine (*A*), thymine (*T*), guanine (*G*) and cytosine (*C*). *A* always pairs with *T* and *G* always pairs with *C*. The strands of the double helix consist of alternating subunits of phosphate and ribose, a five-carbon sugar. In one strand the phosphate links the No. 5 carbon in one sugar to the No. 3 carbon in the adjacent sugar, creating what is denoted a 5′–3′ linkage. In the opposing strand, proceeding in the same direction, the linkage is 3′–5′. Thus each strand of DNA molecule has a 5′ end and a 3′ end.

ify one amino acid. (For example, *ACA* specifies the amino acid threonine.) It is clear that the translation of a gene into a protein molecule might conceivably be repressed, or blocked, at any one of several stages along this complex pathway. It turns out in the case we have studied that repression takes place directly at the DNA molecule, so that the genetic information is not transcribed into messenger RNA unless the repressor is inactivated. Just how is this repression achieved?

Bacteriophage Lambda

The repressor we have studied is a protein molecule manufactured by bacteriophage lambda, a virus that infects the common colon bacterium *Escherichia coli*. The bacteriophage, or phage, particle consists of a single DNA molecule with about 47,000 base pairs, enough for some 50 genes, enclosed in a protein coat equipped with a tail through which the DNA is injected into the bacterial cell. Once the DNA is inside the cell it can follow either of two pathways. It can complete the course of infection by causing the machinery of the bacterial cell to translate the phage genes into proteins. Some of the proteins are enzymes that replicate several hundred



EXPRESSION AND REPRESSION OF GENES is conveniently studied in bacteriophage lambda, a virus that infects the bacterium *Escherichia coli*. When the 50 or so genes of the phage-lambda DNA are translated into protein inside a cell of *E. coli*, the phage multiplies and kills the cell in about 45 minutes. In some cases, however, repressor proteins specified by a particular gene in the lambda DNA, the *cI* gene, get the upper hand and block the transcription of genes on either side, forcing the lambda DNA into a state of dormancy. Normal transcription of lambda DNA into "messenger" RNA and its translation into protein molecules are depicted in *1* and *2*. The transcription is effectuated by the enzyme RNA polymerase, which attaches itself to a promoter region and assembles an RNA chain in the 5′–3′ direction by copying a DNA strand of the opposite polarity. Thus the RNA-polymerase molecules travel in opposite directions, copying the different strands of the DNA as they transcribe into messenger RNA the complete instructions for replicating the phage, beginning with the *N* gene on the left and the *tof* gene on the right. Transcription of these two genes begins just outside the operator. The ribosomes fasten themselves to the emerging "tape" of messenger RNA and translate the encoded message into the protein molecule. (In this diagram the structures are not drawn to scale; the ribosomes in particular are much larger than they are shown.) Under other conditions (*3*) the *cI* repressor gene is transcribed by the same process and translated into repressor molecules. (Specific terminator signals in the DNA prevent RNA polymerase from continuing on into left operator.) Perhaps singly as monomers, but more likely as dimers depicted in *4* or even as tetramers, repressor molecules migrate to binding sites in two operator regions of lambda DNA, blocking access of RNA polymerase to promoter regions of *N* and *tof* genes.

copies of the phage-DNA molecule; other phage proteins package each DNA copy in a protein coat, thus creating multiple copies of the original phage particle. Typically within 45 minutes the bacterial cell, swollen with phage particles, bursts.

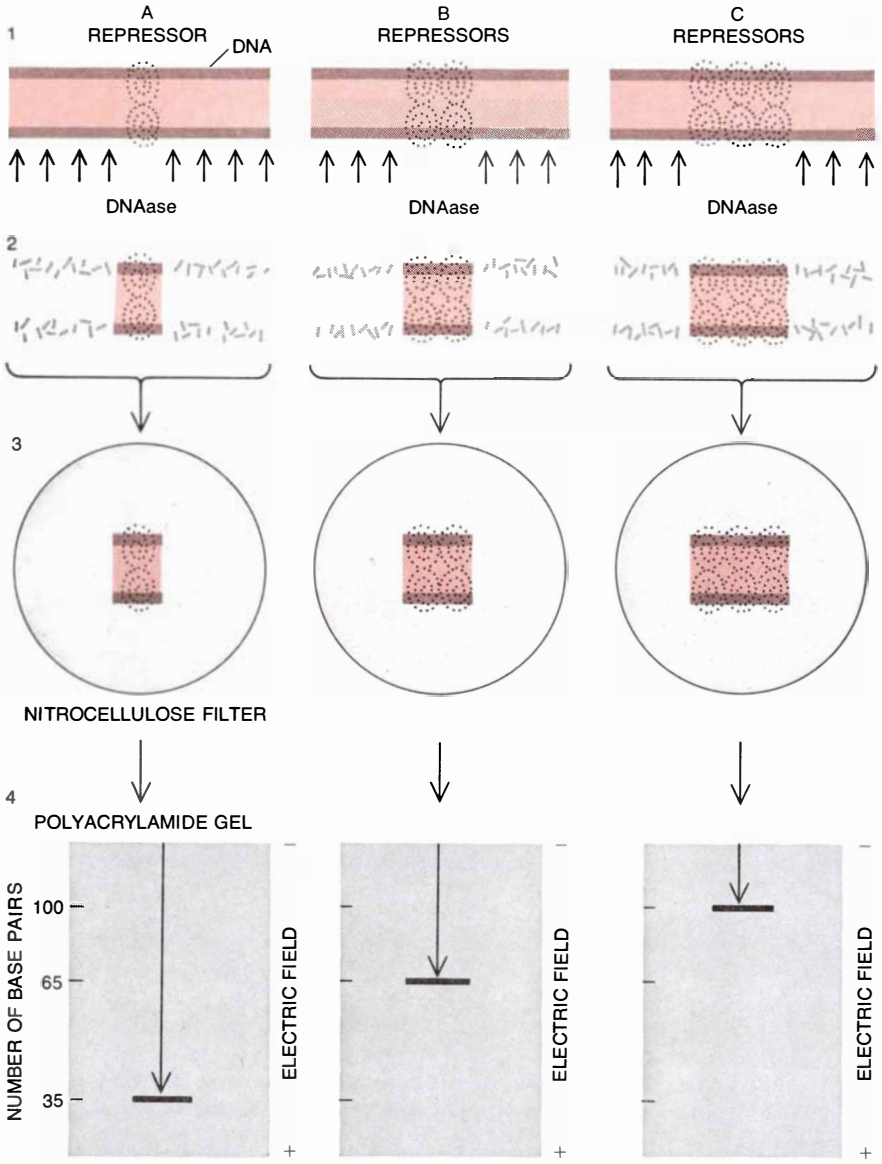
The other pathway is the more interesting one for our purposes. Occasionally after the phage genes enter the bacterial cell they are switched off and the phage DNA becomes integrated into the DNA of the host cell. There it remains dormant, replicating with the bacterial DNA at every cell division and giving rise to a population of *E. coli* cells each of which contains a chain of phage genes. The dormant phage genes are called a prophage.

It has been known for some years that the phage genes are turned off by a specific repressor molecule specified by one of the phage's own genes, the *c*I gene. The repressor actually binds to two separate operators on the phage-DNA molecule, thereby blocking the transcription of two different sets of genes. The turning off of these two sets of genes is sufficient to cause the 40-odd remaining genes, with the exception of *c*I, to stop functioning. The dormant phage genes can be switched on again by a suitable inducing agent such as a low dose of ultraviolet radiation, which causes the repressor to be inactivated.

The two operators, which are separated by some 2,000 base pairs (including the *c*I gene), are designated O_L and O_R, the subscripts denoting left and right. Repressor bound to O_L blocks the transcription of gene *N* and repressor bound to O_R blocks the transcription of gene *tof*. Gene *N* is transcribed to the left beginning near O_L and gene *tof* is transcribed to the right from the opposite DNA strand, beginning near O_R. (DNA chains have a polarity determined by the orientation of the sugar-phosphate linkage in their backbone. In each double helix the linkage is designated 5′–3′ in one chain and 3′–5′ in the other. The numbers 3 and 5 refer to the third and fifth carbon atoms in the five-carbon sugar molecule. RNA is assembled in the 5′–3′ direction, copied from a DNA strand of the opposite polarity.)

One can speculate that it is clearly to the phages' advantage for some of them to go into the prophage, or dormant, stage and to be "revived" at a later time under conditions that may be more favorable for multiplication. The phage genes are reactivated by inactivation of the repressor. Although there are many conditions that result in repressor inactivation, the details of the process are not fully understood.

The lambda-phage repressor is a protein composed of subunits that have a molecular weight of about 27,000. When the concentration of the subunits in a suitable medium is increased, they form dimers and tetramers: two-subunit and four-subunit associations. When the concentration is reduced, they dissociate. Dimers or possibly tetramers must form before the repressor is able to interact strongly with DNA. We shall return to the possible significance of this fact.



SIZE OF REPRESSOR BINDING SITES in operator regions of lambda DNA was established by mixing the DNA with repressor molecules. Wherever repressors bind to the DNA they protect it from digestion when the enzyme DNAase is added to the mixture (*1, 2*). As the ratio of repressor to DNA is increased (*A, B, C*) the segment of DNA protected increases in length, thus indicating that repressor binds to several adjacent sites. The pieces of protected DNA are collected on nitrocellulose filters (*3*), extracted and subjected to electrophoresis in a polyacrylamide gel (*4*). The fewer base pairs present in a fragment, the farther it will migrate when it is subjected to an electric field. Evidently this repressor can cover from 30 to 100 base pairs. If the DNA has been labeled with atoms of a radioactive isotope such as phosphorus 32, location of fragments can be visualized with autoradiogram.

The Isolation of Lambda Operators

Regions on the DNA molecule that are bound to specific proteins can be isolated by virtue of the fact that the enzyme DNAase digests any naked DNA it encounters but leaves intact any DNA that is covered by a protein. This property of DNAase was utilized by Allan Maxam and Gilbert to isolate an operator region from the DNA of *E. coli* and was

Q. When Cadillac made the Seville which car should they have copied?

The Seville was designed to be a smaller-sized luxury car that got good gas mileage. Naturally, they made it roomy. But, even though it's almost 22" longer than the Audi 100LS, the Audi has virtually the same legroom and headroom. And more hiproom. And trunk space.

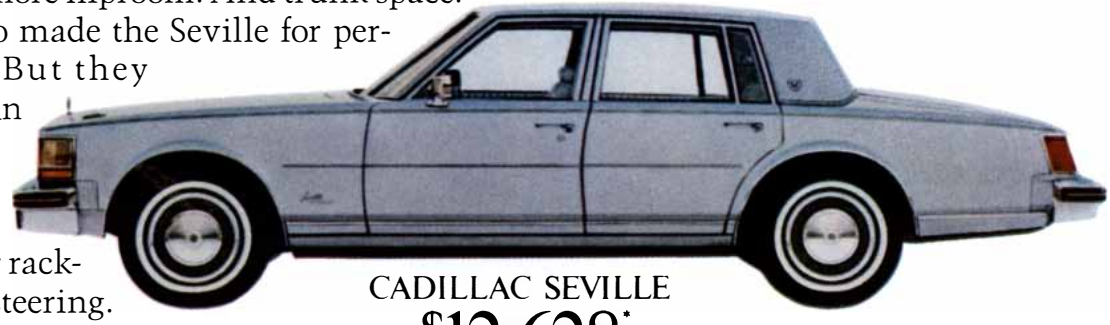
They also made the Seville for performance. But they didn't put in features like front-wheel drive as in the Audi. Or rack-and-pinion steering. Or negative roll radius.

And, for a Cadillac, it got good gas mileage. But it just doesn't come close to Audi's economy. EPA estimated that the standard shift Audi gets a terrific 30 miles per gallon on the highway and 20 miles per gallon in the city.

And, finally, they put a price tag on the Seville, one that was more than $5000 higher than the Audi.

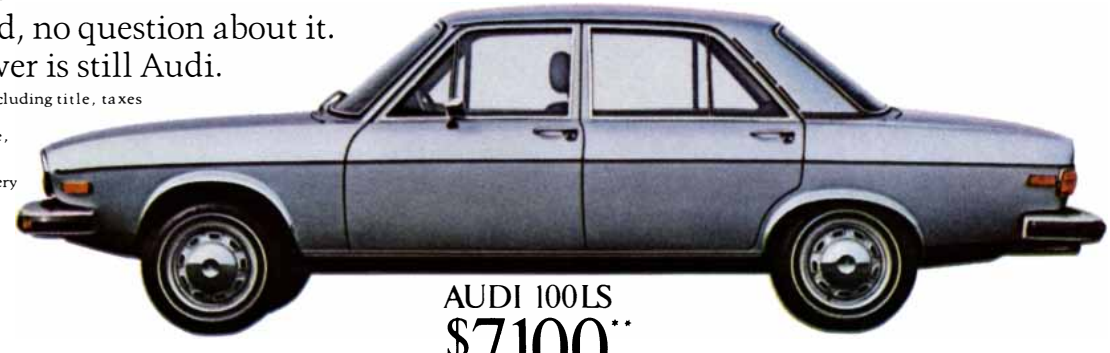
They tried, no question about it. But the answer is still Audi.

HEADROOM FRONT 38.6 REAR 37.0
LEGROOM FRONT 41.6 REAR 38.3
HIPROOM FRONT 51.4 REAR 51.7
TRUNKROOM 12.8 CU. FT.
MILEAGE (EPA EST.) HWY 21 MPG CITY 15
AUTOMATIC TRANSMISSION



CADILLAC SEVILLE
$12,628*

HEADROOM FRONT 38.8 REAR 36.3
LEGROOM FRONT 41.6 REAR 34.6
HIPROOM FRONT 55.7 REAR 55.2
TRUNKROOM 13.38 CU. FT.
MILEAGE (EPA EST.) HWY 30 MPG CITY 20
STANDARD SHIFT
ACTUAL MILEAGE MAY VARY DEPENDING ON TYPE OF DRIVING, DRIVING HABITS, CAR'S CONDITION AND OPTIONAL EQUIPMENT.



AUDI 100LS
$7,100**

*Based on sticker price excluding title, taxes and destination charges.

**Suggested retail price, standard shift, East Coast P.O.E. Local taxes, transportation and other delivery charges, additional. Automatic transmission optional at extra cost.

The answer is Audi

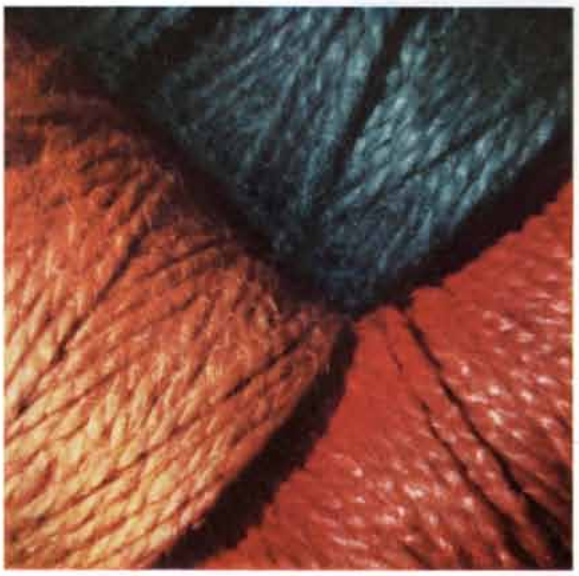
These days, couldn't your business use a small miracle?

Closed, the Polaroid SX-70 Land camera is just about as big as an appointment book. Open, it's a very valuable asset.

The SX-70 can accurately record anything from 10.4 inches to infinity. Or give you a life-size image with the optional close-up attachment.

And since operating a business demands a lot of attention, it's good to know that operating the SX-70 doesn't.

Just focus, frame and push the button. The camera's electronic system gives you the right exposure.



Instantly you have a developing picture in your hand. There's nothing to peel or throw away. More important, there's nothing to pack and send away. Then wait for.

You know you have the picture you want because you see it right away.

You can watch the sharp, crisp images come to life in true-to-life colors. It takes only minutes.

But if things are moving quickly, the SX-70 is ready to shoot again every 1.5 seconds. So you can record every step of an event as it happens.

And since SX-70 prints are encapsulated in durable plastic, you can drop the photos into your pocket. They're tough enough to take it.

And stable enough to last. The metallized dyes make Polaroid instant prints extremely fade-resistant.

The SX-70 is a miracle of time, convenience and dependability. And a very wise investment.

The SX-70 System from Polaroid

Not everybody can do what you're doing now.

Today in our country, just about 21 million Americans over 16 can't even read a want ad.

The 5 million we're trying to help are part of the next generation.

Kids who will learn to read if they want to. And want to read if they're given the incentive. And the books.

Let's face it.

Unless you can read, your opportunities for moving ahead in this world are pretty much limited to the unemployment line.

That's why RIF (Reading Is Fundamental) is helping kids help themselves with 400 local, community projects all across the country. In this Bicentennial year, our goal is to put 25 million books into the hands of 5 million kids who might not read without them.

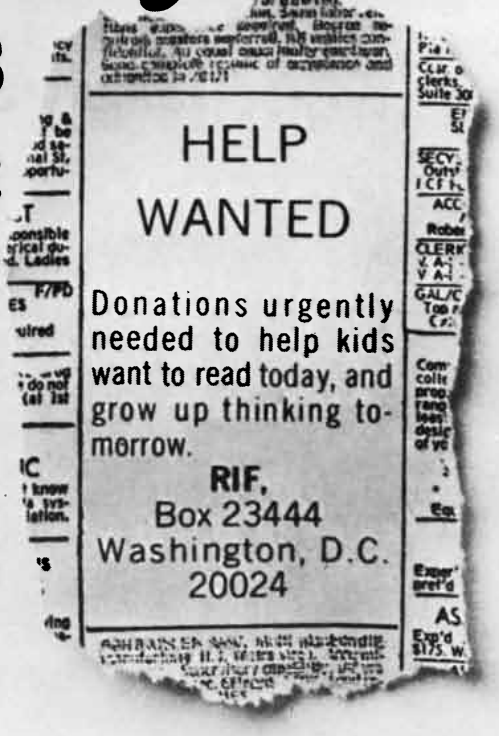
Our problem is money. We need more of it to keep going. And keep helping.

It's tough to ask for money. And we know it's tough to give it. But think how much tougher it would be if you couldn't even read this ad.

Send your tax-exempt donation to RIF, Box 23444, Washington, D.C. 20024.

If America is going to grow up thinking, Reading Is Fundamental.

5 million kids want you to read this ad because they may never be able to.



If America is to grow up thinking,
READING IS FUN•DAMENTAL

later utilized by Vincenzo Pirrotta, then a postdoctoral fellow working with one of us (Ptashne), to isolate the operators on the DNA of the lambda phage to which the repressor becomes attached.

By growing phage particles in a nutrient medium of *E. coli* cells containing the radioactive isotope phosphorus 32 Pirrotta obtained molecules of phage DNA in which the radioactive atoms replaced many nonradioactive phosphorus atoms. The highly radioactive DNA was mixed with purified repressor and then with DNAase. The protected segments of operator DNA were recovered simply by passing the mixture through a nitrocellulose filter. The repressor, like many other proteins, binds tightly to the filter whereas free DNA and DNA-digestion products pass through it. The operator fragment bound to repressor is retained in the filter and can be washed out with a detergent solution. The operator fragment isolated in this way was found to be surprisingly large. A protein the size of the lambda-repressor dimer or tetramer should cover only 15 to 30 base pairs of DNA, but Pirrotta found that roughly 85 base pairs were protected from DNAase digestion by repressor.

We continued these studies of the lambda operators by trying to discover why such an unexpectedly large stretch of DNA is covered by repressor. One possibility we considered (the right one, it turned out) was that the lambda operators may have more than one repressor binding site. We reasoned that if the operators do have more than one binding site, the size of the fragment protected from DNAase digestion should depend on the ratio of repressor to operator in the digestion mixture.

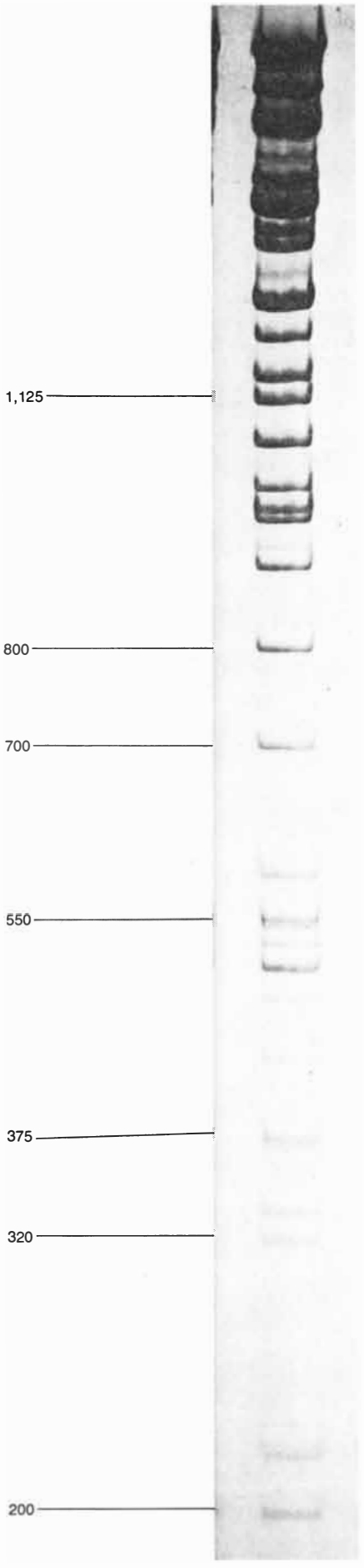
We tested this possibility as follows. We repeated Pirrotta's procedure but varied the ratio of repressor to operator in the DNAase digestion mixture. We then determined the size of the operator fragments by subjecting them to electrophoresis in a polyacrylamide gel. The gel acts as a molecular sieve; under the influence of an electric field smaller DNA fragments migrate through the gel faster than larger ones. Since the operator fragments were labeled with radioactive phosphorus their position at the end of electrophoresis could be determined by placing the gel on photographic film and making an autoradiograph.

We discovered not only that the size of the operator fragment protected by repressor increases as the ratio of repressor is increased but also that the size increases in discrete steps. At low ratios of repressor to operator a single DNA fragment about 30 base pairs long is recovered. At the highest ratios a single fragment about 100 base pairs long is obtained. At intermediate ratios the fragments are of several intermediate lengths. Moreover, this interesting result was obtained whether we used lambda DNA containing only the left operator or lambda DNA containing only the right operator. These experiments and related ones led us to conclude that each lambda operator does in fact have more than one binding site for repressor. Apparently repressor molecules can line up adjacent to each other on the operator, covering a minimum of 30 base pairs and a maximum of 100.

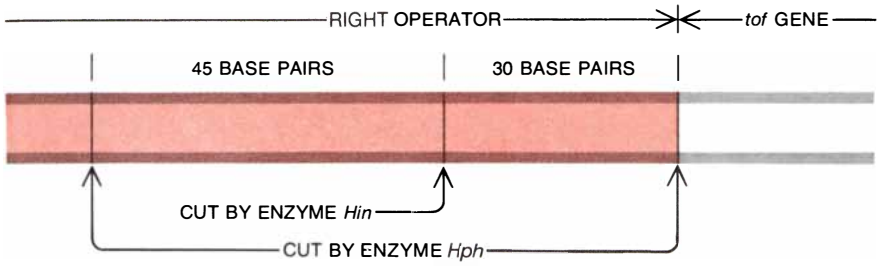
An analysis of the complexity of the DNA sequence in the various operator fragments identified in these experiments revealed two important additional facts about the structure of the two operators. First, the repressor does not bind randomly to any site within the 100-base-pair operator sequence. Rather, it binds initially at a site adjacent to the *N* gene (in the case of O_L) and at one or two sites adjacent to the *tof* gene (in the case of O_R). As the repressor-to-operator ratio is increased secondary sites adjacent to the first sites are filled. Second, the base sequences of the various repressor binding sites are similar but not identical. We obtained strong evidence supporting these two facts by certain experiments we need not review here and confirmed them by the work we shall now describe.

Host-Restriction Endonucleases

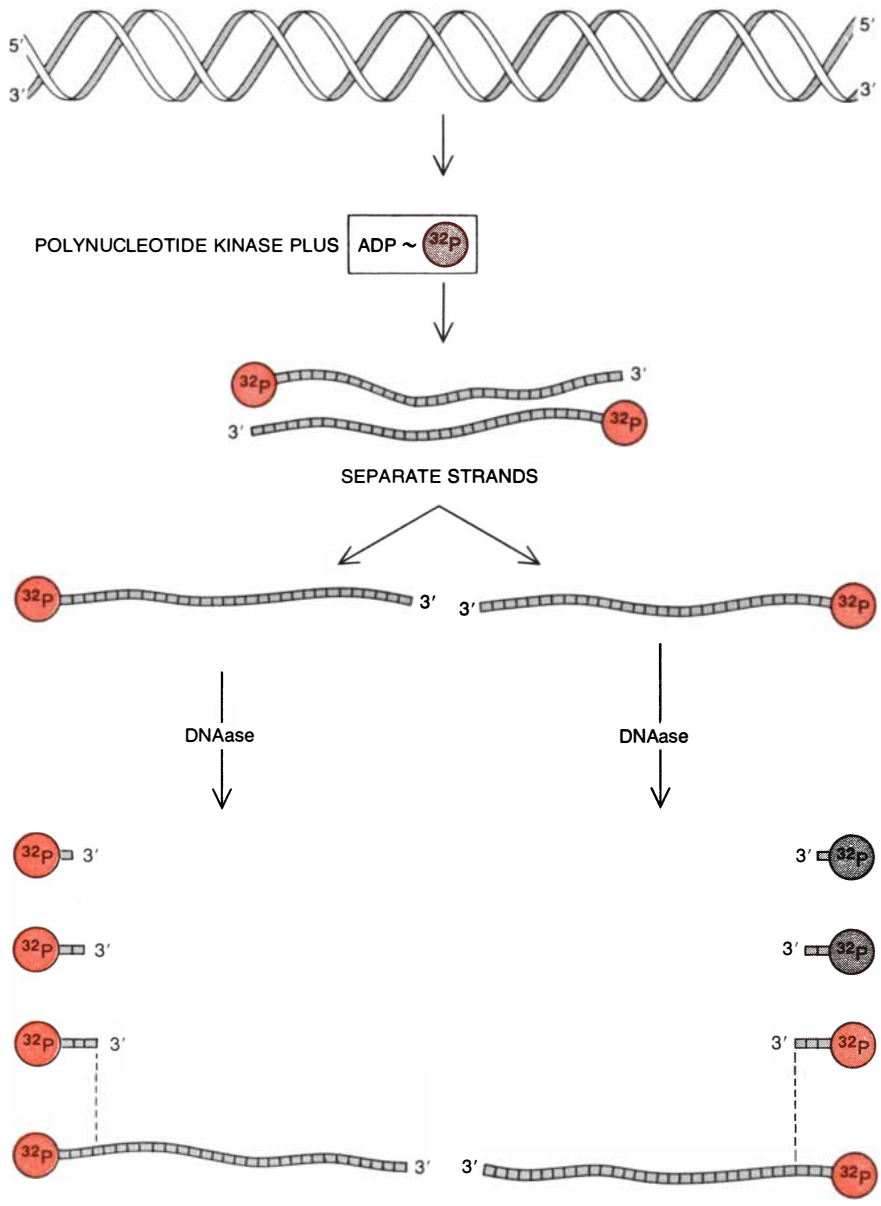
Experiments on the properties of DNA molecules have been revolutionized in recent years by the use of the enzymes known as host-restriction endonucleases. Most of these enzymes, which are widely distributed in the bacterial kingdom, have the remarkable property of recognizing certain base sequences in DNA molecules and cutting the DNA within those sequences. For example, an enzyme (abbreviated *Hin*) isolated from



AUTORADIOGRAM shows the result of digesting lambda DNA with a host-restriction endonuclease, an enzyme (*Hin*) isolated from the bacterium *Haemophilus influenzae*. The enzyme cuts DNA when it encounters a particular sequence of six bases. It cuts lambda DNA into about 50 fragments whose sizes are established by electrophoresis. Numbers beside autoradiogram show number of base pairs in representative fragments.



HOST-RESTRICTION ENDONUCLEASES were used to cut lambda-phage DNA in various ways until fragments were obtained in which the base sequence of the operator could be determined. An enzyme (*Hph*) that was isolated from the bacterium *Haemophilus parainfluenzae* yields a DNA fragment consisting of 75 base pairs that include most of the right operator, which lies next to the *tof* gene. The fragment can be cut in two by enzyme *Hin*.



DETERMINATION OF BASE SEQUENCE in segments of operator DNA is established with the help of enzymes. First, polynucleotide kinase is used to attach an atom of phosphorus 32 (^{32}P) to the 5′ end of each strand in an operator segment including perhaps 30 bases. The strands are separated and mixed with endonuclease, a DNAase that degrades DNA starting from 3′ ends. By removing samples at intervals fragments from one base to 30 bases in length are obtained. Subsequent steps are described in illustration on opposite page.

the influenza bacterium *Haemophilus influenzae* cuts DNA in the middle of the sequence *GTTGAC* paired (on the other chain of the double helix) with *CAACTG*. Another enzyme (*Hpa*) isolated from the bacterial strain *Haemophilus parainfluenzae* cuts DNA in the middle of the sequence *CCGG* paired with *GGCC*. We need not be concerned here that such enzymes may help to protect bacterial species from foreign DNA's (hence the term restriction enzymes) or that some restriction endonucleases act in a more complex way than *Hin* and *Hpa*. The important point is that because the recognition sequences are short they tend to appear at many specific sites on DNA molecules. For example, the enzymes *Hin* and *Hpa* both cut lambda DNA at about 50 sites, generating about 50 specific DNA fragments, ranging in length from fewer than 100 base pairs to several thousand. (Remember that the total length of the lambda DNA is about 47,000 base pairs.)

We designed an experiment with lambda DNA and host-restriction enzymes that we hoped would yield two specific fragments: one containing the left operator and the other the right operator. We reasoned that since the enzymes *Hin* and *Hpa* cut lambda DNA into segments with an average length of 1,000 base pairs and since the operators are separated by some 2,000 base pairs, the enzymes should cut at least once between the operators and, of course, somewhere on either side of the operators. The result should be two fragments with one operator in each. We further reasoned that we could isolate the two fragments by mixing the digestion mixture with repressor and passing the mixture through a nitrocellulose filter. Presumably only those fragments bearing an operator would bind to the repressor molecules and would be trapped in the filter.

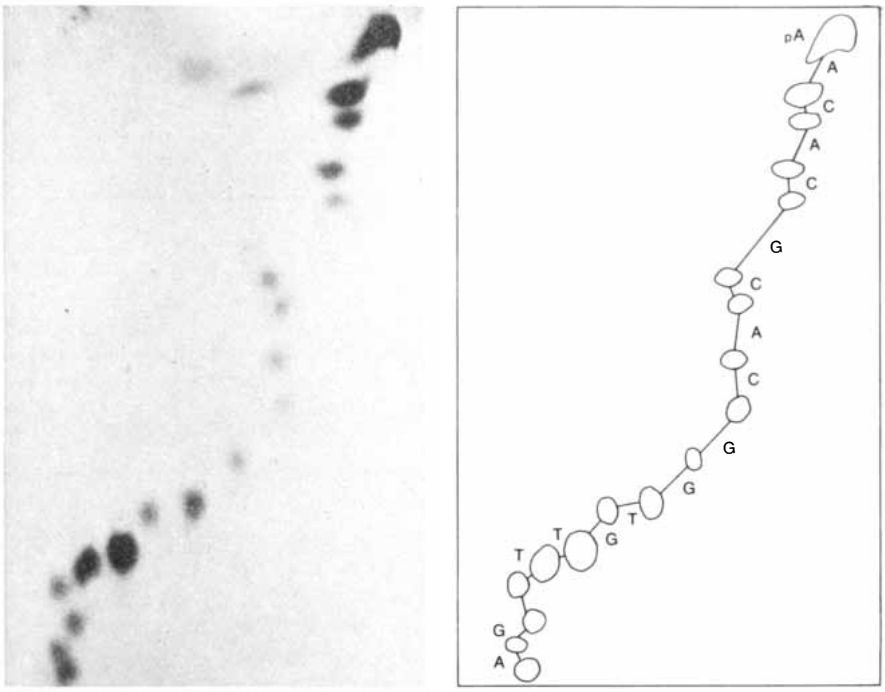
We performed this experiment using the enzyme *Hpa*. When we examined the trapped pieces of the lambda-DNA molecule, we found, as we had hoped, two specific fragments, one bearing the left operator and the other the right operator. When we repeated the experiment with the enzyme *Hin*, however, the results were strikingly different. Now we recovered not two fragments bound to repressor molecules but four. We interpreted this result to mean that *Hin*, unlike *Hpa*, cuts the lambda DNA within each operator, thereby splitting each operator into two fragments that independently bind repressor. Our conclusion implied that the largest fragment of op-

erator, 100 base pairs long, isolated in the DNAase digestion experiment would be cut by the *Hin* enzyme. That prediction was soon verified.

The experiments with host-restriction enzymes enabled us to state conclusively not only that each lambda operator has multiple repressor binding sites but also that the sites can function independently. The latter conclusion was demonstrated by the sites' ability to bind repressor even when they are separated with the cutting of the lambda DNA by *Hin*.

The Base Sequence of Operators

The ability of the lambda repressor to recognize and to bind to a number of different sites within each of the two lambda-DNA operators provides an unusual opportunity for studying the molecular basis of specific interactions between protein and DNA. With the hope of being able to identify the sequence of bases within each operator that interact with repressor, we traveled to the Medical Research Council Laboratory of Molecular Biology in Cambridge, England, in the spring of 1973. There Frederick Sanger, George Brownlee, Bart Barrell and their co-workers had developed novel methods for determining the sequence of bases in RNA molecules. Just before our visit Sanger and his co-workers reported their first success at developing methods for doing the same thing with DNA molecules. In collaboration with Barrell and John Donelson, a visiting American postdoctoral fellow, we determined the base sequence of the binding site in the left operator of the lambda DNA that had the highest affinity for the repressor protein.



DIFFERENTIAL MOBILITY OF FRAGMENTS is exploited to separate the DNA pieces containing from one base to 30 or 40 bases produced by the method described in illustration on opposite page. The mixture of fragments is first subjected to a special type of electrophoresis (*not shown*), which separates them in one dimension according to base composition, with fragments rich in *T*'s traveling fastest and fragments rich in *C*'s traveling slowest. The separated fragments are then carefully transferred to one edge of a glass plate coated with an ionic resin to be separated in a second dimension by homochromatography. The edge of the plate is dipped in a solution containing RNA molecules of assorted lengths, which compete with the DNA fragments for ionic sites. The shorter DNA fragments are bumped along faster than the longer ones. The result of the two-dimensional separation is shown in the autoradiogram at the left. Reading upward, each spot represents a DNA fragment one base shorter than the fragment responsible for the spot below it. The angle of displacement between any two spots reveals which base is present in the lower fragment but absent in the upper one and hence also reveals the sequence in which bases were removed when the strands were digested one base at a time with DNAase. Thus the letters in the drawing at the right represent sequence of bases in a central region of right operator; if bases are read downward, they represent sequence along one strand in the 5′–3′ direction. This method of determining base sequence was devised by Frederick Sanger and co-workers at Medical Research Council Laboratory of Molecular Biology in Cambridge, England.

On returning to Harvard some eight months later one of us (Maniatis) established the base sequence of most of the right operator with the help of Andrea Jeffrey, a research assistant, and Dennis Kleid, a postdoctoral fellow. The methods used to determine the base sequences can be described very briefly. First, we took advantage of the fact that Richard J. Roberts and his colleagues at the Cold Spring Harbor Laboratory had collected and characterized a substantial number of host-restriction endonucleases, samples of which they had isolated from various strains of bacteria. With these enzymes, which were generously provided by Roberts, we proceeded to dissect lambda DNA in and around the right operator. We found that *Hph*, an enzyme isolated from *Haemophilus parahaemolyticus*, neatly excises a 75-base-pair fragment that incorporates most of the right operator. As one would expect, the fragment is itself cleaved by the enzyme *Hin*. The cleavage specifically yields two DNA fragments, one of which is 45 base pairs in length and the other 30, each incorporating part of the right operator.

By combining some new methods of our own with methods developed by workers in Sanger's laboratory (including Edward Ziff and John Sedat, who were then visiting American postdoctoral fellows) we soon developed a fast and accurate technique for determining the base sequence in small lengths of DNA. Remember that DNA chains have a polarity depending on whether the backbone linkage is 5′–3′ or 3′–5′. Our technique involves the use of the enzyme polynucleotide kinase to attach a radioactive phosphorus atom to the 5′ end of each DNA chain. The chains are then separated from each other and subjected to electrophoresis under conditions such that the base composition of each chain determines its relative mobility in the electric field.

We now determine the sequence of bases along each chain with an ingenious method developed by Sanger's group. Each chain is mixed with exonuclease, a DNAase that degrades DNA one base at a time starting from the 3′ end, the end that does not bear the radioactive phosphorus atom. Samples are removed from the digestion mixture at regular intervals and the enzyme is inactivated; hence we recover all the partial products of the degradation. Starting with a chain 30 bases long we end up with as many as 30 different DNA fragments having from

SEQUENCES OF BASES have now been determined for 60-odd base pairs in the left operator of the lambda-phage DNA and 70-odd base pairs in the right operator. The presumed repressor binding sites, each 17 base pairs long, are in colored rectangles. Base sequence of the O_L3 site has not been completely determined. Binding sites are similar in base sequence; moreover, each site has considerable rotational symmetry around its central base pair. This is most readily seen by reading the base sequence on opposite strands starting from opposite ends of a binding site and proceeding toward the middle. Within each site one set of these "half-site" sequences

one to 30 bases, all labeled at the 5′ end with radioactive phosphorus.

The fragments are now separated from one another by electrophoresis; they are placed along one edge of a sheet of cellulose acetate and subjected to the electric field. Under the influence of the field the fragments migrate at a speed influenced by their base composition. *T*'s give rise to the fastest migration and *C*'s to the slowest; the overall order is *T*, *G*, *A*, *C*. Thus at the end of the electrophoresis molecules rich in *T*'s, regardless of their size, will have traveled farther than those rich in *G*'s, and so on.

The distributed molecules are carefully transferred to one edge of a glass slide coated with a thin layer of ionic resin. There they are further separated along an axis at right angles to the electrophoresis axis by the procedure called homochromatography. The plate, clamped vertically with the fractionated DNA fragments lined up along its lower edge, is dipped in a solution containing RNA molecules of various lengths. As the RNA molecules compete with the DNA molecules for sites in the layer of ionic resin, the DNA molecules are displaced upward; the shorter the DNA molecule is, the farther it travels. The final position of each DNA fragment is revealed on photographic film by autoradiography.

Since the fragments are labeled with radioactive phosphorus, they appear as dark spots running up the film. Remarkably, from the angle of displacement of each fragment from its neighbor one can deduce the identity of the base removed by each step in the DNAase digestion [*see illustration on preceding page*]. By using both chains of the double-chain DNA fragment, each labeled at its 5′ end, we can determine the probable sequence of 30 to 40 base pairs quite quickly. Although the sequence assignments made in this way are not completely reliable, they can be verified and the ambiguities can be resolved by further manipulations. The important general point is that these methods enable us to determine the base sequence of any double DNA chain up to about 40 base pairs in length.

Our efforts in Cambridge and at Harvard have now yielded the base sequence of a large portion of the left and right operators of phage-lambda DNA [*see illustration above*]. The sequence of the right operator was also determined by Pirrotta at the University of Basel, who used a method different from ours. It is gratifying that the two approaches yield the same sequence.

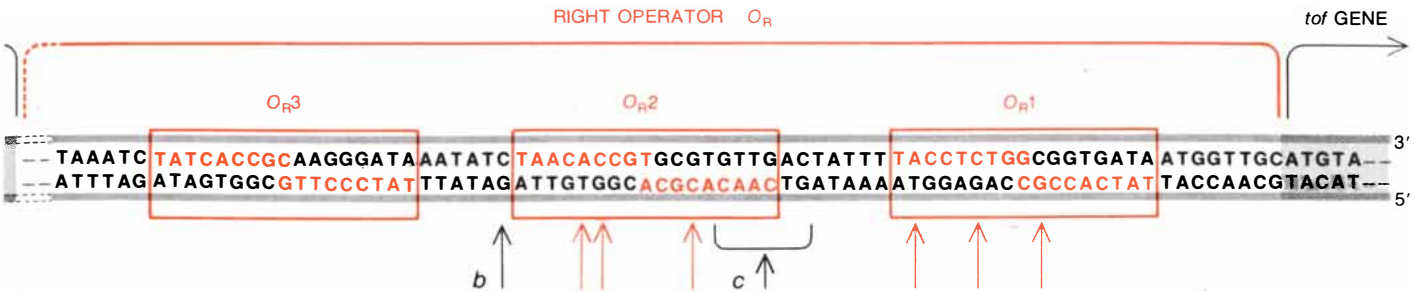
What do we learn by determining the sequence of bases that constitutes the lambda operators? As was first noticed by Keith C. Backman, a graduate student working with us at Harvard, there are base sequences in both operators that are strikingly similar. Presumably they are the sites recognized by the repressor. Each of the sites is exactly 17 base pairs long; moreover, they are separated from one another by "spacers," strings of three to seven bases that contain only, or nearly only, the base pair *AT*.

Mutations in Binding Sites

One of the strongest indications that the closely related 17-base-pair sequences are indeed the sites recognized by repressor comes from a study of mutations that change a base pair in the operator region. Kleid, Zafri Humayun and Stuart M. Flashman, working with one of us (Ptashne), have now determined the sequence of 12 such mutants, many of which were selected and studied by Flashman. Ten of the mutants change the sequence of bases in the sites we have called O_L1, O_L2, O_R1 and O_R2; all of them decrease the affinity of that portion of the operator for repressor. Two mutations located in two different spacer regions, and a third probably located in another spacer, have no effect on repressor binding. Instead they drastically decrease the efficiency with which the enzyme RNA polymerase initiates transcription of the adjacent genes.

An interesting feature of the 17-base-pair repressor binding sites is that each has a partial internal symmetry. What is meant by this is as follows. One reads off the sequence of bases on opposite strands of the binding site starting at opposite ends and proceeding toward the middle. One finds that the two sequences of eight bases on either side of the central base pair are more similar than would be expected if the sequences were random. This partial twofold rotational symmetry, as it can be called, is most apparent in the site O_L1, where six of the eight positions are occupied by the identical base. If the sequence were random, one would expect only two of the eight bases to be the same. The other 17-base-pair binding sites are also more symmetrical than would be expected by chance, although they are less symmetrical than the O_L1 site. That site, perhaps significantly, is also the site with the highest affinity for repressor.

One way to compare the various repressor binding sequences with one another is to consider only the "half-site" sequences obtained by reading the opposing chains in the 5′–3′ direction, beginning at the ends of each 17-base sequence and proceeding toward the middle. If any site were perfectly symmetrical, the two half-site sequences compared in this way would be identical. In fact, certain symmetrically arranged positions in every half-site are identical [*see top illustration on page 76*]. For example, at position No. 2 we always find the base *A;* at positions No. 4 and No. 6 we always find the base *C*. Positions No. 5 and No. 8 are usually occupied by an *A* and a *G* respectively. We infer that the identity of the bases at these positions cannot be changed without strongly af-



is identified by letters in color. Nine colored arrows identify base pairs within repressor binding sites that have been altered in mutant lambda DNA. In every case the mutation abolishes the affinity of the repressor for that binding site. Three black arrows (*a*, *b*, *c*) show location of mutations that do not affect repressor binding but do interfere markedly with the ability of RNA polymerase to recognize promoter regions for gene *N* and genes *c*I and *tof* respectively. (Exact location of the mutation indicated by the arrow *c* is uncertain.) The base sequence of the *c*I gene, which codes for the repressor, has not been determined; it is about 1,000 base pairs long.

fecting the affinity of the site for repressor. Evidently the requirements of the other positions are not so stringent, although the favored sequence is *TATCACCG*, with either *C* or *G* as the ninth base (that is, the middle base of the 17-base sequence). The 10 operator mutations that have been found, each of which abolishes affinity for repressor, are located at positions No. 2, No. 5, No. 6 and No. 8.

Is there any reason to expect that the base sequences recognized by repressor would be symmetrical? Some years ago Walter Gierer of the Max Planck Institute for Biophysical Chemistry in Göttingen suggested that symmetrical sequences in DNA might form looped structures and that the loops might be recognized by repressors. We now have good reason to believe repressors bind to DNA in linear form; thus we prefer to think that symmetry of the operator reflects a symmetry in the structure of the repressor. As we have mentioned, the lambda repressor binds to DNA as a dimer or tetramer made up of identical subunits. On the basis of general considerations of protein architecture we suspect the repressor is largely, if not entirely, symmetrical; therefore we would not be surprised to find symmetry in the operator to which it binds.

As was first noted by Hamilton O. Smith and Kent W. Wilcox of Johns Hopkins University, the sites recognized by some restriction endonucleases, although much shorter than those recognized by repressors, also have a twofold symmetry. Moreover, as Gilbert, Maxam and their colleagues have shown, the operator recognized by a different repressor, the *lac* repressor, is highly symmetrical. (The *lac* repressor turns off the genes in *E. coli* that specify enzymes for the utilization of the sugar lactose.)

Our analysis of the lambda operators has emphasized the internal symmetries in the repressor binding sites. It must be said, however, that there is no direct evidence that the repressor actually interacts with the DNA symmetrically. Moreover, we know of cases where proteins recognize DNA sequences without there being any apparent symmetries. In fact, there may be a systematic asymmetry in the various lambda-repressor binding sites. Each 17-base-pair site has on one side the sequence *TATCACCGC* or a sequence closely related to it, together with another half-site where the sequence is more variable. It is possible that repressor protein bound to its operator DNA is somewhat deformed and not perfectly symmetrical.

An appraisal of the role of symmetry in operator-repressor interaction may call for the chemical synthesis of various operator sequences and a study of their interactions with repressor. In any case we believe each repressor will recognize its own favored DNA sequence; we have no reason to believe the sequences of operators recognized by different repressors will be similar to the sequences of the lambda operators. In particular, the base sequence of the *lac* operator, the only other operator whose sequence is known, bears no obvious relation to the sequences of the lambda operators. Moreover, we cannot yet describe in any detail how the repressors recognize their operator targets in DNA. For example, the 17 base pairs of a binding site occupy nearly two full turns of the double-chain DNA helix. One can only guess at what the repressor protein "sees." Perhaps some simple rules govern the recognition of base sequences by proteins, but if they do, they remain to be discovered.

How the Repressor Works

Our experiments with lambda repressors and operators, taken together with the results from several other laboratories, have revealed how the repressor turns off its target genes. At the University of Basel, Alfred Walz and Pirrotta have shown that in the absence of repressor the enzyme RNA polymerase binds tightly to, and protects from DNAase digestion, a 45-base-pair sequence that includes most of one of the repressor binding sites. The same sequence includes about 20 base pairs of the beginning of the gene the RNA polymerase transcribes into messenger RNA. Many other cases have been found where RNA polymerase covers about 20 base pairs on either side of the beginning of a gene. (The covered region does not include all the bases required for RNA-polymerase recognition because, as was found by Russell A. Maurer, a graduate student working with us, and by others, mutations some seven bases to the left of that region severely impede the action of RNA polymerase. The entire DNA region required for RNA-polymerase binding is called the promoter, and its exact extent is not known.) RNA polymerase apparently recognizes some aspect of the base sequence near the beginning of many genes. When the enzyme is supplied with the appropriate substances on which to act, it copies one chain of the DNA sequence into messenger RNA.

The RNA-polymerase molecule is some 20 times heavier than the repressor molecule and is therefore several times larger in volume. From the DNA region covered by RNA polymerase one can see that the enzyme competes with the repressor for its binding site. Therefore when repressor is bound to the operator, RNA polymerase cannot bind to the promoter region. The same effect is seen with the *lac* repressor.

Why does each lambda operator have more than one repressor binding site? The *lac* operator has only one repressor binding site, so that multiple binding sites do not seem to be the general rule. One possibility is that multiple sites allow for graded control. For example, if

	WILD-TYPE SITES									MUTANT BASES
O_L1	T	A	T	C	A	C	C	G	C	A OR T
	T	A	C	C	A	C	T	G	G	G
O_R1	T	A	T	C	A	C	C	G	C	T
	T	A	C	C	T	C	T	G	G	G, A
O_L2	C	A	A	C	A	C	C	G	C	T
	T	A	T	C	T	C	T	G	G	
O_R2	C	A	A	C	A	C	G	C	A	A
	T	A	A	C	A	C	C	G	T	G, A
O_R3	T	A	T	C	C	C	T	T	G	
	T	A	T	C	A	C	C	G	C	
O_L3	T	A	T	C	A	C	C	G	C	
	T_9	A_{11}	T_6	C_{11}	A_8	C_{11}	C_6	G_9	C_5	
	C_2		A_3		T_2		T_4	C_1	G_4	
			C_2		C_1		G_1	T_1	A_1	
									T_1	

OPERATOR HALF-SITES are compared to reveal their similarity. The 11 sequences are those designated by colored letters in illustration on preceding two pages. Each sequence is written left to right in the 5′–3′ direction. Letters in color at left indicate nine sites where 10 different mutations have been found. Mutant bases that replace normal bases and abolish repressor affinity are listed in the column at the right. Tabulation at lower left summarizes frequency with which each base appears at each position in the 11 half-sites. One can see that the second, fourth and sixth positions are invariant.

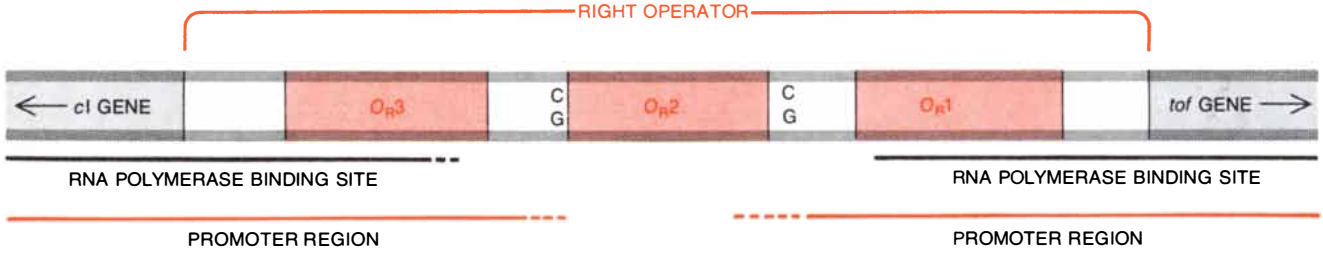
only one operator site is occupied by repressor, some gene transcription might occur, whereas if all sites were filled, transcription would be abolished. Flashman and Barbara Meyer, another graduate student, have recently shown that the maximum repression of transcription of the *tof* gene requires that repressor be bound to both O_R1 and O_R2. Thus mutation of either O_R1 or O_R2 decreases the effect of repressor on the transcription of *tof*, and mutation of both sites has a stronger effect. Significantly, two mutations at one site do not have as strong an effect as two mutations at two sites. Similar experiments show that the maximum repression of gene *N* requires repressor bound to both O_L1 and O_L2. We imagine that occasional unnecessary expression of the *lac* genes is not harmful to the cell, whereas occasional expression of the lambda genes could be lethal. We may therefore speculate that the lambda system, involving multiple repressor binding sites at each operator, has evolved stricter controls than the *lac* system, involving only a single repressor binding site.

Recently Meyer has discovered a most remarkable function for O_R3: the third repressor binding site in the right operator. She found that RNA polymerase recognizes not one promoter region in the right operator but two regions. Moving to the right from one promoter region, RNA polymerase transcribes the *tof* gene. Moving to the left from the second promoter region, it transcribes the *c*I gene, the gene that codes for the repressor. Although the exact starting point for the transcription of the *c*I gene is not known, it is probably just to the left of O_R3. We have mentioned the fact that transcription of *tof* begins just to the right of O_R1. Therefore we have two genes, *c*I and *tof*, transcribed in opposite directions and separated by three repressor binding sites.

Meyer has found that repressor not only turns off the transcription of *tof* but also turns off transcription of *c*I. In other words, repressor regulates its own level in the cell. From analysis of the effects of mutations in the right operator we deduce that repressor turns off the transcription of *c*I primarily by binding to O_R3. Because repressor binds more weakly to O_R3 than to O_R1 and O_R2 it allows a higher level of *c*I transcription (and hence a higher level of repressor protein) than it does of *tof* transcription [*see illustration below*].

Even this description of the role of the reiterated repressor binding sites is probably incomplete. For example, Meyer has found that relatively large amounts of repressor turn off the transcription of both *tof* and *c*I, whereas smaller amounts suffice to turn off the transcription of *tof*. On the other hand, the smaller amounts of repressor actually enhance the transcription of *c*I, the repressor's own gene. This positive effect of repressor on the transcription of its own gene had been predicted by work on whole bacterial cells infected with bacteriophage lambda; the work was done by, among others, Louis Reichardt, who is now at the Harvard Medical School. We do not yet know what molecular mechanisms are involved.

We suspect that as investigations progress even more sophisticated roles in gene regulation will be assigned to the system we have been describing. We have described some features of the interaction of two proteins, lambda repressor and RNA polymerase, with the sequences in and around the lambda operators. We know, however, that at least two other proteins, the products of genes *N* and *tof*, are themselves regulatory proteins that almost certainly recognize sequences in the operators. Although our understanding is far from complete, these studies have begun to reveal how complex patterns of gene regulation can be described in terms of specific interactions between proteins and DNA.



RNA-POLYMERASE BINDING SITES overlap at one end of the left lambda operator and both ends of the right operator. For simplicity only the right operator is shown. Repressors produced by the *c*I gene bind preferentially to O_R1 and O_R2, blocking expression of the *tof* gene. As long as no repressor is bound to O_R3, RNA-polymerase molecules have access to the binding site next to the *c*I gene and keep generating more repressors. (The exact location of the RNA-polymerase binding site next to the *c*I gene is not known.) Eventually, however, O_R3 is occupied by repressors and access to the *c*I gene is blocked. The ability of RNA polymerase to bind to lambda DNA is influenced by the integrity of bases lying somewhat outside the binding site. Thus mutations in the base pair *CG* to the left of O_R2 make it impossible for polymerase to transcribe the *c*I gene. Location of this gene within right operator was first noted by Gerald Smith and his colleagues at University of Geneva. A similar mutation in a *CG* pair between O_L1 and O_L2 (*not shown*) prevents transcription of *N*. A mutation believed to affect a *CG* pair to right of O_R2 blocks transcription of *tof*. Total sequence of bases required for polymerase recognition is termed a promoter; exactly how many bases are involved is not established.

We're drilling for plastic.

When somebody mentions crude oil and natural gas, most people think of gasoline, motor oil, and fuel for home and industry.

Diamond Shamrock sees more than these products in a barrel of crude oil or a thousand cubic feet of natural gas. Because crude oil and natural gas are also raw materials in the production of chemical feedstocks. At our petroleum manufacturing complex in the Texas Panhandle, facilities are now under construction to furnish these vital feedstocks for our future growth in plastics, and many other products.

Plastics used to make building and construction materials, home furnishings and household goods, consumer goods such as rainwear, sporting goods and toys, automobile and transportation equipment,

medical goods, packaging for foods, hard and soft goods, and a multitude of other things.

Long before most people were aware of the impending energy shortage, Diamond Shamrock significantly increased its budget for domestic oil and gas exploration and production. Today, we have expanded our net leasehold acreage by 52%, and brought important new sources of crude oil and natural gas into production.

Offshore in the Gulf, Diamond Shamrock owns interests in 54 tracts, covering 244,300 acres of Federal leases offshore Texas and Louisiana. Natural gas is flowing from 34 wells drilled from two platforms and drilling operations are underway from two others. Six other platforms are under construction. In shallower state waters offshore Texas, Diamond Shamrock owns major interests in two production platforms already in operation.

Onshore Diamond Shamrock has assumed a leading position in the Rocky Mountain area, principally in Wyoming.

We've stepped up development of gas fields in the Anadarko Basin of the Texas Panhandle and we are actively evaluating promising areas in Alabama, Arkansas, Florida, Kansas, Louisiana, Montana and Texas. Long-range, we are preparing for participation in such frontiers as offshore Atlantic, the Gulf of Alaska, Bristol Bay, and other domestic areas.

In addition to gasoline sold through branded service stations in 10 Southwestern and Rocky Mountain States, other Diamond Shamrock petroleum products include: diesel fuel, lubricants, jet-turbine fuel, liquefied petroleum gases, fertilizers, asphalt and sulphur.

And when our expanded facilities in the Texas Panhandle are completed in 1976, feedstock from oil and gas operations will be going to our chemical plants in the Houston area.

For more information about Diamond Shamrock, the international corporation that's turning industries around with chemistry and energy, write: Diamond Shamrock, 1100 Superior Avenue, Cleveland, Ohio 44114.

Because when we drill for oil and gas, we're drilling for more than just gasoline, motor oil, or heating fuel.

Diamond Shamrock

Trade your hog

You loved that car of yours when it was new, didn't you?

All bright and shiny and everything?

How could you know the world would change? How could you know it would use too much gas, too much space, too much everything?

How could you know that there would be a VW Rabbit?

We knew 5 years ago. That's when we started—from absolute scratch—to design the one and only car that honestly makes any sense today.

It's almost as if we re-invented the wheel. We re-thought every detail, every old concept of car design to get to the new concept of the Rabbit.

Even the key is padded for safety.

For example, size:

The Rabbit is a big car. Not to look at. Outside, it parks in almost no space.

But inside, there is more glass area than in a Lincoln Continental Mark IV; as much legroom and headroom as some mid-sized cars; and (with the rear seat down) more luggage space than a Cadillac Fleetwood.

Seat belts put themselves on.

Quite an achievement.

Another example: safety.

The Rabbit helps make you a safer driver.

Front-wheel drive gives you better tracking.

Rack-and-pinion steering gives you better handling.

If one front tire blows, negative steering roll radius helps you to a straight stop.

If one brake circuit fails, a second circuit is still there.

A totally new rear axle makes the car more stable (and safer) on rough roads.

The deluxe Rabbit also has an utterly unique safety belt that puts itself on as you sit down.

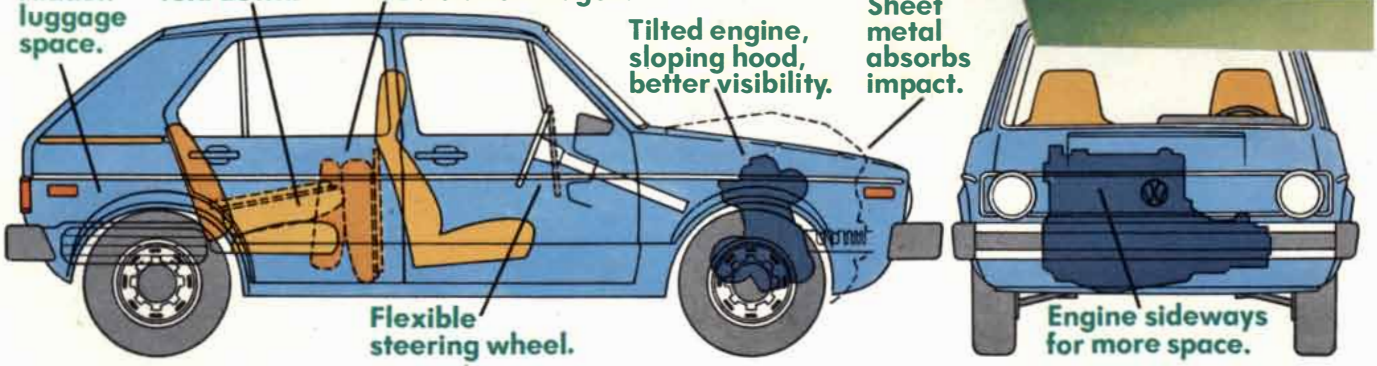
Still another example: economy and performance.

The EPA estimates that the Rabbit with stick shift got 39 mpg on the highway, 25 mpg in the city.

(Actual mileage may vary, depending on type of driving, driving habits, car's condition and optional equipment.)

Economy alone is a major accomplishment. But economy plus great performance is close to a miracle.

We've done it.



for a Rabbit.

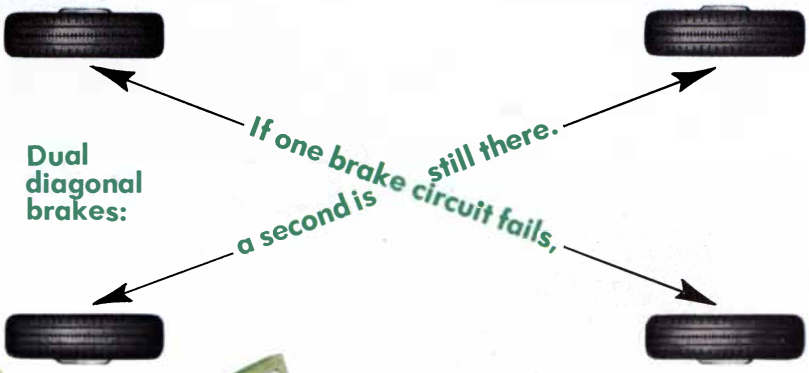
0-50 mph in 8.2 seconds. No other car gives you this kind of power with so much economy.

The Rabbit zips from 0 to 50 mph in a mere 8.2 seconds. A $21,000 Maserati doesn't do much better.

No other car gives you this kind of power with so much economy.

There are more than 80,000 VW Rabbits hopping around the U.S. already.

It is the most successful new foreign car introduction in history.



The Amazing Rabbit

Our biggest VW sale was to Detroit's Big Three–56 Rabbits. We don't think it was a gesture of friendship or because they liked the colors.

They are tearing them to pieces to find out how we did it. We did it the hard way–from scratch.

But it was worth it.

The Rabbit is a happy car, and we're happy with it.

At first, we were worried about the future of the car.

Now we've come up with the car of the future.

SPEARING STOMATOPOD of the species *Squilla empusa* is found in the shallow waters off the coast of the southeastern U.S. The species is a major predator of shrimp. Specimen in this photograph is 11 centimeters (about four and a quarter inches) long.

SMASHING STOMATOPOD of the species *Odontodactylus scyllarus* is found in the Indian Ocean and the Pacific. The specimen shown here is 17 centimeters (about six and a half inches) long. This colorful species is commonly imported for display in marine aquariums. The smashing strike of this stomatopod's raptorial appendages has sufficient force to break the glass of an aquarium.

STOMATOPODS

These shrimplike creatures are formidable predators. Two of their forward appendages are adapted either for spearing or for smashing. One blow from a smasher can break the shell of a good-sized clam

by Roy L. Caldwell and Hugh Dingle

Stomatopods, commonly known as mantis shrimps, are predatory crustaceans that live in the shallow waters of tropical and subtropical seas. They dwell in burrows and cavities and feed on other marine animals such as annelid worms, snails, clams, shrimps, crabs and fishes. Although stomatopods can be so numerous that in some waters they are a major predator, they are seldom seen because of their retiring habits. The most an exploring diver is likely to see of a stomatopod is a pair of eyestalks peering out from a mud burrow or a cavity in rock or coral. Retiring though stomatopods are, they are also extremely aggressive, both in capturing prey and in defending their territory; as a group they are among the most pugnacious animals known. The most notable feature of a stomatopod is the two large forelimbs it uses to attack its prey. These raptorial appendages unfold and shoot forward much like the forelimbs of the praying mantis. They are adapted either for spearing or for smashing, and in the evolution of stomatopods the structure and function of these two different types of appendage have interacted closely with other aspects of the animals' morphology and with their patterns of behavior.

Stomatopods are not closely related to shrimps or other marine crustaceans. They appear to have split off from the leptostracan stock some 400 million years ago. Originally they were filter feeders, using their thoracic appendages as food strainers. Fossil evidence indicates that some 200 million years ago one forward pair of appendages, the second maxillipeds, were evolving into large folded limbs. ("Stomatopod" is from the Latin for "mouth-foot.") These animals were the first true stomatopods. During the Jurassic period, from 190 million to 135 million years ago, the stomatopods radiated into the four modern families: the Squillidae, the Lysiosquillidae, the Gonodactylidae and the Bathysquillidae. The Bathysquillidae are found in deep water and have rarely been taken alive, so that nothing is known about their behavior. The remaining three families are found in shallow waters, and we have collected data on more than 40 species distributed around the world.

There are currently about 350 known species of stomatopods, and additional species are being discovered every year. The species range in length from 15 to 335 millimeters (half an inch to 13 inches), and many of them are brightly colored. They can be divided into two functional groups: the spearers and the smashers. The spearers consist of all the Squillidae, Lysiosquillidae, Bathysquillidae and several genera of the Gonodactylidae. The smashers consist of only a few genera of the Gonodactylidae.

Among the spearers the last joint of the forelimb is equipped with sharp spines, from three to 17 of them. This spined dactyl is used to spear soft-bodied prey such as fishes and shrimps. The strike is made with the dactyl open and is one of the fastest animal movements known. It is completed in from four to eight milliseconds, and the velocity of the movement is more than 1,000 centimeters per second even under water. Once the prey is stabbed it is pulled toward the stomatopod's mouthparts and ripped to pieces by its sharp, serrated mandibles and by its third, fourth and fifth maxillipeds, which are equipped with hooks. The spines of the dactyl are barbed at the tip, so that the stabbed prey is unlikely to slip off. In many species the second joint of the raptorial appendage also has spines, which help to secure the prey when the dactyl is folded.

Most spearers dig their burrows in mud or sand and wait in ambush at the entrance. For example, *Squilla empusa,* a stomatopod that is common off the coast of the southeastern U.S., digs a hole or trench and often lies in wait covered with silt so that only its eyestalks are exposed. The species is abundant in commercial shrimp beds and is believed to be a serious predator of shrimps.

Many of the larger spearers that feed on fishes have greatly elongated raptorial appendages, which give them a reach of half their body length. *Harpiosquilla harpax,* a 250-millimeter (10-inch) stomatopod of the Indian Ocean, can strike 130 millimeters. We have seen this species catch a 110-millimeter fish and devour it in four minutes.

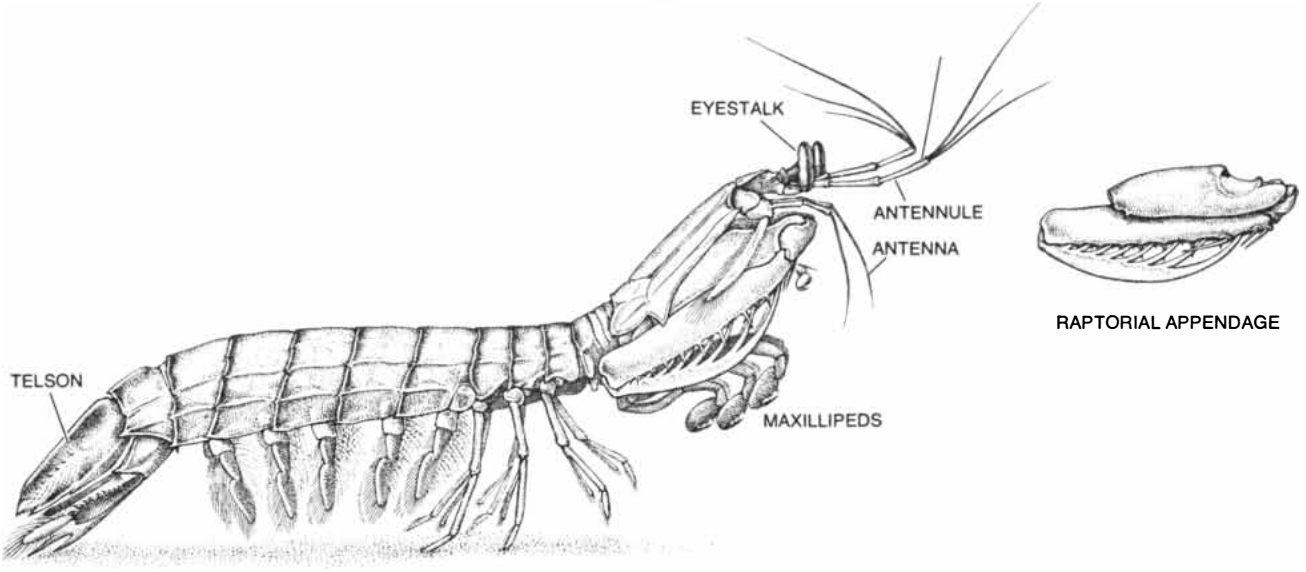
The smashing stomatopods have a dactyl with few spines or none, but in most of them the heel of the dactyl is greatly enlarged. During the strike the dactyl remains folded, and the prey is struck by the blunt heel. These stomatopods normally feed on armored animals such as snails, hermit crabs, clams and crabs, which they batter to pieces. The strike of a large species of smasher such as *Hemisquilla ensigera* (250 millimeters long) has a force approaching that of a small-caliber bullet. A large specimen of this species from southern California broke an aquarium wall consisting of a double layer of safety glass. Even small species 80 millimeters long can break the wall of an ordinary glass aquarium.

Most smashers live in cavities in rock or coral. They visually select their prey and stalk it. When a stomatopod attacks a crab from the rear, its first blow stuns the crab; subsequent blows knock off the crab's legs and claws and smash its carapace. When the stomatopod attacks a

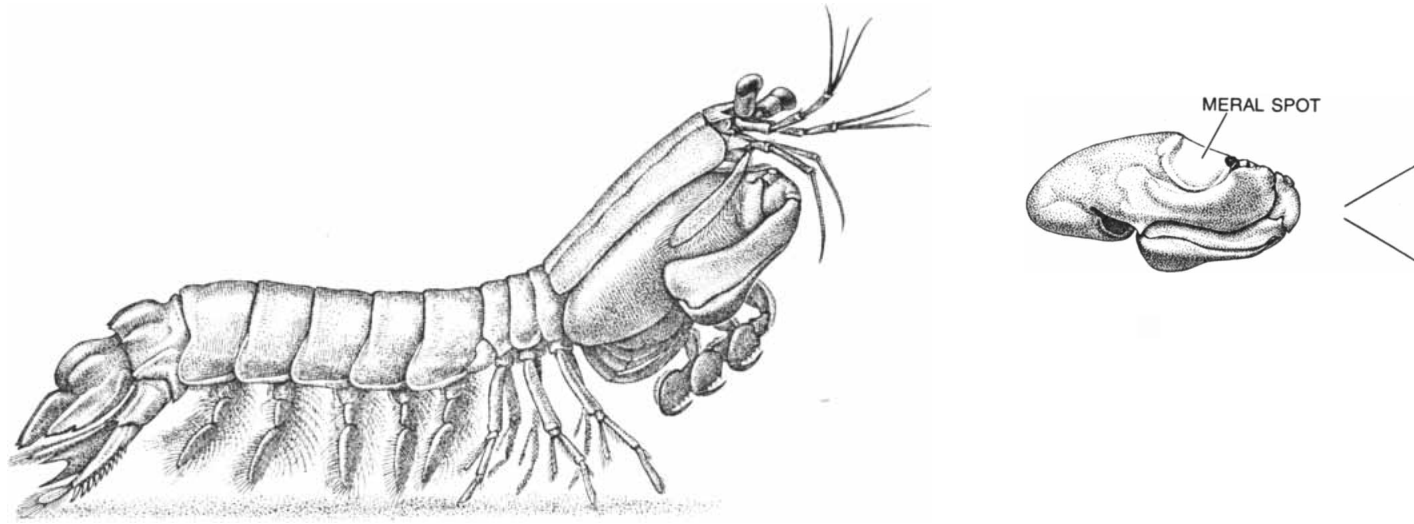
crab from the front, its first blows are usually directed at the crab's claws. When the claws are broken, further blows crush the crab's carapace and break its legs. The stomatopod then drags the battered hulk to its dwelling cavity, where it breaks the crab into small pieces and picks the tissue from the exoskeleton. Snails, clams and hermit crabs living in snail shells are simply seized and taken to the cavity. The stomatopod wedges the prey against the wall with its third, fourth and fifth maxillipeds and repeatedly strikes the shell with its raptorial appendage until enough of the shell is broken away for it to pull out the soft tissue.

The feeding efficiency of an intertidal species, *Gonodactylus chiragra*, was studied in Thailand by one of us (Caldwell). The species lives in rock cavities and forages over the reef flat. After it has captured and eaten its prey, it dumps the broken shells outside its dwelling cavity. The broken shells made it easy to identify the various items in the animal's diet. These stomatopods normally ate snails, hermit crabs and clams. Some of them were captured and were offered different sizes of snail in the laboratory. Although they were quite capable of breaking the shells of snails up to 30 millimeters long, they preferred smaller snails. They also preferred snails to hermit crabs that lived in shells of the same size. Most of all these stomatopods preferred clams; it takes only a few strikes to break open the shell, and the amount of edible tissue obtained for the effort is much greater than it is with other prey animals.

The number of strikes the stomatopods needed to break open the shell of snails and hermit crabs of various sizes was recorded, together with the amount of tissue they got out of the shells. With this information the amount of edible tissue the stomatopod could get per strike with prey of various kinds and sizes could be calculated. The results showed that the stomatopods were



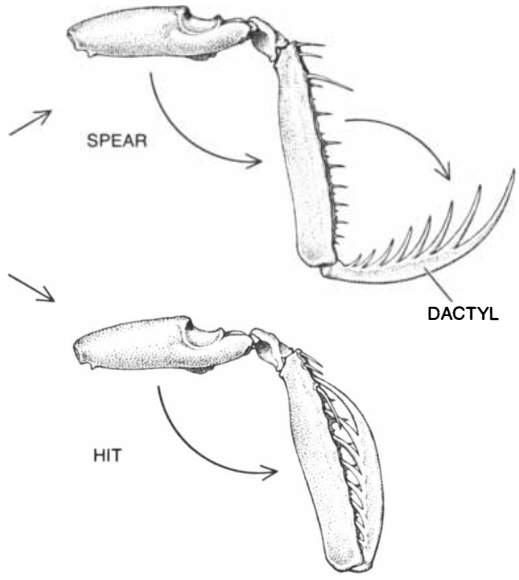
SPEARER of the species *Harpiosquilla harpax* is shown in a typical attitude. This species can grow to a length of 25 centimeters (10 inches). Like all spearers, it has two modes of striking with its raptorial appendages (*right*). In capturing prey it uses a spearing



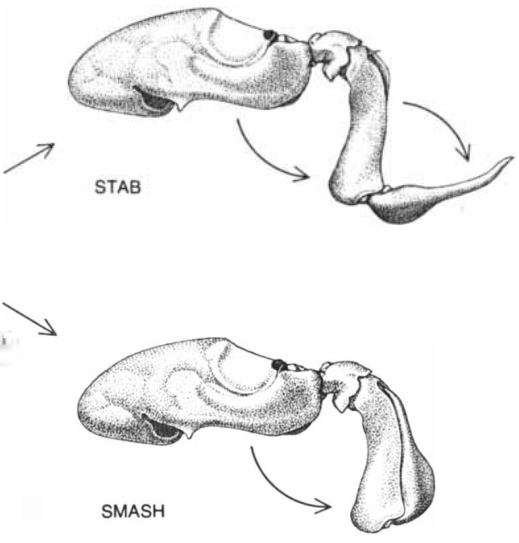
SMASHER of the species *Gonodactylus chiragra* can grow to a length of about 10 centimeters. As do all smashers, it uses its raptorial appendages both in combat with other stomatopods and in attacking prey. If the prey is soft-bodied, the smasher will change

choosing the prey that gave them the greatest amount of tissue per strike.

Another interesting observation made in the course of the study was that individual stomatopods show considerable plasticity in learning how to open various kinds of shell. Each animal developed its own style of attacking shells and used a favorite rock as an anvil. Some would always break the apex of the shell first, others would break the lip and still others would shear off the whorls on one side. Studies of other species of smashers have shown that when they are given novel food items, they are at first inefficient in breaking open the shell but with experience their efficiency improves.



strike. In combat with another stomatopod spearers usually hit with the dactyls closed.



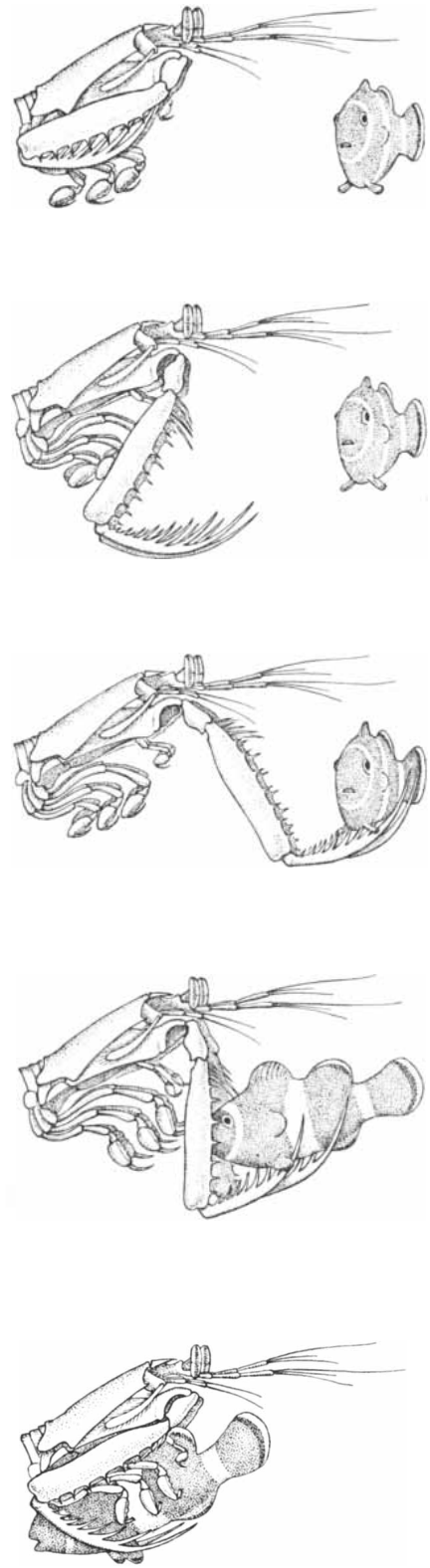
from a smashing strike to a stabbing one. It also sometimes stabs in intense combat.

The power of the smashing strike is so strong that the smasher could conceivably break off a dactyl. Some newly captured smashers are missing one raptorial appendage or both appendages, or have a dactyl with a broken heel. Such damage is not, however, the result of feeding activity but of fights with other stomatopods. When a raptorial appendage of a stomatopod is severely damaged, the animal cannot simply shed the limb, as a crab can. It literally tears the limb off with its maxillipeds, and a new appendage regenerates over the next two or three molts. Intact dactyls show surprisingly little wear. Although a smasher may make hundreds of strikes per day at hard snail shells, the heels of its dactyls show very little erosion, even over a period of several months between molts.

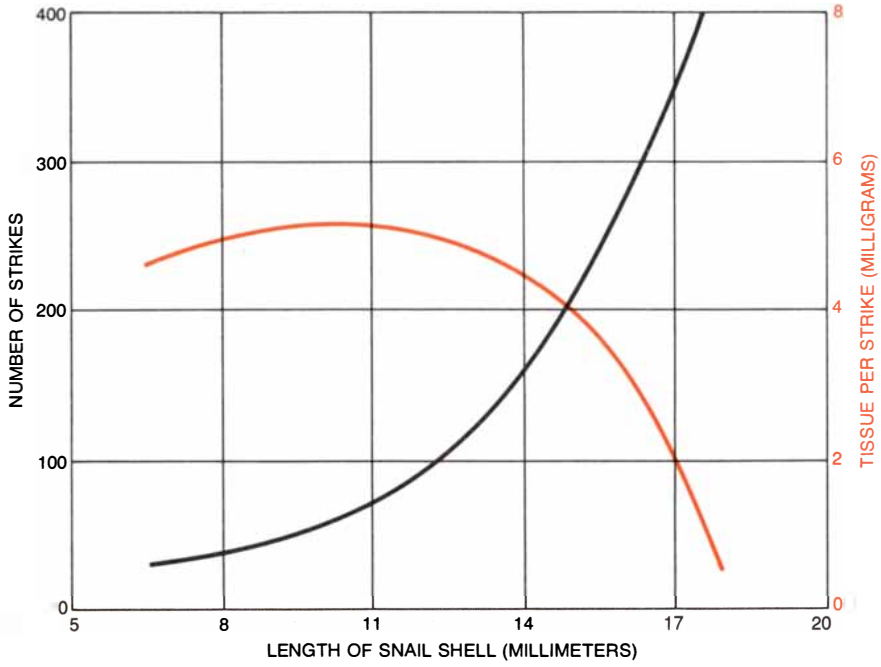
The tip of the smasher's dactyl, as distinct from the heel, is sharp but not barbed like the spined dactyl of spearers. On occasion a smasher will strike its prey with its dactyl tip open, usually after the prey has been struck at least once by a smashing blow. That can happen if one holds a stomatopod in one's bare hand, a point that has been painfully driven home to us. Apparently the animal gets some feedback information about the softness of the target and changes its mode of attack.

Most stomatopods are territorial and vigorously defend their dwelling space. The strike of a stomatopod is potentially lethal to other stomatopods and is usually somewhat modified when it is directed against a potential competitor of the same species or a closely related one. Spearers normally strike at another stomatopod with a closed dactyl, which is much less likely to inflict serious injury than the piercing strike of an open dactyl. Smashers also attack an intruder with a closed dactyl, but they normally direct the blow at the intruder's body armor and thus rarely inflict a fatal injury. Under certain conditions, however, many smashers will stab an intruder with an open dactyl. If blood is drawn during a fight, some smashers will try to stab and kill their opponent.

When two smashers are fighting, most of the blows are delivered to the telson, or tail shell, which is heavily armored. Spearers do not have a powerful smashing strike, and in combat they direct their blows at all parts of their opponent's body. The telson of spearers is more fragile, even though in some species it is armed with sharp spines. The telson of smashers has numerous large ridges with heavy, blunt projections instead of sharp spines. The ridges appar-



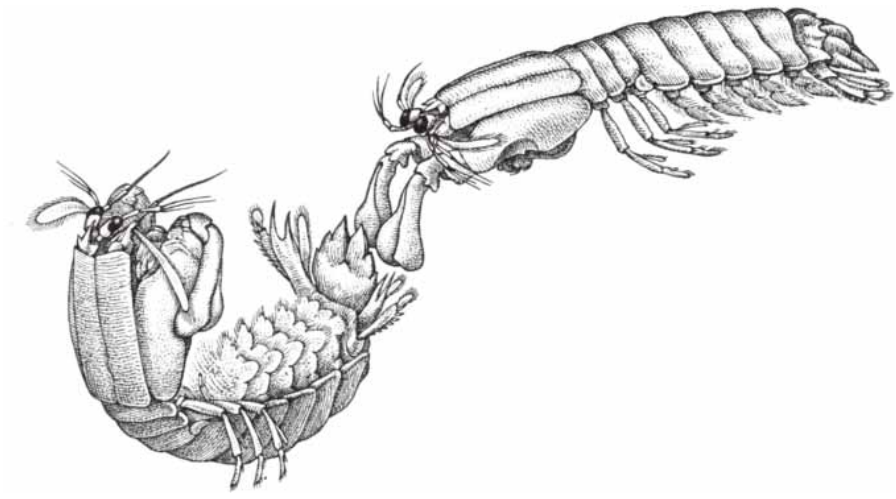
RAPTORIAL STRIKE of a spearer is depicted. When the prey, in this instance a fish, swims into range, the stomatopod strikes with both raptorial appendages. The strike is one of the fastest animal movements known, attaining a velocity of more than 1,000 centimeters per second. Spines on the second joint of the raptorial appendage secure the prey when the dactyl is folded. The captured fish is pulled to the mouthparts and ripped apart by hooks on the maxillipeds.



FEEDING EFFICIENCY of a five-centimeter *Gonodactylus chiragra* is shown in terms of the number of strikes it requires to completely break open snail shells of various sizes and the amount of tissue it obtains per strike. The stomatopod strikes on the average once every 30 seconds and can take three hours to break open larger shells. When offered snails of various sizes, *G. chiragra* selects the sizes that give it the highest amount of tissue per strike.

ently give the telson structural support, enabling it to withstand crushing blows. A survey of the telson morphology of various species of stomatopods shows that as the crushing power of the strike increases there is a concomitant increase in the armoring of the telson.

Smashers that have evolved heavy telson armor employ a unique defensive behavior when they are fighting. The animal being attacked lies on its back and coils its telson upward, so that blows are taken on the telson's outer surface. When there is an opening for a counterattack, the animal momentarily lowers the telson, lunges forward and strikes its opponent, which has coiled in response so that the blows fall on its own telson. After the attacking stomatopod has struck a blow, it quickly resumes its coiled posture. The armor of the telson, together with the shock-absorbing action of the coiled abdomen, minimizes physical damage.



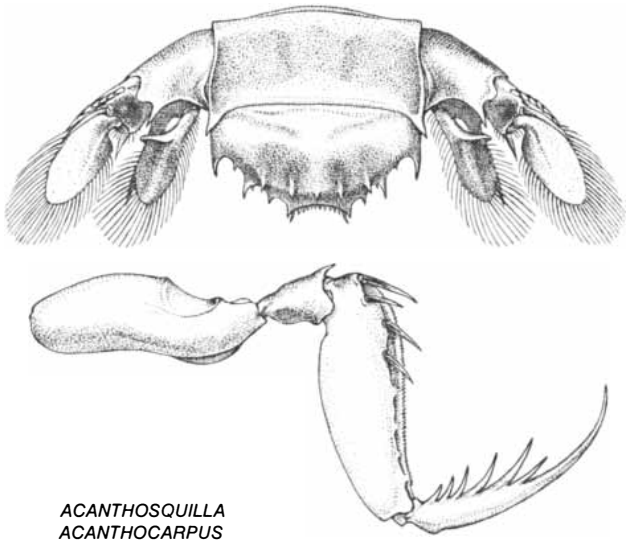
COILED DEFENSIVE POSTURE of the smasher is shown at the left. The stomatopod lies on its back and presents its armored telson to the attacker, which has just uncoiled and is lunging to strike. After it has struck a blow it will recoil and the opponent will lower its telson, lunge forward and strike. This ritualized fighting minimizes physical damage.

A few species of smashers, for example *Gonodactylus chiragra,* engage in such ritualistic combat less frequently than others. Instead the animal tries to circle around its opponent and kill it by striking at the head or the thorax. The killing species tend to be found in areas where there is intense competition for suitable cavities. Their unrestrained attack behavior gives them a competitive edge over species that use the ritualized coiling form of combat, particularly since they are the largest and most powerful species in the community.

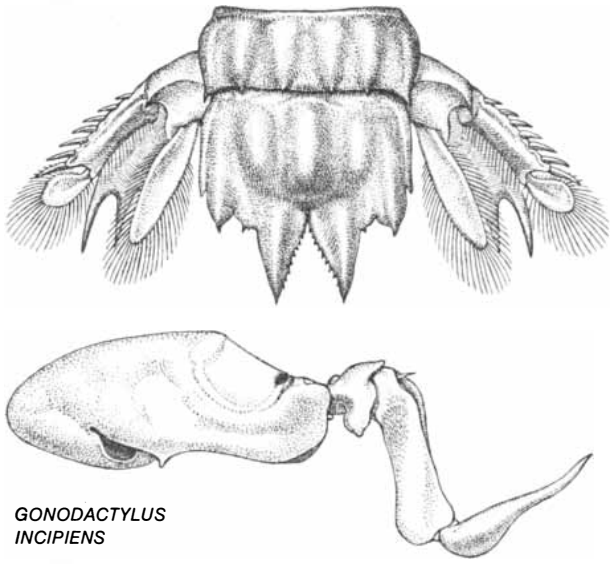
Smashers also use their armored telson to defend their domicile. They block the opening of the cavity with the telson and thereby bar the entry of competitors or predators. The species of one genus, *Echinosquilla,* have evolved a telson that mimics a small sea urchin, so that the telson serves to camouflage the opening as well as to block it.

When a smashing stomatopod is attempting to evict another stomatopod from a cavity, it will coil with its telson directed at the entrance. The telson provides protection against the strikes of the cavity's occupant. *Haptosquilla glyptocercus,* which has the heaviest armor of any species we have studied, even uses its telson to evict the occupant of a cavity. It forces its way into the cavity tail first and then stops, plugging the entrance. It holds that position for several minutes, receiving repeated blows on the telson from the occupant of the cavity. Periodically the intruding animal withdraws from the entrance for a few seconds and then reinserts its tail. This behavior may continue for several hours until the occupant flees. We do not know whether the eviction results from a reduction of the oxygen in the blocked cavity, from a fouling of the water in the cavity by aggressive defecation or from prolonged stress.

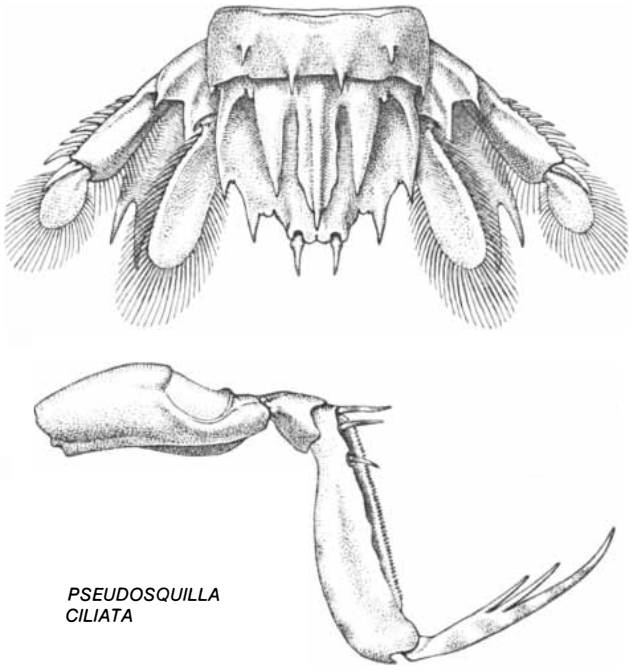
One of the largest stomatopods studied by one of us (Caldwell) is the smasher *Odontodactylus scyllarus.* The power of its blow is impressive. In the laboratory we have seen this species break open the shell of very large snails such as the fighting conch. *O. scyllarus* rarely engages in combat, however. If it did, its heavy blows could be fatal to an attacking stomatopod. Instead it presents a series of ritualized threat displays with its raptorial and posterior appendages. Interestingly, its telson is not heavily armored, and it has spines similar to those found on spearing stomatopods. It is a plausible hypothesis that this species has not evolved an armored telson and a more potent attack behavior because



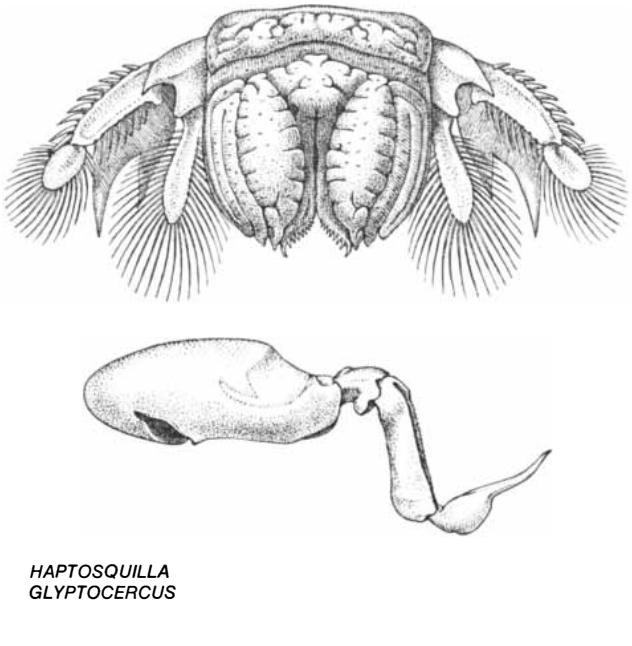
ACANTHOSQUILLA
ACANTHOCARPUS



GONODACTYLUS
INCIPIENS



PSEUDOSQUILLA
CILIATA



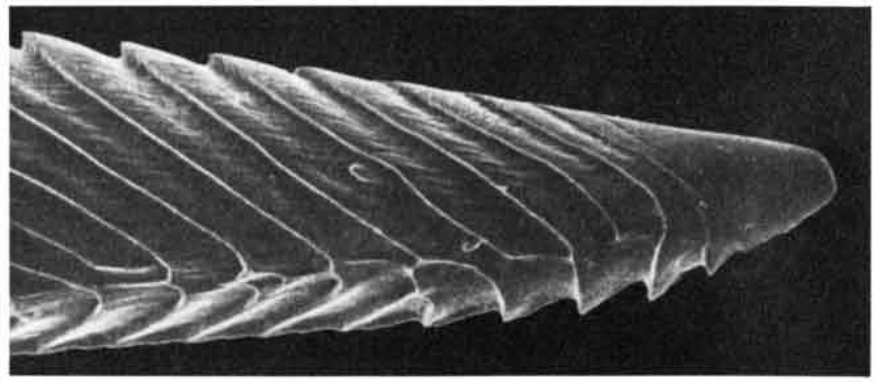
HAPTOSQUILLA
GLYPTOCERCUS

TELSONS AND RAPTORIAL APPENDAGES of two species of spearers (*left*) and two species of smashers (*right*) are depicted. The telson of spearers is relatively fragile. In some species it is armed with sharp spines. The raptorial appendages of spearers typically have spines, whereas the raptorial appendages of smashers have few spines or none. The telson of smashers is armored with ridges and blunt projections. Species of smashers that have powerful smashing appendages tend to have a heavily armored telson.

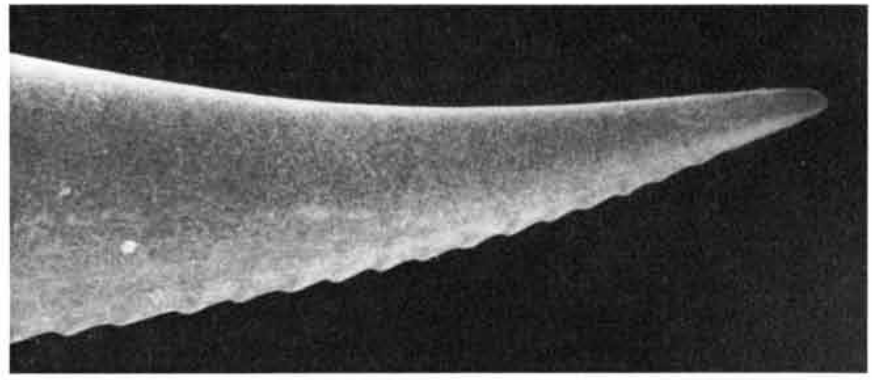
no amount of armor could protect it from the smashing strike of an opponent of the same species.

Most species of stomatopods have a conspicuous threat display that involves a posturing of the antennules, antennal scales and maxillipeds and the spreading of the raptorial appendages to expose a depression called the meral spot [*see bottom illustration on next page*]. The effect is to increase the animal's apparent size and to dramatically display its offensive weaponry. The meral spot of many smashers is brightly colored and edged in black. Species of stomatopods that make frequent use of a threat display typically have the most conspicuous meral spot. The meral spot of most spearers is rather dull in color. In general smashers use the meral display much more often than spearers do. The close association of frequent threat display and conspicuous coloration is another example of the coevolution of behavior and morphology in stomatopods [*see illustration on page 88*].

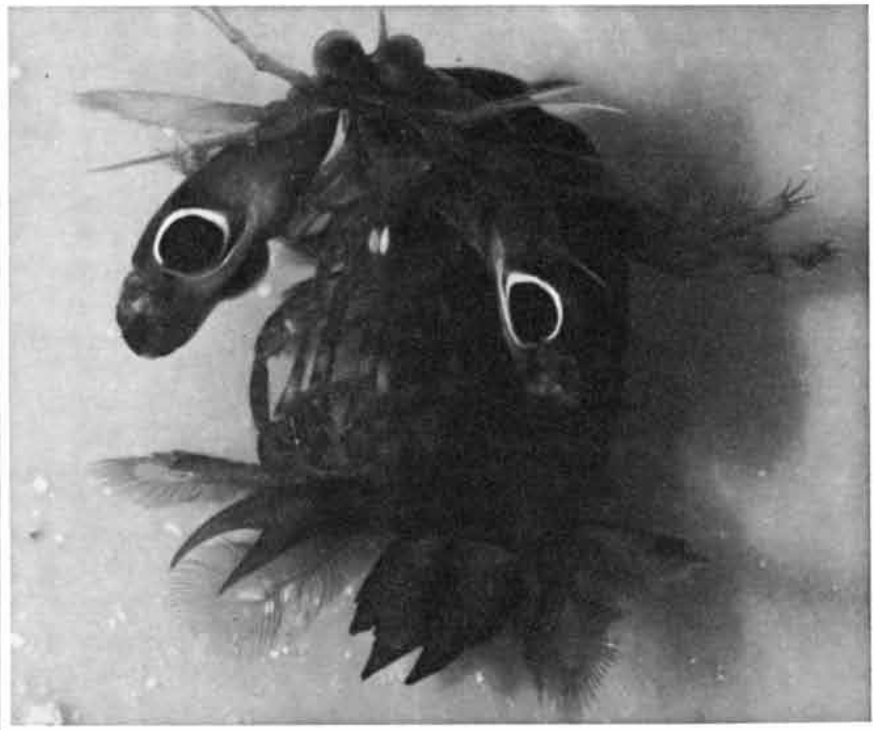
When several species of a stomatopod genus are found in the same area, each species usually has a distinctive meral spot. This color-coding apparently serves for species recognition in both aggressive and reproductive behavior. For example, two species common off the coast of Florida have the same body size and similar body coloration. They are often found in the same area, and the only way to readily tell them apart is by the color of the meral spot. One species, *Gonodactylus oerstedi,* has a purple spot; the other, *G. bredini,* has a white spot. We have studied four similar species of smashers in the waters off Thailand; all live on the



BARBED TIP of the dactyl spine of a spearer, *Squilla empusa,* is shown enlarged 125 diameters in this scanning electron micrograph. The barbs prevent prey from slipping off.



DACTYL TIP of a smasher, *Odontodactylus scyllarus,* is shown enlarged 125 diameters. Although the tip is occasionally used to spear prey, it does not have barbs. Both of the micrographs on this page were made by L. Jackson of the University of California at Berkeley.



CONSTANT THREAT DISPLAY of *Gonodactylus smithii* in the course of a fight frequently enables it to bluff its way through encounters with larger stomatopods. It has the most conspicuous meral spots and widest meral spread of any stomatopod studied by the authors.

edge of a reef and are distinguished by their meral spots, which in this instance are white, orange, yellow and purple. In the Gulf of Siam three species of spearers live side by side on mud flats. They are quite similar in appearance except for their meral spots, which are dark gray, light blue and yellow.

The threat display not only makes a stomatopod appear larger; it also signals a readiness to attack, and this in turn can inhibit attack by another animal. In one series of experiments we placed two individuals of the species *Gonodactylus zacae* in a container and observed their behavior the first time the two animals came in contact. When the animals were of equal size, not a single attack occurred when one of the animals gave a threat display during the initial contact. Attack was common when one of the animals did not give a threat display.

In many other *Gonodactylus* species smaller individuals tend to give a greater number of threat displays, both when they are matched against a stomatopod of the same size and when they are matched against larger stomatopods. Older ones are more likely to employ a "hawk" strategy, attacking first and then retreating into a defensive posture, with a threat display if necessary.

Gonodactylus smithii has the largest and most conspicuous threat display of any species we have studied. When it meets another stomatopod, its behavior is characterized by almost constant display. It is frequently able to bluff its way through initial encounters with species that are actually more aggressive.

There is evidence that threat behavior in a species is plastic, and that the frequency of threat display varies between communities of the same species. At Eniwetok atoll in the western Pacific *Haptosquilla glyptocercus* lives with two less aggressive species, and none of the three species uses threat display very often. In Thailand *H. glyptocercus* is found with two larger and more aggressive species. There it threatens more than twice as often as it does in the Eniwetok community, both against the other species and against other *H. glyptocercus* individuals.

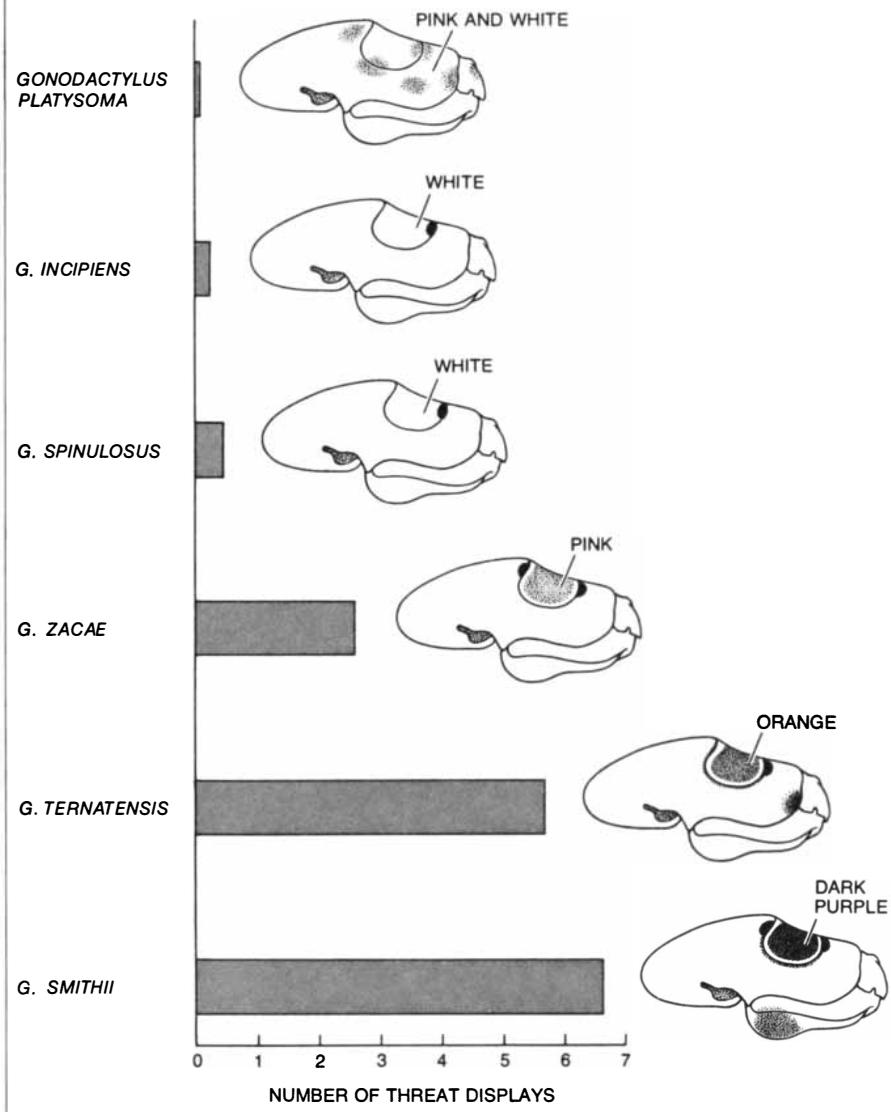
The strike of smashing stomatopods is so potent that the survival of a species calls for the accurate communication of aggressive intent. We have found a positive correlation between the power of a stomatopod's strike and the complexity of its behavioral repertoire. Spearers typically have seven or eight different aggressive displays. These in-

clude acts such as approaching, meral display, striking, grasping and avoidance. Smashers, on the other hand, have complex repertoires that may consist of 30 or more different acts.

Smashers that live in areas where there are a limited number of dwelling cavities tend to be more aggressive and to have more complex patterns of aggressive behavior than smashers in areas where suitable domiciles are abundant. For example, smashers that live in rubble cavities, which are usually found only in limited numbers in any one area, are extremely aggressive, whereas those that occupy cavities in live coral, which are usually found in larger numbers in any one area, tend to be less aggressive. *Gonodactylus platysoma* often does not occupy cavities at all but takes shelter in the large spaces under live coral. It is the least aggressive of the smashers we have studied and has the smallest repertoire of aggressive behavior.

Spearing stomatopods also show a stronger correlation between aggressive behavior and habitat. Spearers that dig long, complex burrows tend to be more aggressive than spearers that dig shallow tubes or trenches. In spearers aggressiveness appears to be correlated with the amount of energy and time put into burrow construction.

We have found that when several stomatopod species live in the same area, the more aggressive species are more abundant than the less aggressive ones. We have three clear examples. Before 1954 *Pseudosquilla ciliata,* a spearer, was the only stomatopod known to



NUMBER OF THREAT DISPLAYS given by six species of *Gonodactylus* are shown. The data were obtained from observation of 40 individuals of each species during a 10-minute confrontation with another stomatopod. It was found that species with conspicuous meral spots tended to give threat displays more often than species with less prominent meral spots.

inhabit cavities in the coral rubble around the Hawaiian Islands. Then a powerful smasher, *Gonodactylus falcatus,* was introduced (probably by Navy barges), and by 1963 the smasher had evicted the spearer from the coral rubble. The spearer is now found only on sand flats, where it constructs burrows or lives under small rocks or marine algae. When *G. falcatus* and *P. ciliata,* which are about the same size, fight in the laboratory, *G. falcatus* always wins.

On the intertidal mud flats at Ang Sila on the Gulf of Siam the spearer *Clorodopsis scorpio* is five times more abundant than another spearer, *Oratosquilla inornata.* Both are the same size, have the same diet and dig their own burrows. The burrow of *C. scorpio,* however, is more complex than that of *O. inornata.* When the two species fight, *C. scorpio* easily defeats *O. inornata,* and it is almost always able to evict *O. inornata* from its burrow.

We have mentioned that *Haptosquilla glyptocercus* is the most aggressive stomatopod in the coral-rubble habitat of Eniwetok but is dominated by other species in Thailand. We also found a relation between aggressiveness and abundance: at Eniwetok *H. glyptocercus* is the most abundant species, and in Thailand the other species are more abundant. Although the relation undoubtedly exists, it is not as simple as it might at first appear. If aggressiveness were the only factor involved, it would be difficult to explain the coexistence of several species.

It may be that the less aggressive species survive by specializing in a microhabitat that differs slightly from the microhabitats of the more aggressive ones. Other factors could also account for coexistence, among them differences in the degree to which the various species are preyed on by other animals, differences in size and differences in rate of reproduction.

The two lines of stomatopods—the spearers and the smashers—point an evolutionary moral. Our studies show an unmistakable association between changes in the form of the stomatopods' raptorial appendages and changes in their behavior, body armor, threat display, type of habitat and competitive ability. The stomatopods are an unusually clear example of the unity of evolution, reminding us that evolutionary change in one part of an organism is only one aspect of a comprehensive transformation of the entire organism and its place in nature.

Paleoneurology and the Evolution of Mind

Tracing changes in the relation between brain size and body size in various groups of fossil and contemporary animals sheds light not only on the evolution but also on the nature of intelligence

by Harry J. Jerison

The mind evolved. Paleoneurology, which deals with the evidence of nervous systems in fossil animals, provides new clues to the nature of that evolution. Integrated with information about the variety of brains and behaviors in living vertebrates and knowledge of how neural tissue is packaged in brains, the fossil record can be interpreted to develop a coherent account of the evolution of intelligence over the 500-million-year span of vertebrate history. That account provides fresh perspectives on the nature of intelligence as a biological phenomenon.

The approach complements more traditional ways of studying the evolution of mind: ethological analyses, based on naturalistic observations of species-typical behaviors of living animals, and psychological studies, which measure the competence of various species in standardized laboratory tests. The traditional analyses build on evolutionary relations among living animals to reconstruct the evolution of behaviors; the variety of present patterns of behavior are projected backward in time to probable ancestral patterns. With paleoneurological data, on the other hand, the approach to the evolutionary history of the brain is direct, and the history is interpreted in the light of present-day relations between brain and behavior.

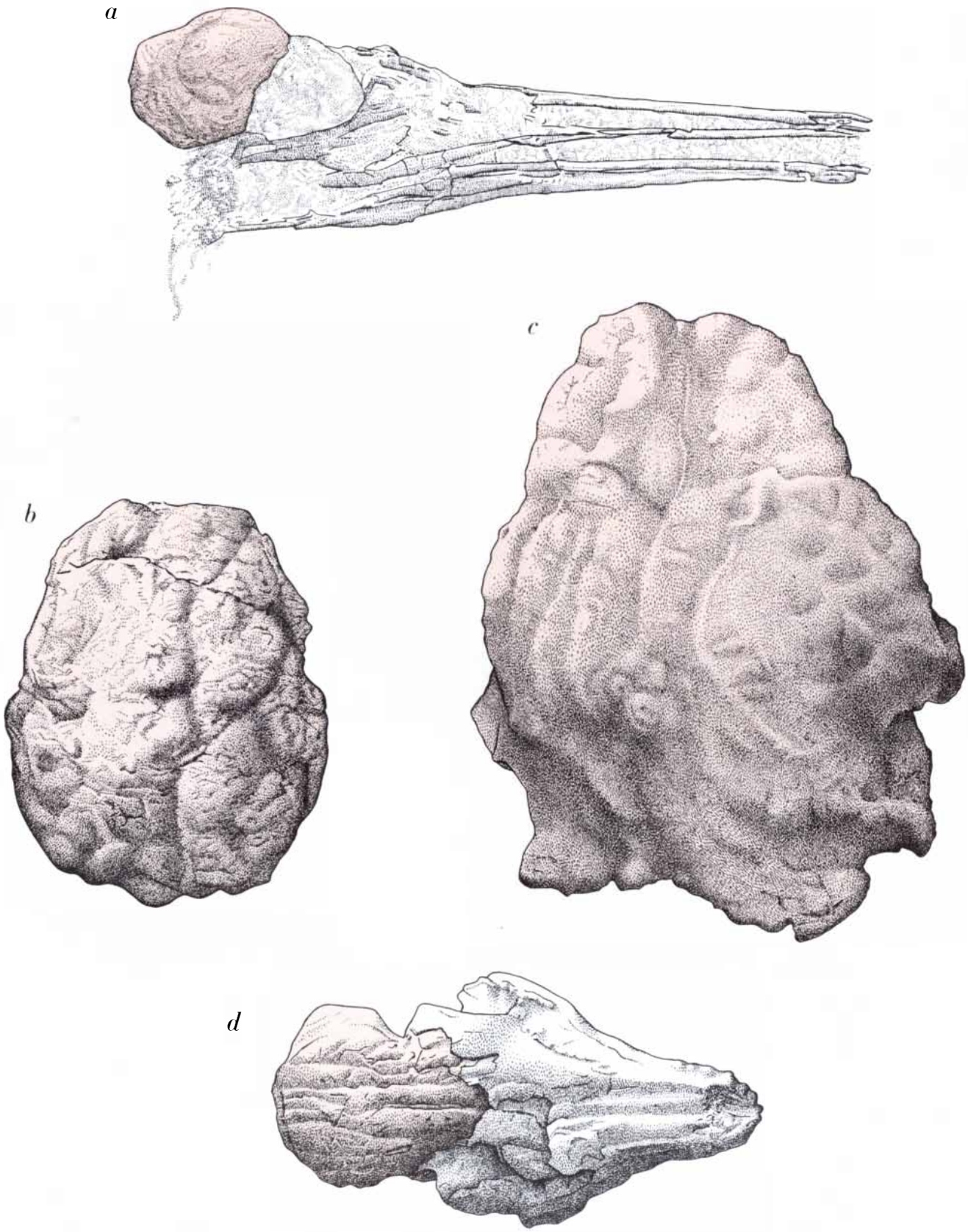
The strategy of the paleoneurological analysis of mind is to identify a morphological trait as a correlate of mind, or biological intelligence. If the mind evolved, certain trends in the evolution of that correlated trait should be evident in the fossil record. For example, since there are obviously different grades in the distribution of intelligence or mind in living animals, it should be possible to measure an increase and a diversification of the morphological trait in successive geological periods.

Charles Darwin was the author of an early statement of one hypothesis that relates mind to morphology when he wrote in *The Descent of Man:* "No one, I presume, doubts that the large proportion which the size of man's brain bears to his body, compared to the same proportion in the gorilla or orang, is closely connected with his mental powers." Some 80 years later Karl Spencer Lashley's more careful and explicit statement of the hypothesis made it applicable to the analysis of morphological data: "The only neurological character for which a correlation with behavioral capacity in different animals is supported by significant evidence is the total mass of tissue, or rather, the index of cephalization... which seems to represent the amount of brain tissue in excess of that required for transmitting impulses to and from the integrative centers."

The index, a measure of relative brain size, is defined as total brain size divided by the two-thirds power of the body size. (The exponent has to do with the relation between surface and volume, as I shall explain further along.) The ratio can be visualized if one plots data reflecting the present diversity of brain and body sizes in vertebrates [*see top illustration on page 94*]. Lashley's statement is borne out by the clear differentiation between "lower" and "higher" vertebrates and by the fact that the arrays of points representing both groups are similarly oriented at a slope of 2/3. It is easy to imagine the higher vertebrates as having evolved from a lower vertebrate grade as the result of the vertical displacement of a set of points in brain : body space. That displacement is in effect what is measured by the index of cephalization. This implies that brain size is determined by a "body-size factor" and an "encephalization factor." Lashley's hypothesis was that only the encephalization factor was involved in the evolution of mind. Birds and mammals are "higher" vertebrates because they are more highly encephalized and are higher on a scale of biological intelligence than reptiles and other "lower" vertebrates. That is an interpretation of the data that agrees nicely with our intuition.

To assemble paleoneurological evidence one needs to know the brain and body size of fossil vertebrates. The brain size is determined from the volume of endocranial casts, or endocasts, that are replicas of the cranial cavity. In some cases natural endocasts are found: mineralized remains of sand or other deposits that replaced the soft tissue in a fossil skull. In other cases the fossil skull serves as the mold for a man-made acrylic or plaster endocast. The body size of the fossil vertebrate can be measured or estimated from models based on skeletal reconstructions. To take one example, the Eocene ungulate (hoofed mammal) *Uintatherium anceps* weighed about 1,250 kilograms; its endocast indicates a brain weight of about 290 grams [*see illustration on page 95*]. By assembling and plotting a large body of data of this kind it is possible to derive "brain : body maps," minimum convex polygons that enclose a set of brain : body points and enable one to analyze major shifts in encephalization without numerical analyses or indexes [*see bottom illustration on page 94*].

In the illustration a polygon for living mammals defines the region in brain : body space associated with the present mammalian grade of encephalization.



ENDOCASTS, casts of the inside of the cranial cavity, are displayed for four fossil animals. These endocasts, each drawn about one and a quarter times actual size, are natural ones: mineralized remains of sand or other deposits that replaced the soft tissues in the fossil skulls. The endocast of *Numenius gypsorum,* a bird of the Eocene (about 40 million years ago), is still in place in the long-beaked skull (*a*). The endocasts removed from their craniums are of *Potamotherium* (*b*), an otterlike Oligocene carnivore (about 25 million years ago), and *Dicrocerus* (*c*), an ungulate, or hoofed herbivore, from the Miocene (15 million years ago). The endocast of *Cainotherium* (*d*), an Oligocene ungulate, has parts of the skull in place. Casts are from the Muséum d'Histoire Naturelle in Paris.

Is the leg mightier than the atom?

Before you say no, keep in mind that we know very little about many forms of energy available to us.

Including good old muscle power.

For too long a time we've relied on oil and gas to serve our needs, and failed to take full advantage of other sources of power.

Including the atom.

But recent events make it clear we must learn about all the options, and how best to apply them.

At Union Carbide we're studying a wide range of energy technologies and resources for the Energy Research and Development Administration.

From something as basic as bicycling to the complexity of controlling nuclear fusion.

For instance, we are learning how to turn coal into oil and gas in a way that is practical economically.

We're deeply involved in nuclear research, particularly in finding ways to make this important source of energy safer and more efficient.

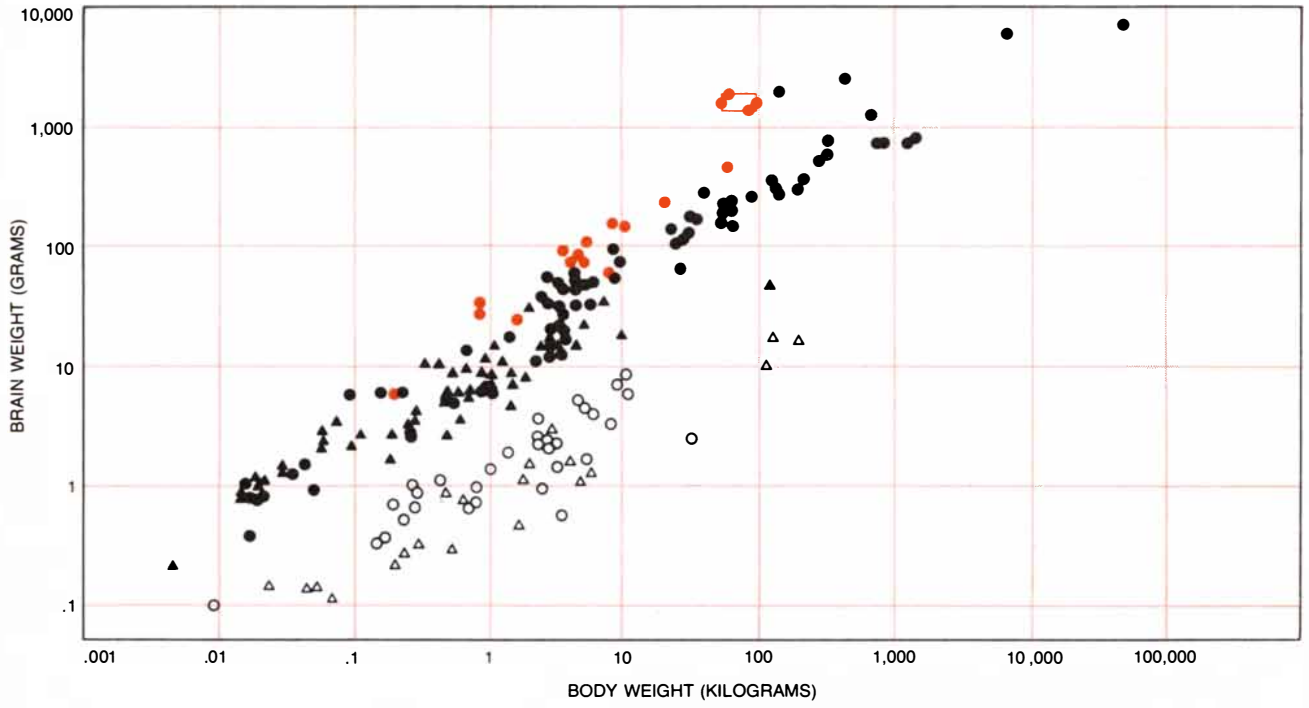
Our work in fusion power, at Oak Ridge, Tennessee, offers the most exciting possibility for the future: the ultimate source of inexhaustible energy.

If we succeed, there will never be another energy crisis.

But for the present, the answer to our energy dilemma is not likely to come from one source, but many. All the way from the leg to the atom.

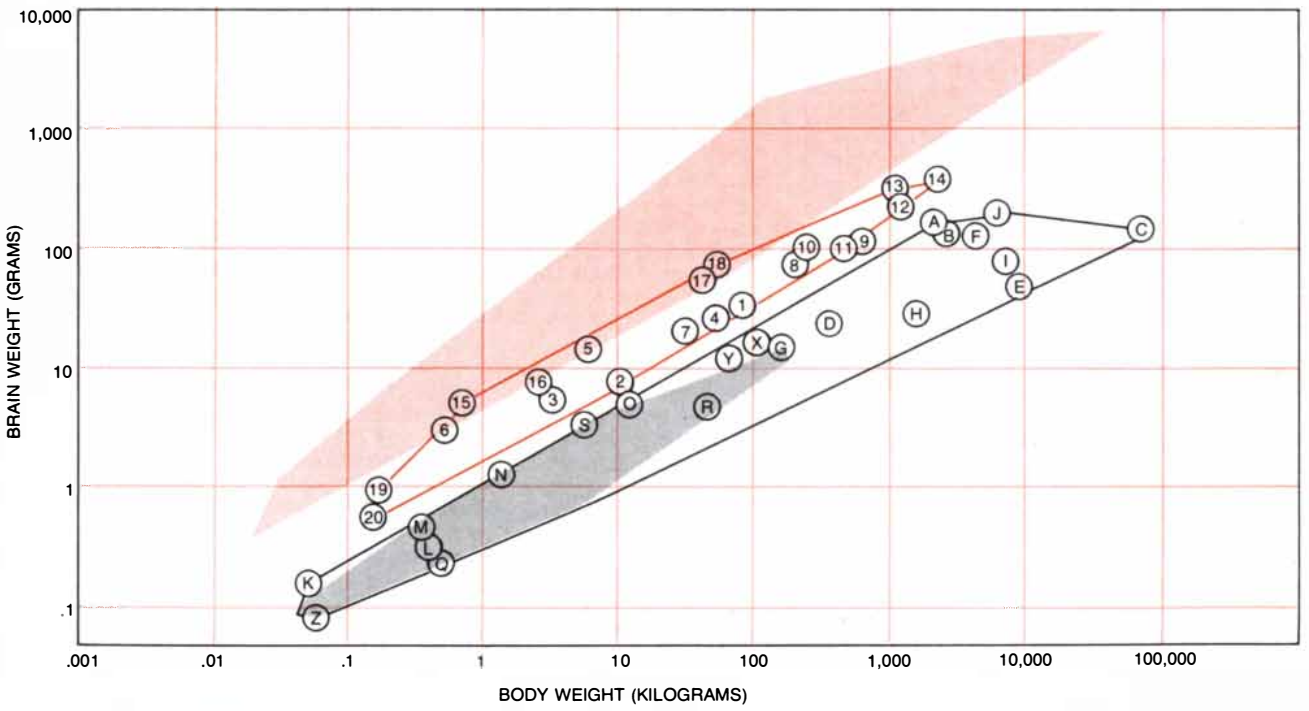
UNION CARBIDE

Today, something we do will touch your life.



PRIMATES
MAMMALS
BIRDS
BONY FISH
REPTILES

BRAIN SIZE is plotted against body size for some 200 species of living vertebrates. The data were collected by George W. Crile and Daniel P. Quiring some years ago. The four colored points connected by a rectangle represent the extreme measurements reported for man, indicating that variation within a species does not loom large in comparison to the distinctions among species. Data fall into two clearly delimited groups, which may be considered to be the lower and the higher vertebrates. In both cases the data fall along a line with a slope of 2/3 on log-log coordinates: brain size varies with the 2/3 power (which is the cube root of the square) of body size.



LIVING MAMMALS
ARCHAIC MAMMALS
LIVING REPTILES
ARCHAIC REPTILES

BRAIN : BODY MAPS are the minimum convex polygons that can be drawn to enclose a set of points representing brain size plotted against body size. Here the data for living mammals and reptiles have been taken from the illustration at top of page. Numbered points are for archaic ungulates and carnivores; oldest is *Triconodon* (*20*). Fossil reptiles are dinosaurs (*A—J*), pterosaurs (*K—O*) and mammal-like reptiles (*Q—S*); two amphibians (*X*, *Y*) and a fish (*Z*) are included.

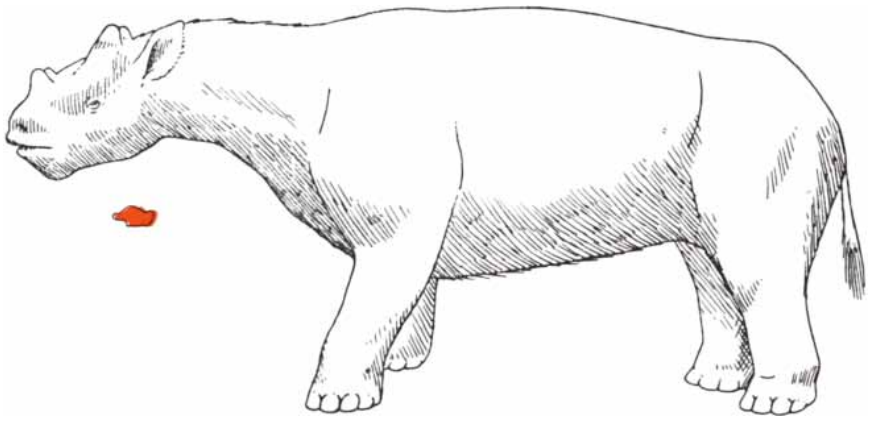
The brain : body data of early fossil mammals are shown in the archaic-mammalian polygon. ("Archaic" designates taxonomic orders that are now extinct, having been replaced in their ecological niches by species from "progressive" orders.) The geologic time of the map for archaic mammals extends from the upper Jurassic period represented by the earliest well-preserved mammalian endocast (*Triconodon mordax*) to the late Eocene epoch with forms such as *Uintatherium*, or a time span of from about 150 to 40 million years ago. The polygon indicates that a single archaic mammalian grade persisted for the entire period. In the reptilian polygon the fossil data on dinosaurs, pterosaurs (flying reptiles) and therapsids (mammal-like reptiles) are entered along with the sample of living reptiles. Two fossil amphibians and one fossil fish are included, so that the reptilian polygon is a reasonable picture of the original lower grade of brain : body relations in jawed vertebrate species, the grade from which birds and mammals advanced when they first evolved.

The brain : body maps lead to a number of interesting conclusions:

1. Living reptiles have not departed significantly from their ancestral condition. A single brain : body map includes all the fossil and living species for which we have data, showing that the adaptive radiation of the reptiles and other lower vertebrates was accomplished without any major advances in relative brain size.

2. Adding fossil reptiles extends the reptilian polygon while maintaining its orientation in brain : body space. This fact and the data on mammals suggest that a fundamental biological process determines the orientation at a slope of about 2/3. Because the coordinate system is logarithmic the slope can be interpreted as providing exponents for the measures of body and brain (2 for the square units of a body surface and 3 for the cubic units of a brain volume) and thus for the exponent of 2/3 in the index of cephalization. That is a straightforward application of principles of physics. It makes sense biologically because of the way sensory and motor "surfaces" of the body are projected on the many structures that add up to the volume of the brain; maps of the cerebral cortex of mammals are often presented as grossly distorted but recognizable body surfaces drawn on a diagram of the brain.

3. Dinosaurs were not unusually "small-brained." The 10 dinosaurs that contributed data on fossil reptiles were the source of the inference about the stability of the reptilian, or lower vertebrate, condition and the uniformity of slope for both fossil reptiles and living ones. The 10 dinosaurs were clearly normal reptiles with respect to brain size.



FOSSIL ANIMALS are reconstructed on the basis of their skeletal remains in order to provide an estimate of body size to compare with the brain size derived from an endocast. The body can actually be modeled and the model's volume can be determined, or the size can be estimated from a drawing. This animal is *Uintatherium anceps*, an archaic ungulate of some 40 million years ago. It weighed about 1,250 kilograms and its brain (*color*) weighed about 290 grams. Data are plotted as point No. 12 in archaic-mammal map on opposite page.

4. Encephalization and the evolution of intelligence occurred independently in birds and in mammals, which evolved independently from two different subclasses of reptiles. (The record on fossil birds is not considered in detail in this article; the 150-million-year-old fossil *Archaeopteryx*, the earliest-known bird, was intermediate between the reptilian grade and the grade of living birds.)

5. The reptilian map and the two mammalian maps are similarly oriented with a slope of about 2/3, and there are vertical displacements that indicate the progressive encephalization of the mammals.

6. Although the step from reptiles to mammals required a certain amount of encephalization (approximately a fourfold increase in relative brain size to transform a reptilian polygon into an archaic mammalian polygon), mammalian encephalization did not progress immediately but remained at a steady level for at least 100 million years. The earliest and the latest of the archaic species of mammals were all at about the same grade of relative brain size, enclosed within a single rather narrow polygon. That stability for such a long period of time suggests a successful response to the selection pressures of a stable new ecological niche.

7. Progressive evolution of encephalization within the mammals came late in their history, in the last 50 million years of a time span of about 200 million years. That evolution transformed the archaic mammalian map into the map of living mammals by another four- or fivefold increase in relative brain size for the average mammal.

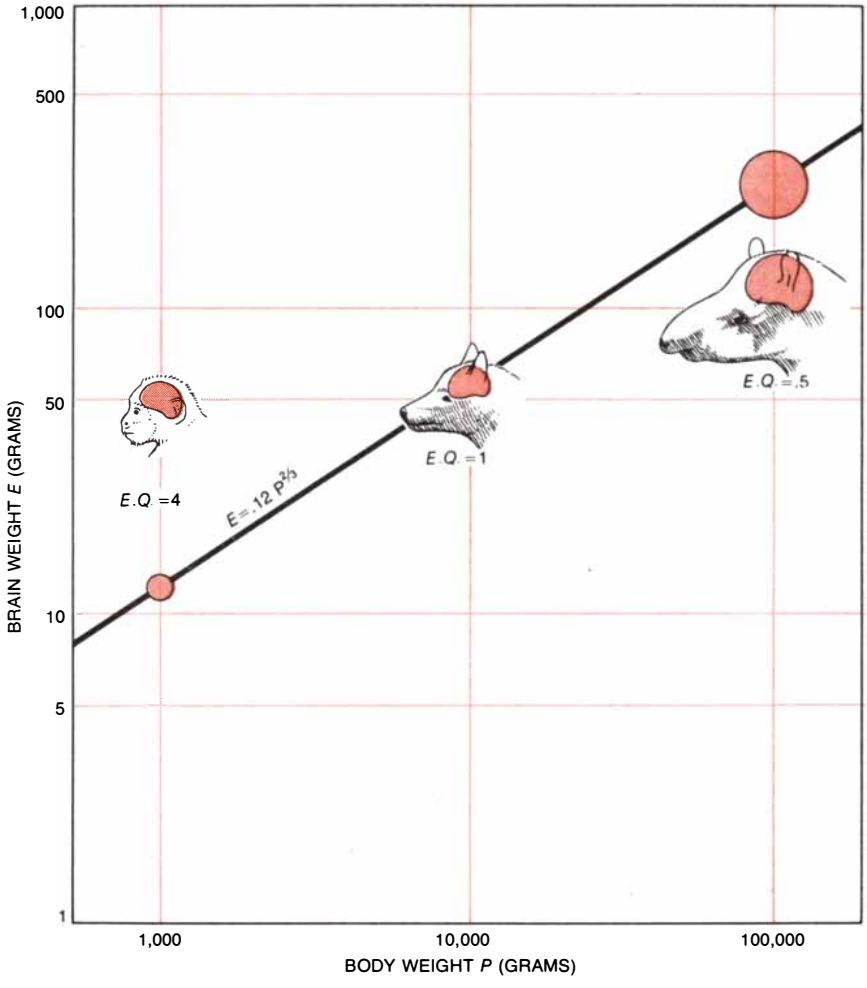
The final steps in encephalization are revealed in the fossil record of mammalian endocasts of the present geological era: the Cenozoic, which began approximately 65 million years ago. Those steps are best measured with a modern version of the index of cephalization: the encephalization quotient (E.Q.), which is the actual size of the brain divided by its expected size for an average living mammal. The expected size is determined by an equation, which states that the brain size equals the 2/3 power of the body size multiplied by a constant (.12) that represents the index of cephalization for an average living mammal [*see illustration on next page*].

In order to analyze the progressive evolution of encephalization in the mammals, I have computed average E.Q.'s and standard deviations for samples of fossil and living ungulates and carnivores and plotted them as a set of normal curves [*see top illustration on page 97*]. The curves have equal areas and can therefore be treated as normal probability distributions, that is, each curve shows the probability of the presence of species of the indicated degrees of encephalization in each assemblage. There was evidently a steady advance in average E.Q. during the Tertiary period (from about 65 million to about three million years ago), and there was also a diversification of E.Q. that was

more or less proportional to the average E.Q. Significantly, the distributions overlap in the lower range of encephalization, indicating that some relatively small-brained species persisted in all the fossil groups and that they still persist among living mammals.

These are results that one would expect if encephalization (and intelligence) evolved as other traits have. Evolution involves morphological and behavioral adaptations to a variety of niches and to the invasion of new niches. The adaptive zones occupied by successive species of carnivores and ungulates during the Tertiary period must have included many niches in which there were selective advantages for species further encephalized. As more of these niches were invaded there would have been diversification, which would in turn have affected the frequency distributions of E.Q.'s [*see bottom illustration on opposite page*].

The gradual change in encephalization throughout the Tertiary period has thus far been demonstrated only for the ungulates and carnivores among the orders of mammals. Although that may be owing simply to a lack of enough data on other orders, the limited evidence on other mammals suggests a different pattern of evolution: When an adaptive zone was entered, the succession of species in that zone attained a particular grade of encephalization rapidly and then maintained it. For example, the earliest-known insectivore endocasts, from about 35 to 40 million years ago, were already at the same grade of encephalization as average living species of insectivores, namely somewhat above the archaic level. Similarly, endocasts and body sizes for "dolphins" of about 18 million years ago, the earliest in this group of cetaceans, show that they had already reached the E.Q. level of comparable living species such as the harbor porpoise. Dolphins, as large-brained mammals, may therefore represent an evolutionary picture very different from the more or less equally encephalized human species. As we shall see, the evolution of the hominid brain to its present size is a relatively recent phenomenon, having been completed only within the past million years.

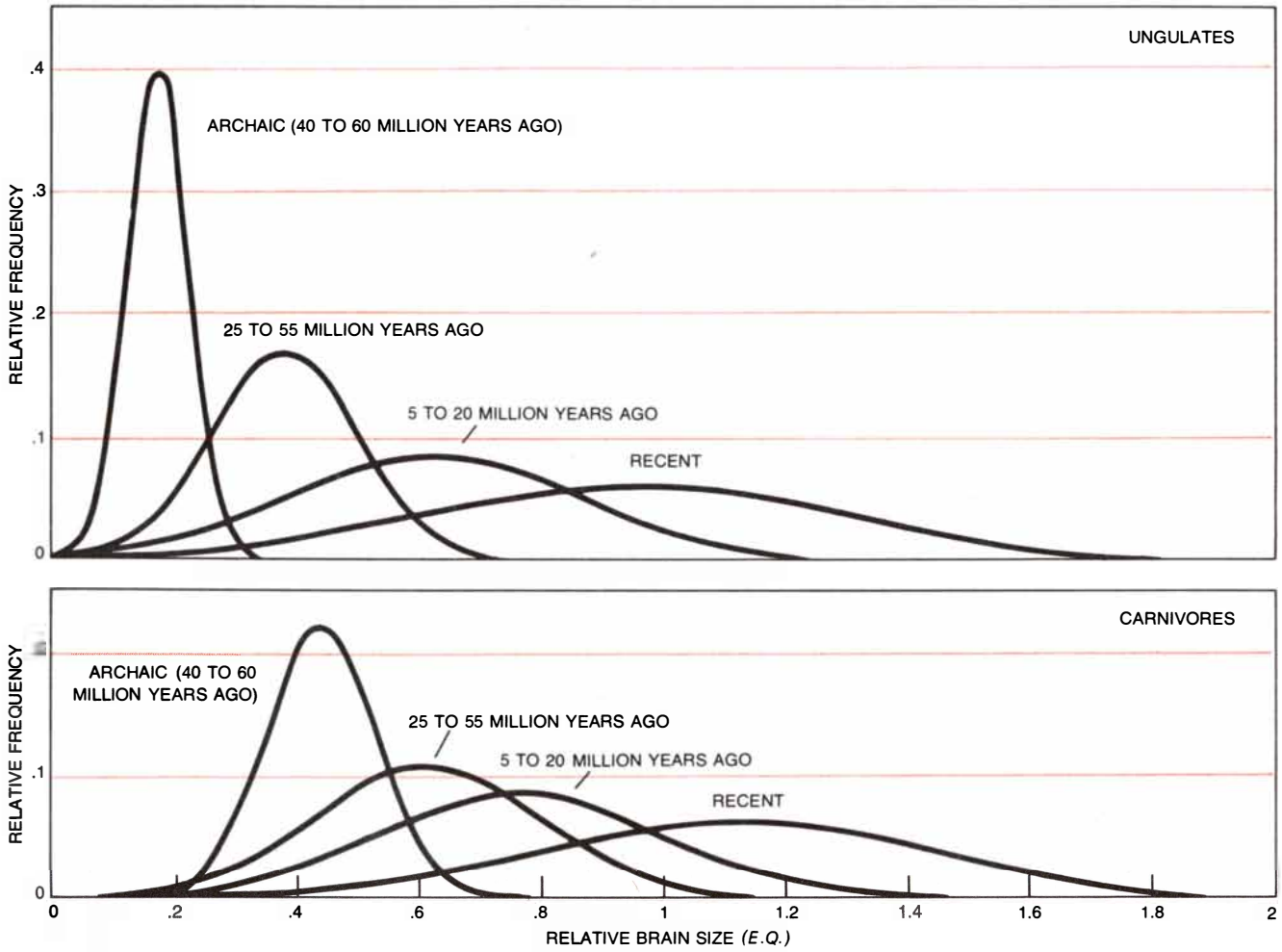


ENCEPHALIZATION QUOTIENT (E.Q.) is the ratio of an animal's actual brain size to its "expected" brain size. The expected sizes, represented by the diagonal, are given by an equation: brain size (*E*) equals .12 (the index of cephalization for an average mammal) times the 2/3 power of the body size (*P*). A hypothetical "smart" monkey has a brain four times as large (48 grams) as the expected size (12 grams) for its body size; its E.Q. is said to be 4. An "average" dog has an E.Q. of 1 and a "stupid" tapirlike animal has an E.Q. of .5.

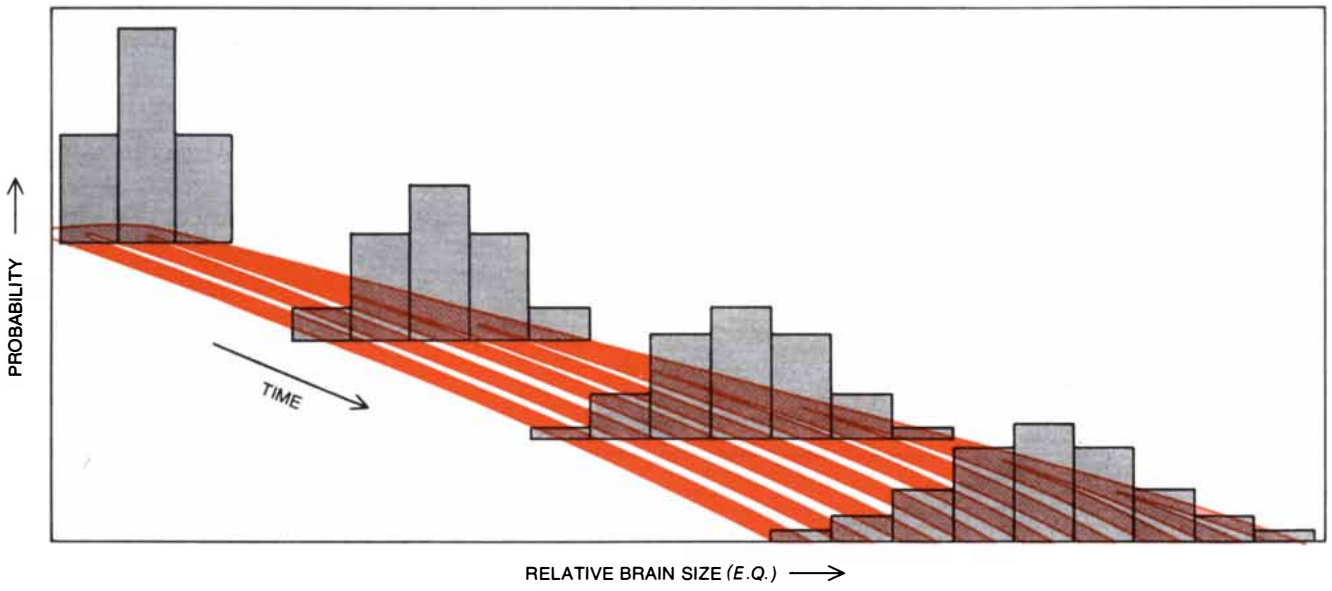
The fossil evidence on the primate brain is not as complete as that for carnivores and ungulates, but it is orderly and involves several striking features. The primates have always been large-brained mammals. Early Tertiary prosimians had values of E.Q. ranging from .55 to 1.75. This can be contrasted with the range for all other early Tertiary progressive land mammals on which I have data, which is from .19 to .92. Living monkeys and apes are about twice as encephalized as other living land mammals (a mean E.Q. of 2.1 for simians as compared with 1.0 for average mammals). Primates of about 50 million years ago were also about twice as encephalized as their land-mammal contemporaries.

Encephalization within the hominids is measurable by cranial capacity rather than E.Q. (The comparison of cranial capacities is equivalent to the analysis of encephalization in species that are similar in body size, since it amounts to scaling vertically in brain : body space.) An australopithecine grade—a volume of about 500 milliliters—had been achieved in the early Pleistocene epoch, two or three million years ago, and probably as early as the Pliocene, five million years or more ago. A pithecanthropine grade has been recorded in a unique 800-milliliter endocast from Lake Rudolph in Kenya that is almost three million years old; the better-known true pithecanthropines are about a million years old and their cranial capacities range from about 750 to 1,250 milliliters [see "The Casts of Fossil Hominid Brains," by Ralph L. Holloway; SCIENTIFIC AMERICAN, July, 1974].

The earliest endocast for *Homo sapiens* is about 250,000 years old, and man's present cranial capacity ranges from about 1,000 to 2,000 milliliters. The important evolutionary fact is the rapidity and recency of the increase in encephalization in the hominid lineage. There is no evidence of a change in encephalization in any other mammals in



CHANGING DISTRIBUTIONS of relative brain size (E.Q.) are plotted for ungulates (*top*) and carnivores (*bottom*). Each curve gives the distribution of brain sizes for various species during specified periods; recent species are those living today. There is a steady increase in brain size and a concomitant diversification, with small-brained species persisting, shown by the flattening of successive curves. Note that these are "between species" curves. Differences within a species are usually not reflected in behavior.



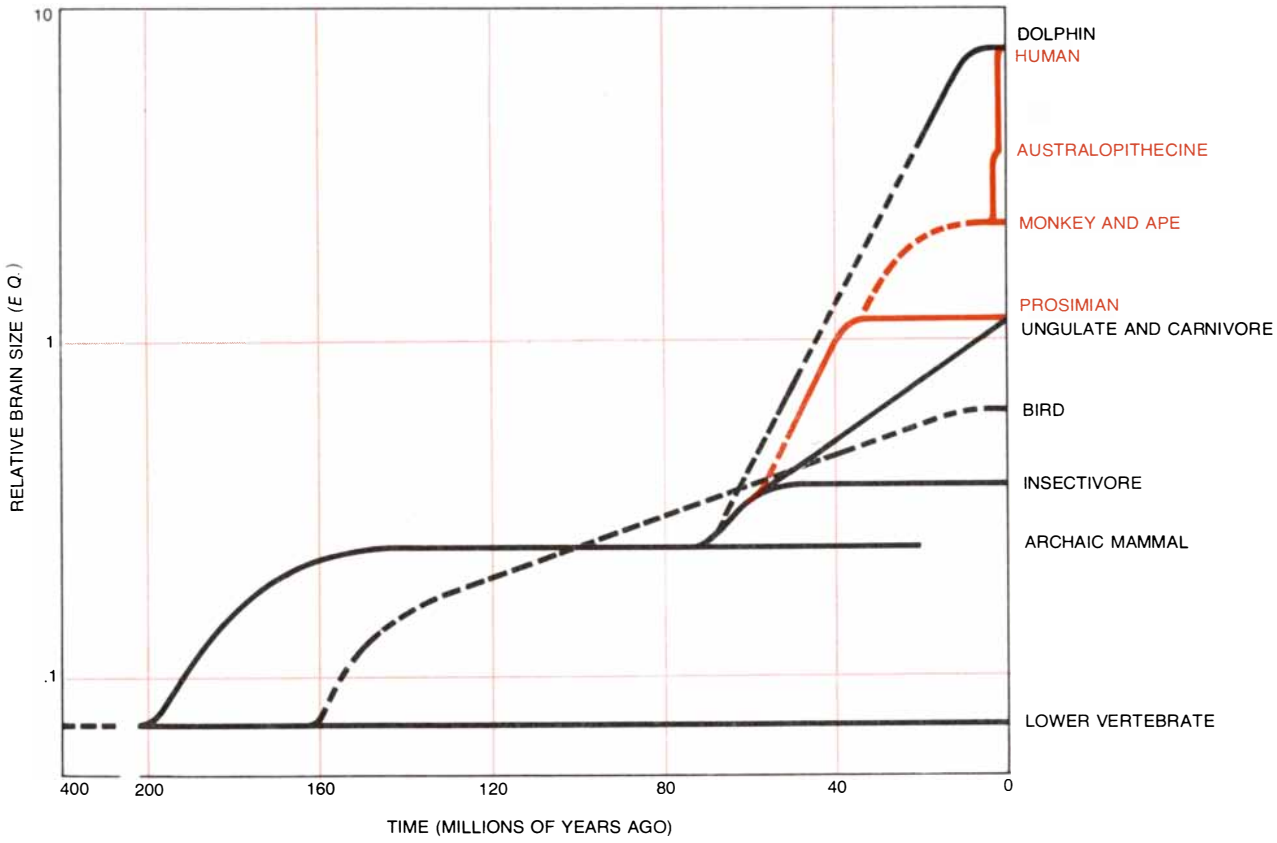
DIVERSIFICATION IN BRAIN SIZE reflects an adaptive radiation with respect to encephalization. The diverging pathways (*color*) suggest the increasing number of ecological niches making diverse demands on the brain. Niches appeared that required more encephalization, and other niches were preserved that made smaller demands. Animals adapted to a wider range of "encephalization niches" were able to survive. The bars reflect schematically the fact that distribution depends on discrete numbers of niches.

the past five million years [*see illustration below*]. In the remainder of this article I shall discuss some critical events in the history of the vertebrates that provide clues to the causes of encephalization and also to the nature of biological intelligence.

Vertebrates do not live by brains alone. That is the message from the insignificant encephalization in the many successful species of lower vertebrates and from the stability of relative brain size in the archaic mammals for at least 100 million years. The advance from fish to amphibian about 350 million years ago illustrates the conservatism in brain evolution. Here was an invasion of terrestrial niches, possibly the most demanding new adaptive zone in vertebrate history, yet according to the paleoneurological evidence there was no increase in encephalization. That was possible because the earliest amphibian required only minor alterations in the patterns of neurological and behavioral organization of its immediate ancestor among the bony fish. Sir James Gray has shown that even the adaptations for movements on land could be "conservative." The legs of an amphibian served the same function as the inertial force of water for a swimming animal, providing a fulcrum that enabled early amphibians to be little more than fish that swam on land.

In the same sense, I believe, the earliest mammals were probably only slightly modified from their reptilian predecessors: they were "reptiles" that were active at night. This view, which explains the evolution of enlarged brains in early mammals, is consistent with the general evolutionary history of reptiles of the late Paleozoic and early Mesozoic eras, about 250 million years ago. The mammal-like reptiles called therapsids were the dominant forms during the late Paleozoic era, and they were replaced during the early Mesozoic by the ruling reptiles, notably the dinosaurs. The dinosaurs and therapsids were clearly in a kind of competition for the normal niches of land reptiles, and the mammal-like reptiles lost the competition, becoming completely extinct by the mid-Mesozoic era, about 150 million years ago. The niches for which the therapsids competed and lost were for large diurnal animals that used vision as the normal sense for receiving information about events at a distance. The earliest mammals were a persistent remnant of only slightly modified therapsids, and they survived as small, nocturnal animals. They had improved auditory and olfactory systems and their visual systems were modified, with rod cells taking the place of cone cells. Both changes were appropriate for animals that were active at night.

The evolution of hearing and smell to supplement vision as a distance sense is sufficient reason for the evolution of an enlarged brain in the earliest mammals. The reason is to be found in the way neural elements are packaged in vertebrate sensory systems. In the visual system many of the circuits are in the retina, which contains an extensive and complex neural network that allows elaborate analysis of visual information. The corresponding neural elements of the auditory and olfactory systems of



RATE OF EVOLUTION of grade of encephalization varied in different vertebrate groups and at different times. There was actually variability of brain size within groups too, so that the curves are somewhat arbitrary. In the case of cetaceans the highest grade attained is plotted to emphasize that the cetaceans that reached the dolphin grade did so long before the primates reached even an australopithecine grade; the recent and rapid evolution of the hominid brain is notable. The broken lines indicate gaps in data.

living vertebrates are in the brain proper [*see illustration on next page*].

In quantitative terms the effect of different modes of packaging is enormous. A small, highly visual lizard such as the American chameleon *Anolis* has in its retina at least a million sensory cells (cones) and about 100,000 ganglion cells, where the fibers of the optic nerve originate. Still other retinal neurons are involved in higher-order processing in that lizard's retina, and they must be almost as numerous as the sensory cells. The ear of *Anolis,* on the other hand, has only a few hundred sensory cells and other neural elements external to the brain. An auditory system analogous to the visual system would presumably have to have about as much integrative circuitry as there is in the retina, so that an "auditory" animal the size of *Anolis* would need space for almost a million integrative neurons and their dendritic fields to analyze inputs from the ears. There is no space for these in the middle and inner ears; the obvious place to package the additional material is in the brain itself, and solving the packaging problem would therefore require the enlargement of parts of the brain involved in audition. A similar argument would apply to an "olfactory" animal: its increased dependence on the sense of smell would require expansion of the forebrain systems that contain the integrative neurons of the olfactory system.

The brain of an early mammal with the body size of *Anolis* would have handled distance information in a reptilian way but with audition and olfaction as the receptor systems in place of daylight vision. Such a mammal's brain would have had to be enlarged compared with a normal Mesozoic reptilian brain in order to have space for the new neural networks that evolved to analyze nonvisual information. And so we see that the first expansion of the vertebrate brain may have been primarily a solution of a packaging problem and that it may only incidentally have resulted in the evolution of intelligence. Let us see how this crucial incidental result might have been attained.

The introduction of encephalized and finely discriminative audition and olfaction demands new ways of encoding neural information. Visual information is encoded at a retinal level with a structurally determined spatial code: the optics of the eye and the arrangement of retinal elements provide a grid that labels the location of stimulated cells. No such code is possible for sound or odors. It is difficult for us, the least olfactory of land mammals, to imagine how spatial information could be encoded with the olfactory system, but consider how such coding is accomplished by the auditory system. Animals that use echolocation to identify the source and shape of distant objects do it by translating spatial information into a temporal code. Some such localization of environmental sounds in space must have been accomplished by the early mammals if audition was for them a precise distance sense. In the evolution and functioning of the neural apparatus necessary for such behavior two dimensions of sensory experience had to be encoded by the brain: space and time.

More was needed. Imagine an early mammal coping with life at twilight, sensing stimuli from distant sources. The stimuli are recorded by reptilelike vision (modified toward the mammalian retina with rod cells for night vision), mammalian hearing and mammalian smell, all providing information from the same environmental source. It would obviously be adaptive if the information received from the different sense modalities were given a common code, or label. The integrating code would work, in all likelihood, by the labeling of stimuli in the different modalities as coming from the same object in space at a particular time. And so we have the basic constructs of human conscious experience: objects in space and time. The conscious experience is essentially a construction of nervous systems for handling incoming information in a simple, consistent way.

This leads me to a few simple propositions. Reality, or the real world we know intuitively, is a creation of the nervous system: a model of a possible world, which enables the nervous system to handle the enormous amount of information it receives and processes. (That view is similar to one presented by the English psychologist K. J. W. Craik on the nature of explanation.) The "true" or "real" world is specific to a species and is dependent on how the brain of the species works; this is as true for our own world—the world as we know it—as it is for the world of any species. (That view was made familiar by the German biologist Jakob Johann von Uexküll, who described the "perceptual worlds" of animals and men.) The work of the brain is to create a model of a possible world rather than to record and transmit to the mind a world that is metaphysically true.

Biological intelligence, then, is a measure of the quality of the particular real world created by the brain of a particular species. The world as we know it ourselves, with the self as perhaps its most complex object, represents the human grade of biological intelligence. Different worlds are presumably constructed by different species. A very simple construction of a world may be characteristic of the lower vertebrates. As a matter of fact, no transformation of neural information—no construction—at all may be required in the lower vertebrates. Their behavior is tightly bound to specific stimuli by fixed-action patterns of response, in contrast to an "intelligence" system in which varied patterns of stimuli are transformed into invariant objects. The birds seem to be a special case. In birds the fixed-action pattern is the typical behavioral mode, and biological intelligence may be a little-used capacity. Yet experimental procedures showing that birds are well within the mammalian range of competence in performing standardized—albeit "unnatural"—learning tasks seem to affirm the basic validity of the judgment, based on relative brain size, that birds and mammals are at comparable grades of biological intelligence. Intelligence, in biological perspective, is clearly only one of several dimensions of behavior and is not the most important one for birds.

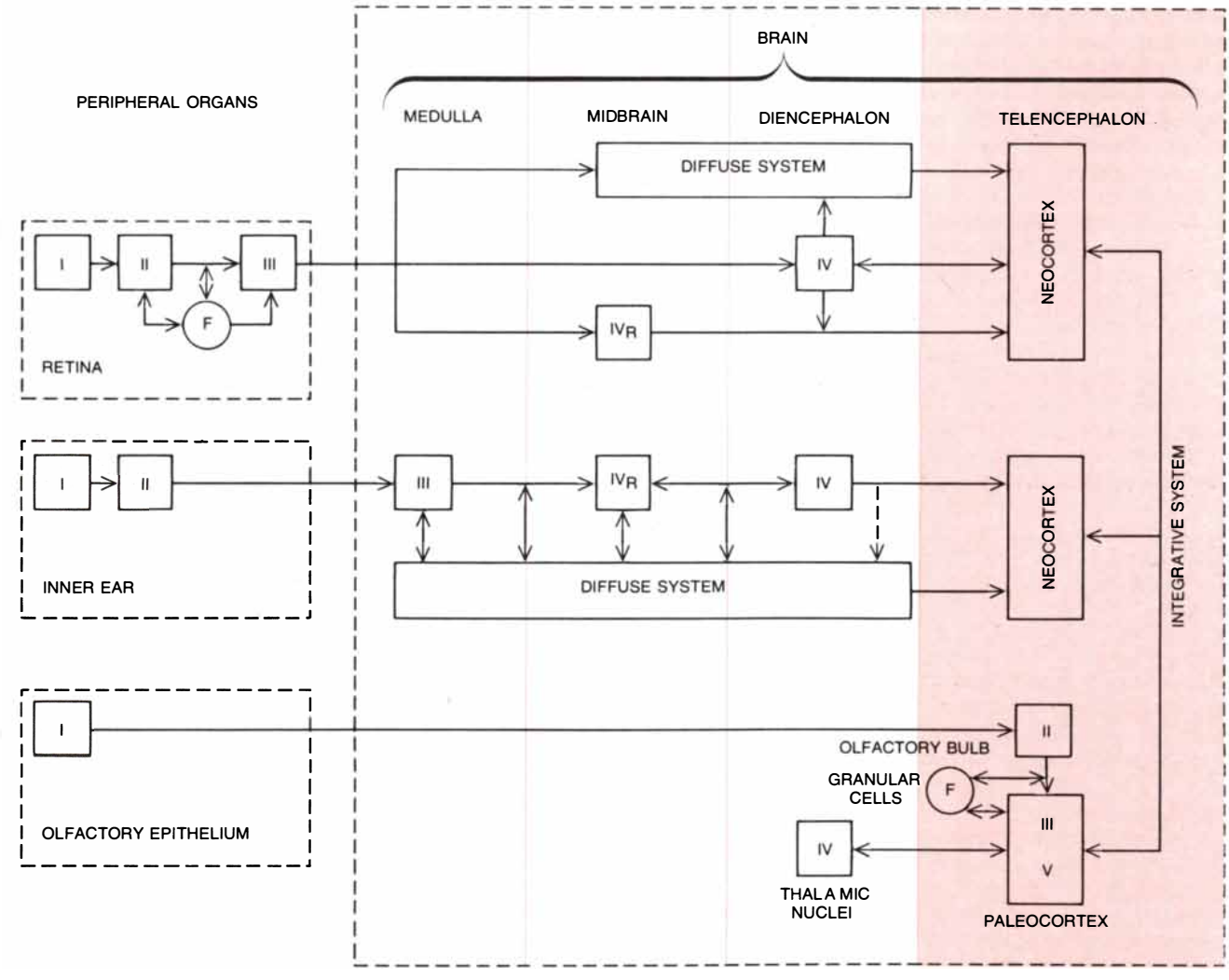
The further encephalization of the mammals beyond the archaic level involved new and peculiarly mammalian adaptations to niches that became available as a result of the extinction of dinosaurs and other ruling reptiles. These were daytime niches, and the mammalian response to them occurred in two stages. The first stage was an adaptive radiation that did not involve encephalization, in which effects were toward increases in body size; this was the archaic mammalian radiation. The second stage was a response to the new availability of daytime niches. It was inevitable that the visual system would evolve in some mammalian species and that those species would be at an adaptive advantage in daytime.

The new mammalian visual system would not be a simple retrogression to reptilian vision. Mammals had by then lost many of the reptilian visual adaptations; their normal vision was nocturnal, based on a new system of receptors (rod cells rather than cone cells) and a different analysis of the information from the receptors. Their vision must also have been much more encephalized than reptilian vision if it needed to be

integrated with auditory and olfactory information. Lost adaptations do not reappear, and it is clear from the record of the mammals that their daytime vision was based on mechanisms different from those of reptiles. Mammalian vision is represented at forebrain levels, in the thalamus and the cortex, whereas reptilian vision has its most significant central representation at the level of the midbrain and is accomplished to a great extent at a retinal level.

What was the nature of the encephalized daylight visual system in progressive Tertiary mammals? A conservative approach would assume that such a system would be modeled after the other encephalized systems of the brain—the auditory, olfactory and night-vision systems—and would therefore involve the temporal encoding of spatial information and also object formation. Because the peripheral information would already have been encoded spatially, the temporal code might result in the creation of "mental images," or memories of objects and of their spatial location. That construction would also involve other sense modalities, and so integrative systems for associating information contributed by the various sense modalities would have to be expanded. That called for still further morphological encephalization.

The step to man can be analyzed in a similar way. Several unusual selection pressures may help to explain the peculiar hominid adaptations and encephalization. It is currently accepted that the early hominids were nonarboreal primates that had invaded a niche comparable to that of predatory carnivores and that they had become modified for life in such a niche. What special problems may have faced such progressive primates, assuming that, like their living relatives, they were noisy social animals with reduced olfactory systems? Their geographic range would have been considerably more extensive than that of any other living primate, and they would have had to cope with that range with a much diminished sense of smell. Olfactory labels, with which wide-ranging social carnivores such as wolves mark a territory and map a perceptual world, would not have been available to the hominids. The development of adequate labels to mark the range may have required the further development of audi-



VISUAL, AUDITORY AND OLFACTORY SYSTEMS of living mammals are diagrammed in simplified schematic form to show that a much more significant fraction of the visual system than of the other two systems is packaged not in the brain but in a peripheral organ. The arrows show the flow of information through successive orders of nerve cells (*I through V*); stages labeled IV_R are parts of reflex control systems; *F* marks feedback loops. The structures in the telencephalon (*color*) are specifically mammalian.

tory, visual and particularly vocal capacities—the last of these acknowledging the fact that primates are noisy animals. Such a development is consistent with the evolutionary model used to analyze encephalization in other mammals. In the case of the primates it suggests the evolution of primitive language as a further solution to the problem of creating a real world that provides an adequate model for the sensory events encountered during an animal's life.

A primary model of that type postulates language as a sensory-perceptual development. The availability of vocal labels can obviously result in the capacity to communicate linguistically, but I perfer to separate the role of language in perceptual activity from its role in communication; this makes it much easier to justify the evolution of language. If there were selection pressures toward the development of language specifically for communication, we would expect the evolutionary response to be the development of "prewired" language systems with conventional sounds and symbols. Those are the typical approaches to communication in other vertebrates, and they are accomplished (as in birds) with little or no learning and with relatively small neural systems. The very flexibility and plasticity of the language systems of the human brain argue for their evolution as having been analogous to that of other sensory integrative systems, which are now known to be unusually plastic, or modifiable by early experience. (Benjamin Lee Whorf and Edward Sapir pointed out many years ago one of the maladaptive features of this flexibility of the language system, which enables different societies to develop different languages and hence different realities, often with catastrophic effects on the interactions of human communities.)

I am proposing here that the role of language in communication first evolved as a side effect of its basic role in the construction of reality. The fact that communication is so central to our present view of language does not affect the argument. It is, in fact, theoretically elegant to explain the evolution of an important novel adaptation in a species by relating it to the conservation of earlier patterns of adaptation. We can think of language as being merely an expression of another neural contribution to the construction of mental imagery, analogous to the contributions of the encephalized sensory systems and their association systems. We need language more to tell stories than to direct actions. In the telling we create mental images in our listeners that might normally be produced only by the memory of events as recorded and integrated by the sensory and perceptual systems of the brain.

Mental images should be as real, in a fundamental sense, as the immediately experienced real world. Both are constructions of the brain, although it is appropriate to encode them in order to distinguish image from reality. The role of language in human communication is special because we have the vocal and manual apparatus to create spoken and written language. In hearing or reading another's words we literally share another's consciousness, and it is that familiar use of language that is unique to man. The point, however, is that it was necessary to have a brain that created the kind of consciousness communicated by the motor mechanisms of language. That new capacity required an enormous amount of neural tissue, and much of the expansion of the human brain resulted from the development of language and related capacities for mental imagery.

The vertebrate brain evolved to control the normal range of behavior within each vertebrate species. It is reasonable to identify the brain of lower vertebrates as being adapted to control fixed-action patterns in response to specific patterns of stimulation, with few requirements for plasticity or flexibility. In the higher vertebrates, the birds and the mammals, plasticity and flexibility are evident in all living species. Yet the birds developed in their own direction, perfecting the fixed-action pattern as the basic behavioral response to environmental requirements. It is among the mammals that more flexible patterns of behavior have been the rule. If one defines intelligence as the capacity to construct perceptual worlds in which sensory information from various modalities is integrated as information about objects in space and time, the evolution of intelligence is most evident in mammals. That capacity was most elaborately developed in the primates, a group of mammals adapted toward adaptability. In the primates skeletal specialization was minimal and adaptations were more completely determined by the enlargement of the brain and the development of learned-behavior mechanisms than they were in any other vertebrates. The trend culminated in man, and we know it as the capacity for imagery, for language and for culture.

MIRAGES

There are several types of apparition to be seen across the sea or the land. Each has an explanation in terms of the optical properties of a fluid medium: the atmosphere

by Alistair B. Fraser and William H. Mach

In 1906 Robert E. Peary, striking for the North Pole, stood on the summit of Cape Thomas Hubbard at the north end of Axel Heiberg Land. To the northwest, at a distance he believed to be about 120 miles, he saw "snow-clad summits above the ice horizon." Later he saw them again, this time from Cape Columbia on Ellesmere Island, and wrote: "My heart leaped the intervening miles of ice as I looked longingly at this land and, in fancy, I trod its shores and climbed its summits, even though I knew that that pleasure could be only for another in another season." That other man turned out to be Donald B. MacMillan, leader of the expedition to "Crocker Land" in 1913. As the group approached the supposed location of Crocker Land (83 degrees north, 103 degrees west) it obligingly appeared before them. MacMillan wrote: "There could be no doubt about it. Great heavens, what a land! Hills, valleys, snow-capped peaks extending through at least 120 degrees of the horizon." They then tramped 30 miles "inland" over the Arctic ice without seeing a thing. Crocker Land was a mirage!

Both Peary and MacMillan had undoubtedly witnessed one of the most spectacular types of mirage, the Fata Morgana. It has been named after the Fairy Morgan (Fata Morgana in Italian), who appears in some of the Arthurian legends as King Arthur's sister. She was credited with the magical power of creating castles in the air. In retrospect that is an apt description of what happens, because the images seen in the Fata Morgana bear no resemblance to the object from which they were formed. Fantastic sights can appear in spite of the fact that the only object in the distance is a barren surface of snow or water.

One of the best early descriptions of the Fata Morgana was written by an Italian priest, Father Angelucci, who related his experience in a letter to a colleague. On the morning of August 14, 1643, he was looking out over the Strait of Messina from the city of Reggio on the southern tip of Italy. As he watched, "the ocean which washes the coast of Sicily rose up and looked like a dark mountain range." In front of the mountain "there quickly appeared a series of more than 10,000 pilasters which were a whitish-gray color," but then "the pilasters shrank to half their height and built arches like those of Roman aqueducts." Before it all vanished castles appeared above the aqueduct, each with towers and windows.

To properly understand how the atmosphere can give rise to such strange apparitions it is first necessary to examine the much simpler types of mirage. A particularly striking example of a common type of mirage was seen and photographed in the spring of 1972 by one of us (Fraser), who watched with fascination as two boys strolled off a beach at Seattle and casually walked out onto the waters of Puget Sound among the sailboats. The scene was so compellingly real that it was easy to believe the records that describe observations of other men apparently walking on water, examples of which occur not only in Christian writings but also in the literature of Buddhism and of the Greeks of the Hellenistic period.

FERRYBOAT "ILLAHEE" is operated on Puget Sound by the Washington State Ferry System. Appearance of Puget Sound ferries is distorted in mirages shown on the opposite page.

This simpler type of mirage differed from the Fata Morgana in that whereas the mountains of Crocker Land and the castles and aqueducts seen by Angelucci were nonexistent, the boys who were perceived walking on the water actually existed. They had, however, been seen by observation through a portion of the atmosphere that, acting as a giant lens, was bending the light rays that passed through it. The image of people walking on water was certainly real, but it was an image, not an object. One could not assume that because the image showed people walking on water the

INFERIOR MIRAGE WITH TOWERING results when temperature and temperature gradient are greatest at the surface and decrease with height. In an inferior mirage the image is displaced downward from the object. Towering means image is magnified.

TWO-IMAGE INFERIOR MIRAGE of a ferryboat appears as a series of towers. Everything below the ship's bridge has vanished. Requirements for this mirage are the same as for an inferior mirage with towering except that surface-temperature gradient is larger.

SUPERIOR MIRAGE results when the temperature increases with height so that the image is displaced up from the object. Magnified portholes and the squashed passenger deck and bridge result when boat is at *A* in the image space of diagram at the bottom of page 106.

MISSHAPEN FERRYBOAT appears to be squashed and to be on the top of a high wall. This is a superior mirage that results because the ferryboat was in approximately the position that is marked *C* in the image space of the diagram at the bottom of page 106.

WALKING ON WATER is the impression this mirage gives. The two boys to the left of the sailboat were actually walking on a sandspit that had been uncovered at low tide in Puget Sound. In this inferior mirage with towering the sandspit and the feet of the boys have vanished below the optical horizon, so that the boys appear to be suspended on the water. Because the figures have also been magnified, they appear to be larger and therefore closer than the people in the boat, whereas they were actually much farther away.

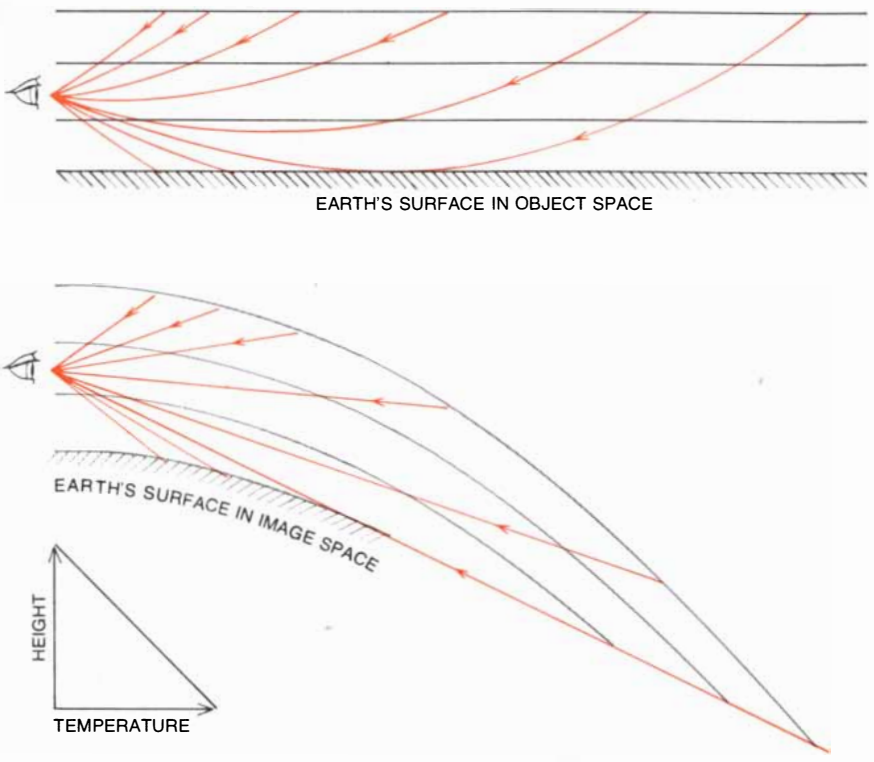
FATA MORGANA MIRAGE appears as a whitish gray wall that breaks into pilasters and, toward the right side, arches. The wall is actually a greatly magnified portion of the sea. The atmospheric lens, however, is so badly out of focus that no detail of the original sea surface can be distinguished on the wall. The same defocusing can also redistribute the brightness of the surface into bright and dark patches that are easily mistaken for real eminences such as buildings and mountains. The mirage was in Puget Sound.

people were actually doing anything of the sort. Usually in viewing one takes for granted this distinction between an image and an object. One knows that the somewhat distorted-looking person who appears on the television screen is just an image and that the object, a perfectly normal individual, is a considerable distance away in front of a television camera. Images seen through the atmospheric lens are no more illusory than the images seen through a telescope or even through a pair of eyeglasses. With the atmospheric lens or any other lens, however, the image may not look the same as the object would if it were seen without the lens.

The word mirage comes from the French verb *se mirer*, to be reflected, and although many of the images seen in mirages resemble those seen in irregular mirrors, the concept of reflection plays no part in mirages. Acting as a lens rather than as a mirror, the atmosphere produces mirages by refraction. The lens is obviously not the kind of lens that would be found in a camera or a pair of eyeglasses. Such lenses, made of glass, have a uniform index of refraction, and the light is made to bend and generate images by their curvature. In the atmosphere the lens has no shape, since both the observer and the object are inside it. The atmosphere causes light to bend as a result of gradual variations of the index of refraction within it. As part of our meteorological research at Pennsylvania State University we have examined this fascinating question of the image-forming ability of the atmospheric lens in an attempt to learn more about the structure and behavior of the atmosphere and to determine how the information carried by light is altered as it passes through the air.

The index of refraction of air depends on the density of the air and the amount of moisture in it. The contribution of moisture to phenomena involving visible light is so small, however, that it can be ignored. The density of the air depends on its temperature and pressure. Since most of the mirages we shall discuss are caused by shallow layers of air over which the change in pressure is slight, we can pretend that the index of refraction depends only on the temperature.

A high temperature corresponds to a low density and a low index of refraction. The stronger the temperature gradient (the greater the change of temperature with distance), the stronger the gradient of the index of refraction (the greater the change of the index of refraction with distance) and the greater the amount of refractive bending of the light. If the temperature in the air is the same everywhere, the light will travel in a straight line. The nature of the atmospheric lens and therefore of the mirage images it produces thus depends on the way the temperature varies in the atmosphere, primarily with height.



OBJECT SPACE AND IMAGE SPACE are contrasted. If light rays (*color*) pass through a region of the atmosphere where the change of temperature with height is constant, as indicated in the temperature profile at bottom left, the rays take a parabolic track. This is what happens in real space, that is, the space occupied by the object one is seeing (*top*). One's interpretation of what one is seeing, however, almost always rests on the assumption that light is traveling in a straight line. Therefore it is best to visualize a mirage by deforming the real space as if the rays were taking straight paths (*bottom*). Then surfaces that are actually flat are represented as being bent (*gray*). That is how space appears to the eye in a mirage; it is the "image space" in terms of which mirages are discussed in this article.

Light will take a parabolic path when it passes through a region of the atmosphere where the change of temperature with height is constant [*see illustration above*]. The curvature of the ray is proportional to the temperature gradient measured perpendicularly to the ray, so that the ray is curved most strongly when it is traveling parallel to the lines of constant temperature. The curvature of the light ray will cause an image, such as a view of a distant boat, to be displaced from the position of the object (the boat itself). Since the ray always bends so that the cold (denser) air is on the inside of the curve, the image is displaced in the direction of the warm (less dense) air.

Most mirages involve viewing over distances of anywhere from about half a kilometer to about five kilometers, so that a mirage is only weakly influenced by the curvature of the earth. For the purposes of this discussion it is therefore convenient to assume that the earth is flat. Over a flat earth any other surfaces drawn at uniform heights would also be flat. If there is a temperature gradient in the air, the light rays that pass through the atmosphere will be not straight but curved.

The space through which the rays pass is the "real" or "object" space, but it is not the space that is perceived by the eye. One's interpretation of what one sees in one's surroundings is almost always predicated on the implicit assumption that light travels in a straight line. To comprehend a mirage it is better to imagine that the "real" space has been deformed so that the rays passing through it are traveling in a straight line and the previously flat surfaces, such as the surfaces of the earth, have become bent. We now have a drawing of space as it appears to the eye; the drawing can be described as a representation of the "apparent" or "image" space. The mirages described here will be discussed mainly in terms of the image space that

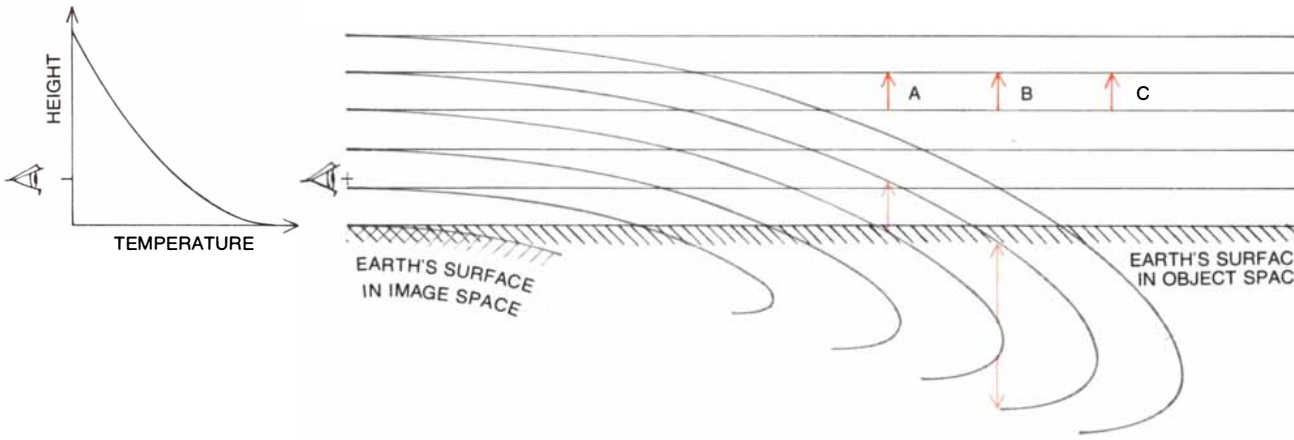
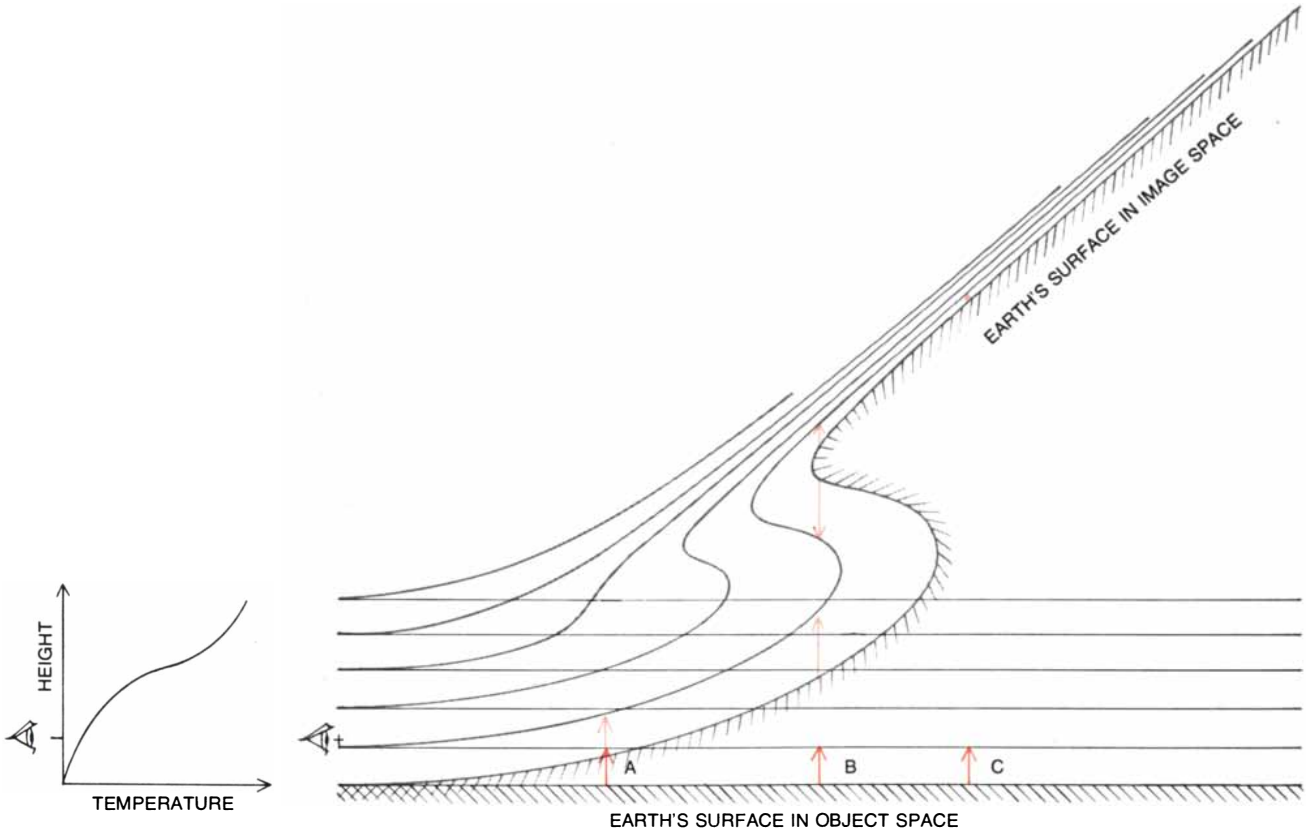


IMAGE SPACE for a two-image mirage is portrayed. When the temperature profile (*left*) has its maximum gradient where the temperature is highest, a two-image mirage can be produced. Because the maximum temperature is also at the ground, the mirage is inferior, so that all image-space surfaces (*gray*) bend down with increasing distance from the observer. Corresponding object-space surfaces (*black*) are horizontal. An object *A* would be seen in image space as a single image, displaced downward and magnified. An object *B* would be seen twice in image space, once erect and once inverted. Object *C* could not be seen unless it was a bit higher.

results from a particular type of temperature (or refraction) profile occurring in the bottom few meters of the atmosphere.

The simplest distinction that can be made is between a superior (literally upper) mirage and an inferior (lower) one. When the temperature increases with height, a horizontal surface such as a body of water will appear to be concave upward. It gives the observer, particularly one who is viewing it with binoculars, the impression of being inside a large, shallow bowl. It is a superior mirage because the image is displaced upward from the position of the object. The phenomenon was known to English sailors as looming.

An inferior mirage occurs when the temperature decreases with height. Then a horizontal surface will appear to be



THREE-IMAGE MIRAGE can result from an inflection-point temperature profile. This profile often exists over an enclosed body of water on a sunny afternoon. The apparent shape of the water surface is a large, flattened *S*. An object *B* located where the surface folds is displaced up and seen as three images. An object *A* will appear as a single image, but if it extends through the depth of the diagram, the central part will be magnified and the top reduced. An object *C* would be seen as a greatly squashed single image.

convex upward. The impression is again of a bowl, but now the bowl is inverted and the observer is on the top of it with the surface curving down in all directions. As a result there is an optical horizon beyond which the surface vanishes as it curves out of the observer's sight. This phenomenon is sometimes called sinking.

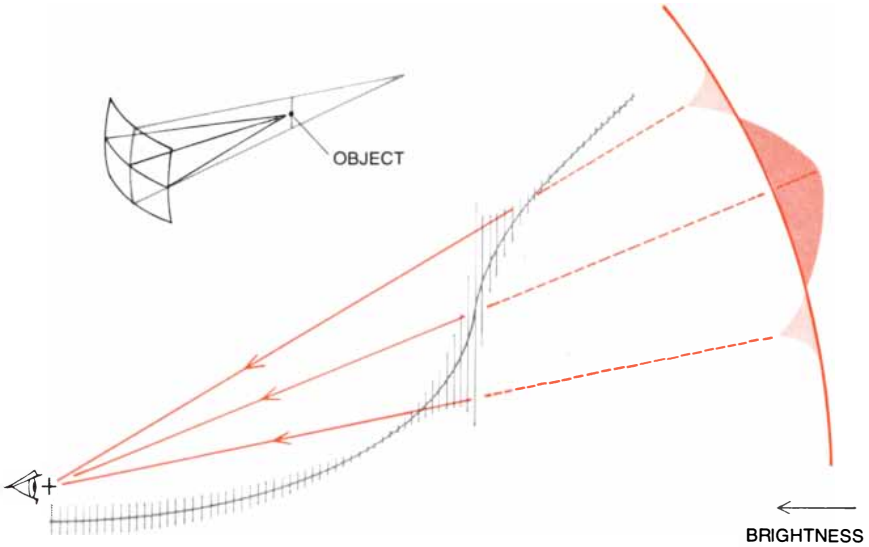
It was an inferior mirage that made those two boys appear to be walking on the waters of Puget Sound. They were actually walking on a sandspit that had been uncovered at low tide, but their feet and the sandspit were beyond the optical horizon and could not be seen. The image of the rest of the boys' bodies was left suspended on the surface of the intervening water.

The temperature profile that gives rise to this effect is quite common over enclosed bodies of water in the early morning. The water retains its heat through the night but the surrounding land cools off. Cool air from the land flows out over the warmer water and is heated from the bottom, thus creating a temperature profile in which the temperature decreases with height.

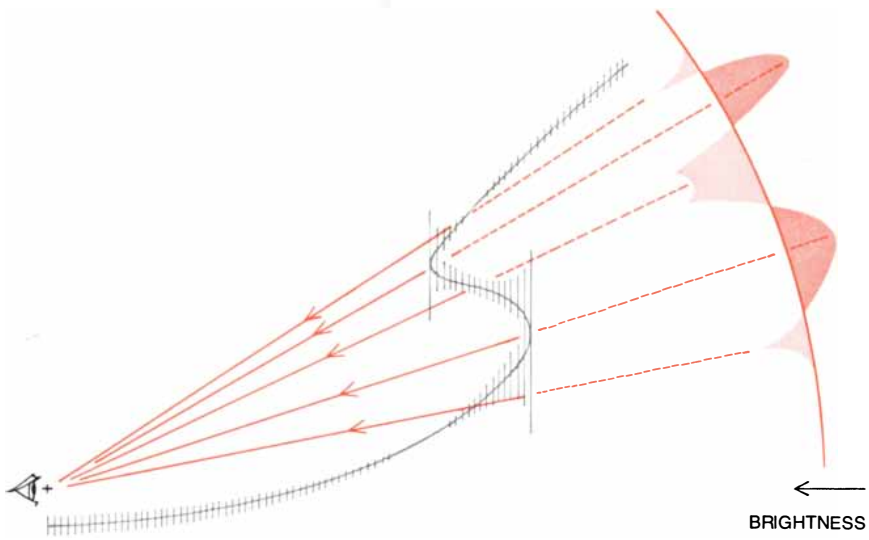
From the viewpoint of a mathematician who wants to calculate the shape of the images in image space the difference between a superior mirage and an inferior one is slight. If he assumes that the temperature increases with height, resulting in a superior mirage, the surfaces bend up with distance and the observer is apparently in the bottom of the bowl. To obtain the corresponding inferior mirage the mathematician need only turn the diagram upside down. Now the temperature decreases with height and the observer is apparently sitting on the top of an upturned bowl. The shape of the bowl is the same in both cases, but there is a world of difference to the observer. With the superior mirage he sees the inside of the bowl, whereas with the inferior one he sees the outside of the bowl and so cannot see beyond its horizon. The meteorologist would likewise claim that a worthwhile distinction can be made between the superior mirage and the inferior one because the meteorological conditions that give rise to an increase of temperature with height differ from the conditions that produce the reverse effect.

Having examined the general characteristics of superior and inferior mirages, let us now look closer at their behavior. One question is whether the image has been magnified or reduced. If the temperature gradient is constant with height, there will be no magnification. The image will have been displaced from the position of the object, but everything at a given distance is displaced up or down by the same amount. For example, a person's head will be displaced by the same amount as his feet, so that he keeps the same proportions. Actually a constant temperature gradient is rarely encountered in the bottom few meters of the atmosphere.

The primary mechanism that determines the structure of the temperature profile near the surface of the earth is the transfer of heat between the surface and the atmosphere. Near the surface, in the bottom few centimeters of the at-



CONDITION FOR THE FATA MORGANA is also an inflection-point profile, but the temperature gradient near the inflection point is slightly smaller than it is for a three-image mirage. The apparent shape of the water surface does not fold over but instead rises up and forms a wall. The surface is blurred because of astigmatism; since the wave front (*upper left*) is not spherical, objects in the distance are out of focus as a blurred vertical line. Amount of blurring is indicated below by gray vertical lines. Astigmatic blurring also redistributes the brightness (*right*) so that center of wall appears dark and other parts bright.



ASTIGMATIC THREE-IMAGE MIRAGE produces an overhanging wall so blurred that no detail can be seen on it. The brightness, however, is redistributed so that the center of the wall is bright. This strip of brightness appears to the eye as a bank of fog and has been called the Fata Bromosa. Internal gravity waves in the atmosphere can cause the image to oscillate back and forth between looking like a vertical dark wall and looking like an overhanging bright wall. The patchy distribution of brightness that results can make observers think they see a wide variety of phenomena, including mountains and elaborate buildings.

mosphere, heat is transferred mainly by molecular conduction and radiation; higher up free or forced convection usually dominates. The most efficient of these mechanisms is convection, and so it requires the smallest temperature gradient to transfer a given amount of heat. It is usual, therefore, for the temperature gradient to be at a maximum near the surface and to decrease with height. The resulting temperature profile has curvature, which changes the magnification of the image.

Stooping, the name applied to an image when it is reduced in size (has a magnification of less than unity), occurs if the temperature gradient decreases as the temperature increases. This condition frequently arises on sunny afternoons over lakes and sounds. Air that has been warmed over the surrounding land is carried out over the cooler water and so is cooled from below. The strongest temperature gradient and the lowest temperature are found at the surface of the water, so that with height the gradient decreases and the temperature increases. We are therefore dealing with a superior mirage: the image is displaced upward from the position of the object. Thus the bottom of the object, which is seen through a stronger gradient than the top is, will be lifted more than the top. The resulting image is squashed.

An observer on a beach can often watch a distant scene becoming increasingly stooped as the day progresses. The lateral dimensions do not change, because the horizontal gradient of temperature is negligible. The view of, say, a distant boathouse would therefore retain a fixed angular width while becoming increasingly small vertically.

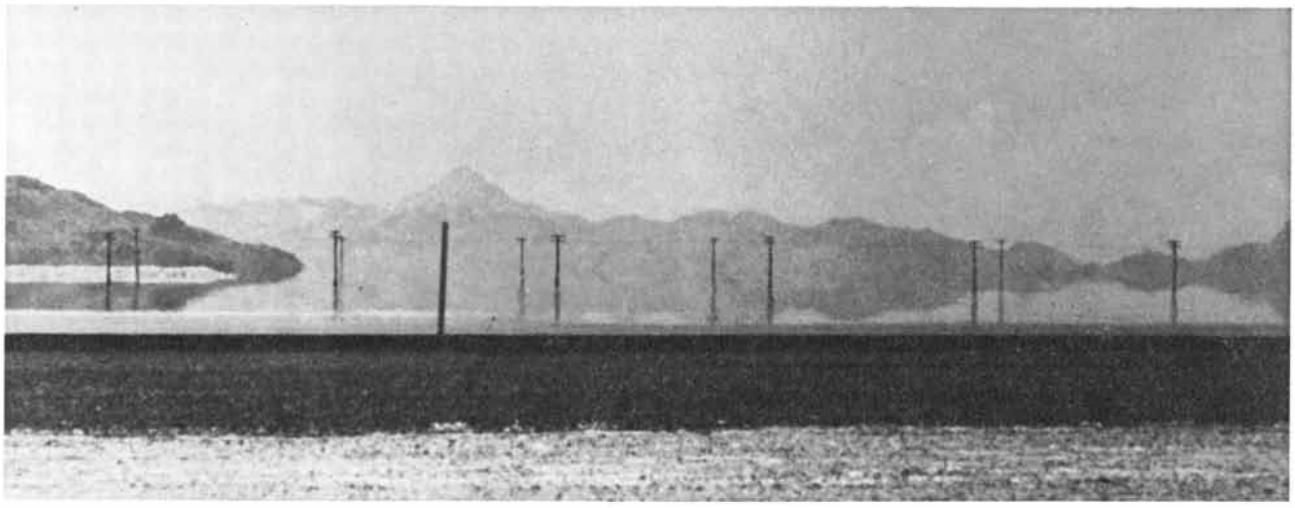
Towering, the term applied to an image when it is magnified, occurs when the temperature gradient and the temperature increase together. Towering almost always accompanies an inferior mirage, since whenever the surface is warmer than the air above it, the air will be heated from below. The maximum temperature and the maximum temperature gradient are hence found at the bottom of the temperature profile, and both quantities decrease with height. Such conditions are to be found over enclosed bodies of water in the early morning, as we have mentioned, and also over sun-heated ground later in the day. Since the temperature increases near the surface, the image of a distant object will be displaced downward, but the bottom of the object will be displaced downward more than the top because the bottom is seen through a stronger temperature gradient. The photograph of the two boys apparently walking on water illustrates the phenomenon. They both appear to be as large as or even larger than the people in the sailboat in spite of the fact that they are considerably farther from the camera. They are towering.

We turn now to the two-image mirage, which is the phenomenon involved in the familiar example of the traveler in the desert who thinks he has sighted an oasis. The "water" in such a mirage is a second (inverted) image of the sky, seen below the horizon and therefore interpreted as if the sky light were reflected by a surface of water. This is a two-image inferior mirage.

The requirements for the existence of a two-image mirage are the same as for towering: the temperature and the temperature gradient must increase together. To give rise to two images instead of a single towering one, however, the temperature profile must have a somewhat greater curvature. In an inferior mirage the effect can be accomplished by an increase in the temperature gradient at the surface of the ground or of the water. A ray of light that travels through this region of strong temperature gradient becomes so strongly bent that it will no longer be able to join the eye with the bottom of some distant object but will instead join the eye with the top of the object to give a second, inverted image. The image is inverted because as the observer lifts his gaze slightly he is looking through a region of the atmosphere that has a weaker temperature gradient, so that the ray is less strongly curved. It will therefore join the eye to a point lower on the object rather than higher, as would usually be expected.

A diagram of the image space for a two-image inferior mirage [*see top illustration on page 106*] shows that the various surfaces do not extend indefinitely into the distance. The bottom surface terminates at the optical horizon, and the other surfaces terminate at a distance that increases with the height of the surface. They vanish at the tip of a noselike fold in the surface that causes a portion of the surface to appear in the diagram twice.

As an object at the height of one of the surfaces moves farther away from the eye it will appear first as a single image, then as a double image, and finally it will vanish from view. A person walking away from you on a desert would therefore slowly vanish from the



DESERT MIRAGE is another example of a two-image inferior mirage. In spite of the strong impression of water in the distance, nothing is there but dry desert and mountains. The bottom part of the mountains has vanished; upper part appears as an inverted image.

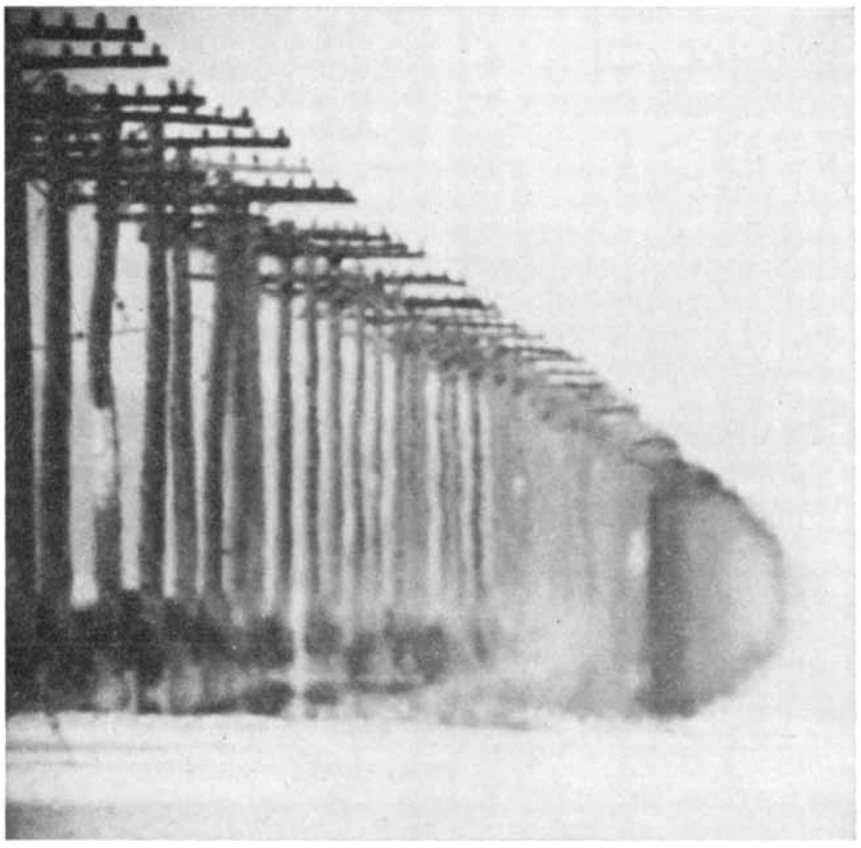
feet up. It would look as if he were wading into the sea, an impression strengthened by the seemingly reflected image of the upper portion of him that was still to be seen. He would ultimately "drown" as his entire body vanished.

The distance to the edge of the "water" is the distance to the optical horizon. It is determined by the last ray that touches the surface of the ground, that is, the ray that is tangent to the surface. As the observer moves forward or backward, so does the optical horizon. The phenomenon is familiar to every driver who watches the "water on the road" recede up the road as he approaches it. Over the flat expanse of a desert the "water" surrounds the observer and moves as he moves. With the coming of evening the desert cools, the temperature gradient decreases, the curvature of the light rays decreases and the "waters" recede from the observer.

It is sometimes difficult for the observer to decide whether the scene before him is actually a reflection from a surface of water or whether it is a mirage, particularly when the two-image inferior mirage is seen over a real water surface (as in the photograph on the cover of this issue). If it is a mirage, what is seen depends strongly on the elevation of the observer's eye. When the observer kneels down, all the surfaces terminate much closer to him, so that distant objects vanish, only to reappear when he stands up. A reflection would not behave in this way.

To obtain a three-image mirage the temperature profile must satisfy the conditions for a two-image mirage (the temperature and the temperature gradient must increase together), and in addition the rate at which the temperature gradient increases must decrease. This is roughly equivalent to saying that the curvature of the profile decreases with increasing temperature.

One example is the "elbow" profile that appears next to a sun-baked wall. When the sun shines on a long, uniform wall, the temperature of the wall rises and the heat is transferred both into the building and out to the surrounding air. A temperature gradient builds up in the first few centimeters outward from the wall. Because the change of temperature is horizontal, the image is displaced sideways, yielding a lateral mirage. An object that is along the wall but close to the eye has a single image. If it is far from the eye, it has two images, and if it is in a narrow zone at an intermediate distance, it has three images.



UTILITY POLES appear to become submerged in water with increasing distance from the observer. The effect is caused by the two-image inferior mirage and can be understood by examining the image space shown in the top illustration on page 106. With increasing distance the poles seem to vanish from the bottom up. As they do so an inverted image of a portion of the poles gives the impression of being a reflection from a surface of water. Perfectly normal utility poles appearing thus distorted here are in Great Salt Lake Desert.

Often on warm spring or summer afternoons a three-image mirage can be seen over an enclosed body of water such as a large lake, a bay or a sound. The images are caused by a temperature profile that has an inflection point. The temperature increases with height, because the warm air from the land has flowed out over the cooler water. As a result of turbulent mixing of the air as it flows over the water, the temperature gradient in the bottom few meters is small, but it increases with height at first and then decreases again. The curvature of the profile therefore decreases with height and indeed vanishes at the point where the temperature gradient is at its maximum: the inflection point.

The apparent shape of the surface of the water, as seen through this strange lens, is that of a large, flattened letter *S* [*see bottom illustration on page 106*]. An object located where the surface folds will be seen three times. A diagram of the image space reveals the remarkable way an image would change as the object moved farther away from the observer.

A temperature profile with an inflection point will not necessarily give rise to a three-image mirage. If the temperature gradient in the vicinity of the inflection point is just a little gentler, the apparent surface will not fold over but will instead rise and form a wall. An object located at the distance of the apparent wall will be seen greatly magnified. That is the simplest manifestation of the Fata Morgana.

If that were all there was to the Fata Morgana, it would have been incapable of generating the many strange images that have been credited to it, such as the castles and pilasters seen by Angelucci and the mountains of Crocker Land seen by Peary and MacMillan. An understanding of those phenomena requires a further discussion of the basic image-forming properties of the atmospheric lens.

Up to this point we have discussed the atmospheric lens in terms of its abil-

ity to produce multiple images of a distant object and to alter the magnification. Nothing has been said about how good the images are in terms of sharpness, although some of the photographs accompanying this article will reveal that the sharpness of the atmospheric lens can vary greatly. Some of the lack of sharpness can be ascribed to shimmering, which is caused by small irregularities in density and temperature that result from turbulence in the air. Also contributing to the lack of sharpness is the astigmatism of the atmospheric lens.

Astigmatism is apparent when the wave front of the light that reaches the eye is not spherical but instead is shaped like a small portion of an ellipsoid [*see upper illustration on page 107*]. A horizontal cross section through the wave front has a curvature different from that of a vertical cross section. One's ability to focus the light on the retina or on the plane of the film in a camera is determined by the curvature of the wave front, so that in this case one can focus to define the image sharply in either the horizontal or the vertical but not in both. In the atmosphere it is only possible to focus on, and thus to sharply define, the horizontal position of the image, because the curvature varies so rapidly in the vertical cross section. A discrete object will appear defocused as a fuzzy vertical line.

The astigmatic wave front gives rise to another curious effect: it can alter the brightness of the image. Consider a point source of light seen in the distance. The vertical defocusing of the image means that the light is spread into portions of the image that would otherwise have been dark. The light energy has been redistributed, leaving the apparent position of the light source darker than it would have been in the absence of astigmatism, whereas the regions above it and below it have become brighter. Because the amount of astigmatism varies greatly in different parts of the mirage, even a uniformly illuminated surface will appear to the eye as being nonuniformly bright.

That effect is the origin of the mirage phenomenon named the Fata Bromosa, or fairy fog. A perfectly flat and uniformly illuminated surface of the sea is deformed in image space to look like a slightly overhanging wall. The blurring effect of the astigmatism eliminates all detail in the "wall," and so the fact that it is an image of the ocean cannot be determined. The redistribution of brightness on the overhanging wall causes the wall to appear much whiter than its sur-

roundings, so that it looks to the eye just like a fogbank out over the water.

The water surface that appears to have an overhanging wall is the result of an inflection-point profile that gives rise to a three-image mirage. If the temperature gradient at the inflection point is decreased a little, the apparent wall no longer overhangs but stands vertical. The astigmatism will still blur detail in the wall, but now it redistributes the brightness so that the wall appears to be darker than the surrounding regions. Only one more element remains to explain Angelucci's observation of the 10,000 pilasters, and that is the presence of gravity waves in the atmosphere.

The temperature profile we have been discussing is one that causes the atmosphere to be stably stratified. If the surfaces of constant temperature (and density) become tipped, they will oscillate back and forth around the horizontal position and give rise to waves. (It is rather like water sloshing back and forth in a swimming pool.) Gravity acts as the restoring force for these waves, which derive their energy from the wind.

For the purposes of this discussion all that matters is that the waves cause the strength of the temperature gradients in our temperature profile to vary slightly in a periodic manner. Looking through one portion of the wave, the apparent shape of the water will be slightly overhanging and thus bright; looking through another portion, the surface will be vertical and thus dark. Each wave will result in the appearance of another whitish gray pilaster. A slight increase in the strength of the mean temperature gradient makes possible the appearance of a white wall with periodic dark windows in it. In fact, slight variations in the shape of the temperature profile and the amplitude of the gravity waves are all that is required to produce any of the details of Angelucci's observations or of Peary's view of "Crocker Land."

Although the simple forms of the mirage only transmit and distort images of objects that actually exist, the Fata Morgana is capable of bringing about such a great transformation that the images seen bear no resemblance to the object that gave rise to them. The Fata Morgana can take a flat, evenly illuminated surface of water, snow or ground and transform it into a wall on which all prior information has been erased. As the brightness on the wall is redistributed, new pictures are created. It is not surprising that some mirages have been credited to fairies.

Immunization against Smallpox before Jenner

Long before he introduced inoculation with cowpox, smallpox was prevented by inoculation with smallpox itself. The procedure was risky, but its spread in the 18th century set the stage for Jenner

by William L. Langer

In the 17th century smallpox, a disease that is now on the verge of eradication, was endemic everywhere in Europe and probably throughout the world. Perhaps the most infectious disease of man, it was in many respects even more loathsome and fearful than that other great killer, plague. Where plague decided between life and death within three or four days, smallpox lasted for two weeks or longer. Its first signs were fever, backache and vomiting. The fever then subsided, and many small bumps the size of birdshot appeared on the skin, particularly on the face, the chest and the arms. Over several days the bumps enlarged as they filled with fluid. The fever then returned and the bumps became inflamed and swollen pustules that broke and formed a soft yellow crust with an offensive odor. Eventually, if the patient survived (and in many epidemics two in five died), the crusts fell off, revealing the characteristic pox, or depressed scars, that gave the disease its name. It was not unusual for the patient also to be left blind.

Although smallpox most commonly attacked children under five, it did not respect age or social status. Scarcely 20 percent of the population escaped it entirely, and it was always more virulent in the cities than it was in the countryside. In 17th-century London it accounted for some 10 percent of all deaths, and it took an even higher toll in other European cities. Like plague it would assume epidemic proportions every five to 10 years. About 1660, just as plague was beginning to die out in Europe, the threat from smallpox increased. Not long afterward the intelligentsia of Europe awakened to a fact already known to common folk: this terrifying scourge was avoidable. Smallpox was in fact the first major disease that was amenable to a form of prophylaxis; it could not be cured but it could be prevented.

Today the prevention of smallpox is usually associated with the name of Edward Jenner. Actually people in many parts of the Old World had practiced one method or another of "buying the pox" long before Jenner's time. Fearful for their children's health, parents would seek out someone with a case of smallpox, preferably a mild one. The smallpox victim and the child would then make contact in such a way as to infect the child. After an incubation period of about a week the child, if he was lucky, would develop a mild case of smallpox and would emerge virtually unscarred and immune to the disease thereafter; the mild induced case gave the same protection that was provided by a severe one. Educated people came to call this practice of folk medicine inoculation, after the Latin *inoculare,* to graft. They also called it variolation; variola, from the Latin *varus,* pimple, was the scholarly name for smallpox.

There were numerous techniques of variolation. The Chinese avoided direct contact with the sick; instead the child was induced to inhale a powder made from the crusts shed by a recovering patient. In the Near East and in Africa fresh material from a diseased patient's pustules was rubbed into a cut or scratch in the skin of the person being immunized. The first scholarly account of peasant immunization practices in Europe appears to be one written by Thomas Bartholin, an anatomist at the University of Copenhagen and later physician to Christian V, king of Denmark and Norway; his note on variolation in the Danish countryside was published in 1675. At that time the practice was also known in rural France and in Wales.

In England 40 more years passed before the folk practices came to the notice of the learned. An Oxford-trained Greek physician then living in Constantinople, Emanuel Timoni, was the reporter. He wrote to the Royal Society of London in 1713 to describe to his fellow members the method of variolation used in the Near East. His note was summarized in the society's *Philosophical Transactions* in 1714. Two years later the same journal published a more detailed analysis of the Eastern practice written by an Italian physician, James Pylarini, who was then serving as the British consul in Smyrna. Variolation was practiced in Smyrna along well-established Near Eastern lines; professional variolators, usually women, took material from a ripe pustule and rubbed it into a scratch or an incision in the arm or leg of the person being inoculated.

We now come to Mary Pierrepont, a well-born English beauty who against her father's wishes eloped in 1712 with Edward Wortley Montagu, a grandson of the first earl of Sandwich. The couple lived quietly in the country until Edward became a member of Parliament for Westminster in 1715 and his wife suddenly found herself one of the most popular hostesses in London and a friend of the leading intellectuals of the capital. That same year Lady Mary contracted smallpox, to the ruin of her beauty. The following year her husband was sent as ambassador to the Turkish court in Constantinople. Lady Mary, who now had an understandable interest in smallpox, was greatly impressed by what she saw of the Turkish practice of variolation. She had the embassy physician inoculate her young son, and on her return to London in 1718 she agitated enthusiastically in favor of variolation. When her daughter reached the age of four

DEVELOPMENT OF SMALLPOX in a child is shown in this series of photographs made under the auspices of the World Health Organization for a poster designed to alert people in smallpox areas to the difference between the symptoms of smallpox and those of chickenpox. Those areas are now limited to two countries: Bangladesh and Ethiopia. The first photograph (*top left*) was made on the third day after the smallpox rash had first appeared; the remaining photographs (*left to right and top to bottom*) were made on fifth, seventh, 10th, 15th and 25th days. In the course of the infection the rash develops into pustules, which in this case have begun to break on seventh day. By 25th day the pustules have largely healed, leaving the depressed scars characteristic of the disease.

in 1721, Lady Mary made sure that she too was inoculated against smallpox. Although she was not the first to report the Eastern practice to the English, Lady Mary deserves great credit both for her courage in having her children inoculated and for her persistent propaganda in court circles.

Lady Mary and those physicians who, having read the Royal Society reports, also favored the practice met with sustained and not entirely irrational opposition. Many clergymen simply denounced variolation as an interference with God's will. The fact remains that a few of those who were inoculated did die. More important, all who were inoculated became potential sources of contagion throughout the period of their illness.

Curiously, this contagious effect had not been foreseen and might have continued to go unnoticed except for an incident in the country north of London. The physician who had variolated Lady Mary's son in Constantinople had by now returned to England and happened to variolate the daughter of a Hertford Quaker. The household staff was large, and in the course of the little girl's illness and convalescence six of the servants (who had never contracted smallpox and therefore had no immunity) came in close contact with the child. All six were stricken, and a local epidemic ensued. Thus although a certain degree of progress was achieved (George I, for example, allowed the royal grandchildren to be immunized), variolation in England remained the exception rather than the rule.

At the same time inoculation against smallpox was initiated quite independently on the other side of the Atlantic. Boston was then the intellectual center of the British colonies in America, and Cotton Mather, the eminent theologian and far-ranging scholar, was a member of the Royal Society of London. Mather read the society's *Philosophical Transactions,* and soon after the appearance of Timoni's note he wrote a friend in England to say that one of his slaves had already told him of variolation as it was practiced in Africa.

In April, 1721, a smallpox epidemic broke out in Boston. Mather tried to mobilize the physicians of the city for a campaign of immunization. With one exception the response was outspokenly negative. The exception was Zabdiel Boylston. On June 26 Boylston inoculated his six-year-old son, one of his slaves and the slave's three-year-old boy, all without any adverse effects. Boylston thereupon proceeded to inoculate more than 200 other Bostonians in spite of a vigorous newspaper and pamphlet campaign in opposition to the practice. Unlike the opposition in England, Boylston's principal adversaries were his fellow physicians. The clergy on the whole supported Mather and Boylston.

The issue was a grave one, because the 1721 epidemic was unusually severe. Almost half of the population of Boston, 5,889 of the city's 12,000 inhabitants (11,720 if one subtracts the 280 immunized by Boylston and two other physicians), contracted smallpox. Of this number 844, or every seventh victim, died. The next smallpox epidemic to strike Boston, a less severe outbreak in 1730, was met with a much more widespread inoculation program.

Meanwhile in England interest in variolation continued to decline until John Kirkpatrick, an American physician from Charleston, S.C., visited London in 1743. He brought to English medical circles details of how inoculation had successfully halted a particularly severe

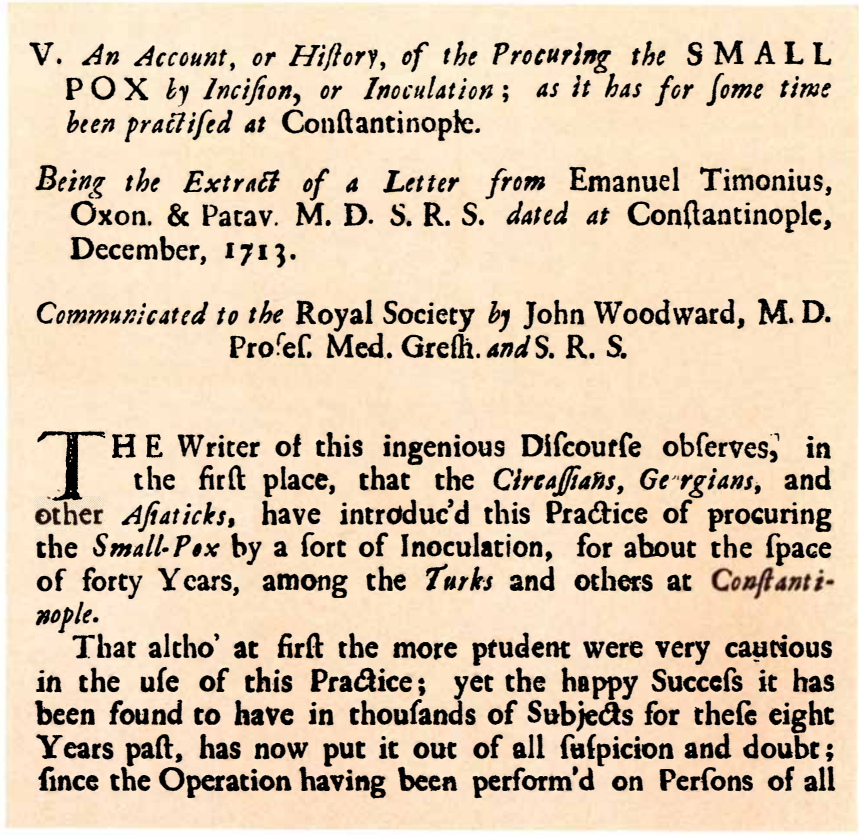
PRO-INOCULATION CARTOON published in 18th-century Russia shows a father and children (*left*) whose clear skin contrasts with the disfiguring smallpox scars of another family (*right*). The pockmarked father complains that the "pretty little children" of the other family refuse to play with his "monsters." His disfigured children scold their father for failing to inoculate them.

outbreak of smallpox in Charleston in 1738. Kirkpatrick gave credit to an improved method of variolation that had reduced the severity of the disease in inoculated patients and had lowered the mortality rate among the immunized to one in 100. (About one in 50 of Boylston's patients had died.)

Kirkpatrick's report reopened the debate over inoculation in England. In 1746 a special facility, the Smallpox and Inoculation Hospital, was founded in London with the patronage of the duke of Marlborough. Although the hospital's capacity was limited, it quickly became a center of study and teaching; many students from abroad went there to learn the techniques of variolation. In 1754 Kirkpatrick's account of the success in Charleston was published. Translated into French and other languages, it soon became virtually a catechism. The next year Robert and Daniel Sutton, father and son, began to practice variolation, first in Suffolk and then in London. The Suttons made many improvements in variolation technique, and before moving to London they reported that they had inoculated 2,514 individuals without a single fatality. Daniel opened a private clinic in the city and soon attracted an imposing clientele of well-to-do Londoners. It is reported that he earned 2,000 guineas in 1764 and 6,500 the following year.

There is nothing implausible in Sutton's financial success; by the middle of the 18th century, as one authority notes, variolation had become a "social institution." Several dozen young physicians, having learned the technique of variolation, set up practice on the Continent and made a good living; the situation was much as it is today, when golf clubs on the Continent think that in order to gain prestige they must have a Scottish professional. The clientele, however, continued to be drawn mainly from the upper classes. For example, in 1756 a Geneva physician, Theodore Tronchin, was summoned to Paris to inoculate the family of the Duc d'Orléans, second in rank only to the king. So great was the excitement that elegant carriages formed a queue at Tronchin's door, much as if their owners were attending a performance at the Comédie Française.

In spite of exceptions such as this one, inoculation in Europe remained virtually an English monopoly. When professional inoculators were called in to train a corps of local physicians in the techniques of variolation (for example, a team was called to Potsdam by Frederick the Great and another to Vienna by the Empress Maria Theresa), it was from England that they were summoned. The French did not begin to catch up with the English in this respect until the mid-1750's, after another smallpox epidemic had broken out. With some interruptions the outbreak continued until the 1790's, and during those years the disease attacked a larger percentage of adults than usual.



V. *An Account, or Hiſtory, of the Procuring the* SMALL POX *by Inciſion, or Inoculation; as it has for ſome time been practiſed at* Conſtantinople.

Being the Extract of a Letter from Emanuel Timonius, Oxon. & Patav. M. D. S. R. S. *dated at* Conſtantinople, December, 1713.

Communicated to the Royal Society *by* John Woodward, M. D. Profeſ. Med. Greſh. *and* S. R. S.

THE Writer of this ingenious Diſcourſe obſerves, in the firſt place, that the *Circaſſians, Georgians,* and other *Aſiaticks,* have introduc'd this Practice of procuring the *Small-Pox* by a ſort of Inoculation, for about the ſpace of forty Years, among the *Turks* and others at *Conſtantinople.*

That altho' at firſt the more prudent were very cautious in the uſe of this Practice; yet the happy Succeſs it has been found to have in thouſands of Subjects for theſe eight Years paſt, has now put it out of all ſuſpicion and doubt; ſince the Operation having been perform'd on Perſons of all

FIRST ACCOUNT OF INOCULATION for smallpox to appear in England was written by Emanuel Timoni and published in the *Philosophical Transactions* of the Royal Society of London in 1714. Timoni, then in Constantinople, described the method there.

Voltaire, who had survived an attack of smallpox at the age of 29, was a staunch advocate of inoculation. The prime mover among the French, however, was not Voltaire but a scientist of great eminence, Charles Marie de La Condamine, a mathematician and geographer whose fame sprang from an expedition to the Amazon in the 1730's. La Condamine went into action during the epidemic of the 1750's. In 1754 he published a number of tracts attacking the clergy's will-of-God position (an argument that appears to have been particularly influential in France) and advocated variolation as a means of controlling the epidemic. The Parlement of Paris, the chief judicial body of France, reacted to popular pressure by appointing a commission of 12 from the Faculty of Medicine of the University of Paris to investigate the issue and draft a report.

After much argument the commission split evenly for and against variolation, but when at last it reported the stalemate, the Faculty of Medicine as a whole had already come to favor this means of preventing smallpox. The practice was sanctioned by the Parlement until in 1763 a fresh epidemic in the city (urban areas were always the worst hit) was attributed to the contagion that inevitably accompanied inoculation. The Parlement responded by forbidding variolation within the Paris city limits, but the inoculators reacted to the prohibition by opening clinics just outside the capital. Among those practitioners, in addition to various English professionals and Tronchin, was a Venetian physician, Angelo Gatti, who lived in Paris and contributed substantially to the study of the disease. Five years later the Parlement rescinded its prohibition.

A roster of the better-known victims of the mid-century epidemics dramatizes the fact that smallpox did not respect the rich and powerful. In 1759 the most notable victim was the empress of Austria, Maria Theresa, stricken at the age of 52. She recovered and later engaged a Dutch physician to inoculate the mem-

bers of the imperial family. In 1774 the most notable victim was the king of France, Louis XV, stricken at 64. After two weeks of agony he died; his malodorous body was hurried to the grave at night for fear that in those troubled times the usual pomp of a royal funeral might provoke a popular uprising. His death so frightened his successor, Louis XVI, that the monarch immediately had the entire royal family inoculated.

Perhaps inspired by Maria Theresa's experience or perhaps because of the hideous appearance of her own pock-marked late husband, Czar Peter, Catherine the Great summoned a London physician, Thomas Dimsdale, to St. Petersburg in 1768 to inoculate her and her son and heir, the Grand Duke Paul. This Dimsdale did without mishap in the fall of 1768. The grateful Catherine not only bestowed the title of baron on Dimsdale but also gave him a cash gift of 10,000 pounds (and another 2,000 pounds to cover expenses) and established an annuity of 500 pounds for him. This must surely be the highest compensation paid by anyone in exchange for immunity against smallpox. In 1781 Dimsdale made a second journey to Russia to inoculate Catherine's grandsons, Constantine and Alexander. During that visit he also immunized many other members of the Russian aristocracy, both in St. Petersburg and in Moscow.

By the eve of the French Revolution the method of inoculation had been so improved that a small pinprick and a drop of thin pus inserted by lancet served the purpose. It would probably be safe to say that everywhere in Europe a majority of the well-to-do had been variolated; after all, it was an excellent investment for those who could afford it. Some students of 18th-century population growth have suggested that the widespread practice had noticeably reduced the overall mortality rate in Europe and was thus responsible, at least in part, for the phenomenal increase in population at that time. There can be no certain judgment on the point; it is impossible, first, to establish the death rate for a time before the initiation of the census and, second, to speak with any assurance of the extent of successful immunization. The suggestion does, however, raise a further question: To what extent was variolation practiced among urban and rural workers? The urban poor in particular were the traditional incubators of endemic disease, and they made up the vast majority of the population in all cities. If they were not inoculated, there could be no hope of eradicating smallpox.

The well-to-do were not blind to the problem. Public clinics were opened and in some instances offered free inoculation for children. In England, Jonas Hanway, a well-known philanthropist and a trustee of the Foundling Hospital, suggested to his peers that they employ no servants who had not had smallpox or been inoculated against it. The younger Sutton, whose clinic was perhaps the most elegant in London, offered to variolate 400 children, 100 at a time, and care for them for a full month afterward if rich patrons would subscribe a guinea per head. Nothing came of his proposition, presumably for lack of subscribers.

As for the urban poor themselves, they seem to have shown no great interest in the issue. Naturally suspicious of approaches by their social superiors, they may also have held the pessimistic view that smallpox was an integral part of life. Perhaps they had no reason to think otherwise. Those of them who favored variolation would have looked to itinerant inoculators or even to amateurs for treatment. Such practitioners could not have been relied on to do an up-to-date job. More important, it would have been impossible in the urban slums to isolate the newly inoculated. Inoculated children would not have felt the effects for a week or so and would have continued to be in contact with others. It is reasonable to suppose they would thus have infected other children, who would then have come down with a full-blown case of the disease. If new contagion of this kind was commonplace, it is not hard to understand why many believed smallpox was constantly gaining ground. Contemporary writings contain eloquent testimony to a veritable panic over the disease that seized all Europe in the last quarter of the 18th century.

Much later, in 1852, Thackeray, recalling his life early in the 19th century, noted that at that time hundreds died or rose from their sickbed terribly scarred and disfigured. "In my early days," he wrote in *Henry Esmond*, "this pestilence would enter a village and destroy half the inhabitants: at its approach, it may

Their Ages.	Perſons inoculated.	Had the Small-Pox by Inoculation.	Had an Imperfect ſmall Pox.	Had no Effect.	Suſpected to have died of Inoculation.
From 9 months to 2 years old.	06	06	00	00	00
2 to 5	14	14	00	00	00
5 to 10	16	16	00	00	00
10 to 15	29	29	00	00	00
15 to 20	51	51	00	00	01
20 to 30	62	60	00	02	01
30 to 40	44	42	00	02	01
40 to 50	08	07	00	01	00
50 to 60	07	06	00	01	02
60 to 67	07	07	00	00	01
Total	244	238	00	06	06
Inoculated by Drs. *Roby* and *Thompſon* in *Roxbury* and *Cambridge*.	36	36	00	00	00
Total	280	274	00	00	06

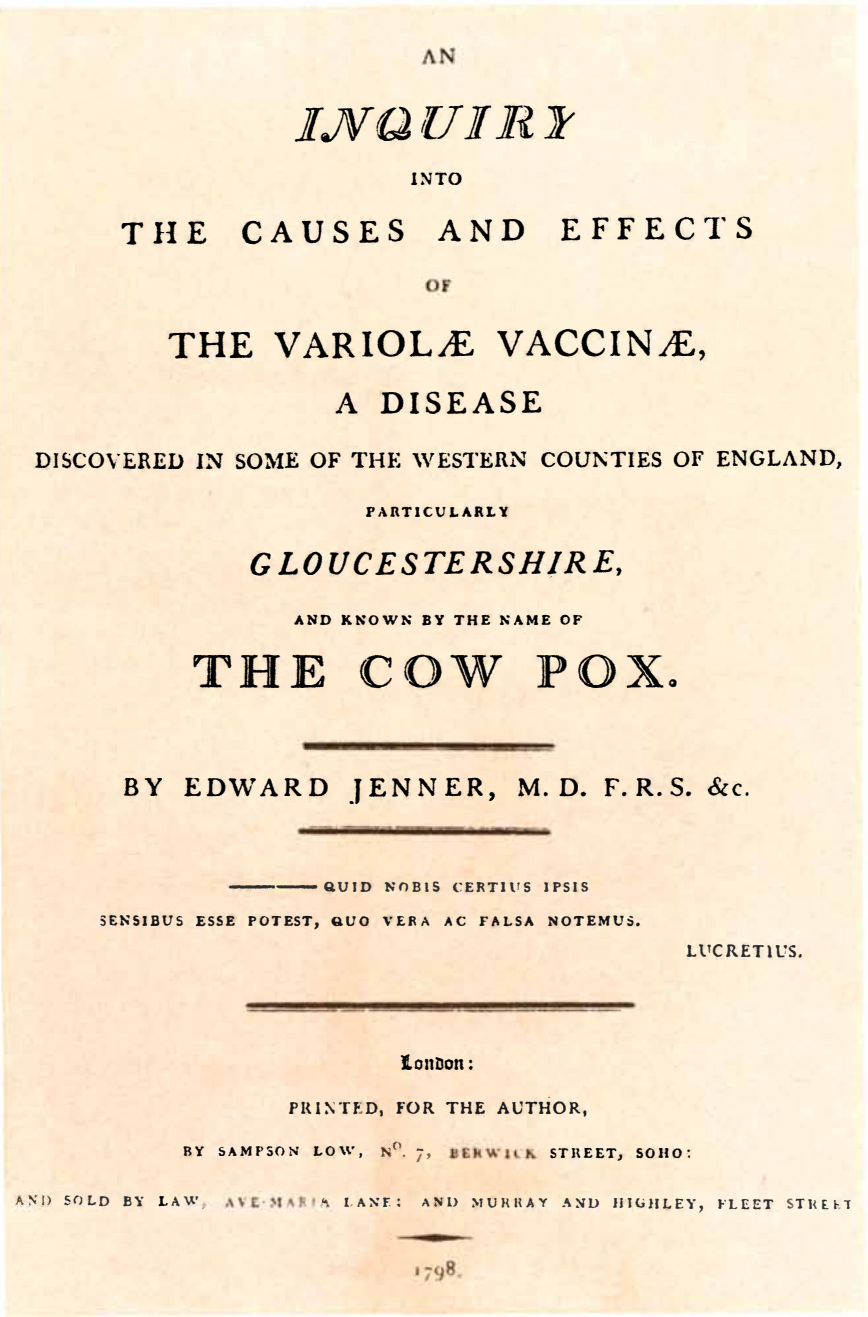
RECORD OF INOCULATIONS performed in Boston and vicinity by Zabdiel Boylston and two colleagues during the smallpox epidemic of 1721 appears in an account published by Boylston. Those inoculated numbered 280 aged between nine months and 67 years. Six of them died of smallpox. Six others had no reaction; they may already have had the disease.

well be imagined, not only the beautiful but the strongest were alarmed and those fled who could." There is more than enough contemporary evidence to show that Thackeray's remarks were not exaggerated.

Under the circumstances many physicians sought ways to prevent the infection that resulted from variolation. Jenner, a physician with a preference for a country practice, noted in the 1790's that the dairymaids in his neighborhood rarely, if ever, contracted smallpox, even in the absence of inoculation. After years of careful observation and experiment he published in 1798 a 70-page pamphlet: "An Inquiry into the Causes and Effects of the Variolae Vaccinae." (Variolae vaccinae was cowpox, the bovine form of smallpox.) In this classic report Jenner advanced the notion that inoculation with cowpox, which he considered a mild form of smallpox, would provide the same immunity to smallpox that variolation did. What was more important, the disease produced by vaccination, as opposed to variolation, would be so extremely mild that the vaccinated individual would not be a source of infection.

No better news could have reached the world in the midst of the Napoleonic Wars. Jenner's thesis was accepted at once by the great majority of physicians, and within a year or two thousands had been vaccinated. The news raced around the world faster than news had ever traveled before; Jenner was hailed everywhere as mankind's greatest benefactor, and his discovery was characterized as the most important medical advance of all time.

It is not my purpose here even to touch on the history of vaccination in the 19th and 20th centuries except to note that for all its benefits the procedure was not quite as simple and the immunity it conferred was not as long-lived as Jenner and the early vaccinators had supposed. Nonetheless, it seems appropriate to close this review of the antecedents of Jenner's great discovery with a note on its instant adoption in the U.S.



AN

INQUIRY

INTO

THE CAUSES AND EFFECTS

OF

THE VARIOLÆ VACCINÆ,

A DISEASE

DISCOVERED IN SOME OF THE WESTERN COUNTIES OF ENGLAND,

PARTICULARLY

GLOUCESTERSHIRE,

AND KNOWN BY THE NAME OF

THE COW POX.

BY EDWARD JENNER, M. D. F. R. S. &c.

——— QUID NOBIS CERTIUS IPSIS
SENSIBUS ESSE POTEST, QUO VERA AC FALSA NOTEMUS.
LUCRETIUS.

London:

PRINTED, FOR THE AUTHOR,

BY SAMPSON LOW, N°. 7, BERWICK STREET, SOHO:

AND SOLD BY LAW, AVE-MARIA LANE: AND MURRAY AND HIGHLEY, FLEET STREET

1798.

INOCULATION WITHOUT CONTAGION, using cowpox pustules instead of smallpox pustules as the source of inoculation material, was proposed by Edward Jenner in 1798. The title page of his pamphlet on the subject is shown here. The quotation is from *De rerum natura*, by Lucretius, a Roman poet of the first century B.C., who suggests that in distinguishing between truth and falsehood man has no more certain criterion than his senses.

At the close of the 18th century a highly imaginative (although somewhat eccentric and testy) Harvard professor of medicine, Benjamin Waterhouse, played a role not unlike the one played by Mather and Boylston in the opening decades of the century. Waterhouse had been educated in Europe and maintained a regular correspondence with his friends there. Having received a copy of Jenner's pamphlet almost as soon as it could arrive by ship, Waterhouse was at once converted to Jenner's view. He wrote to friends in England asking for some of the cowpox material used for vaccination. It arrived in July, 1800, and he immediately vaccinated his children and servants.

Waterhouse promptly notified Thomas Jefferson, who was then vice-president, of Jenner's discovery. That always inquisitive statesman at once sought Waterhouse's good offices in securing an ample supply of the vaccine. Jefferson and his son thereupon vaccinated their entire household of 60, slaves included. As president, Jefferson went on to use his influence to establish the new procedure in Philadelphia and the other cities of the young republic. Before Jefferson's retirement to Monticello in 1809 the number of people who had been vaccinated on both sides of the Atlantic could be reckoned in the millions. The campaign to eradicate smallpox, now coming to a close after nearly two centuries, had begun in earnest.

MATHEMATICAL GAMES

A breakthrough in magic squares, and the first perfect magic cube

by Martin Gardner

"In my younger days, having once some leisure (which I still think I might have employed more usefully), I had amused myself in making... magic squares."

—From a letter by Benjamin Franklin

Two major breakthroughs have been made in the study of magic squares and magic cubes: all order-5 magic squares have been counted, and the first perfect magic cube has been constructed. I am pleased to be the first to publish both results here. So that the magnitude of the two achievements can be fully appreciated, let us take a brief look at the history of magic squares.

Although some of the greatest mathematicians have done work on magic squares, and even though such work leads into the theories of groups, lattices, Latin squares, determinants, partitions, matrices, congruence arithmetic and other nontrivial areas of mathematics, the most enthusiastic square constructors have been amateurs. The famous Franklin square, an ingenious 16-by-16 matrix that Benjamin Franklin called "the most magically magical of any magic square ever made by any magician," is itself the topic of many articles and monographs. The literature on magic squares in general is vast, and most of it was written by laymen who had become hooked on the elegant symmetries of these interlocking number patterns.

A standard magic square, as few readers need to be told, is a square array of positive integers from 1 through N^2, arranged so that the sum of every row, every column and each of the two main diagonals is the same. N is the "order" of the square. It is easy to see that the magic constant is the sum of all the numbers divided by N. The formula is $(1 + 2 + 3 + \ldots + N^2)/N = \frac{1}{2}(N^3 + N)$.

The trivial square of order 1 is simply the number 1, and of course it is unique. It is equally trivial to prove that no order-2 square is possible.

There are eight ways to arrange the digits 1 through 9 in an order-3 array that is magic. It is traditional, however, not to count rotations and reflections. When they are excluded, the order-3 square is unique. To appreciate the gemlike beauty of this most ancient of all combinatorial curiosities, consider all the ways that its constant, 15, can be partitioned into a triplet of distinct positive integers. There are exactly eight:

$$9 + 5 + 1$$
$$9 + 4 + 2$$
$$8 + 6 + 1$$
$$8 + 5 + 2$$
$$8 + 4 + 3$$
$$7 + 6 + 2$$
$$7 + 5 + 3$$
$$6 + 5 + 4$$

Now, in the order-3 square each of eight lines of three numbers must total 15: the six orthogonals (rows and columns) and the two diagonals. The eight lines exactly match the number of triplets we have available. Since the center number belongs to a row, a column and both diagonals, it clearly must be a digit that appears in four of the eight triplets. The only such digit is 5. We therefore know that 5 is the central number.

Consider 9. It belongs to only two triplets, so that we cannot place it in a corner since each corner cell belongs to three lines. Consequently it must go in a side cell. Because of the square's symmetry it does not matter which side cell we choose, and so let us put it above the 5. For the top corners, on each side of 9, we have no choices except 2 and 4. Again it does not matter which digit goes where, since one arrangement is merely a mirror reflection of the other. The rest of the square follows automatically. We have by this simple construction proved its uniqueness.

The completed square, in the form shown in the top illustration on page 120, is the *Lo shu* of ancient China. According to legend the pattern was first revealed on the shell of a sacred turtle that crawled out of the Lo River in the 23rd century B.C., but today's Chinese scholars trace references to it back no further than the fourth century B.C. From then until the 10th century the pattern was a mystical Chinese symbol of enormous significance. The even numbers were identified with yin, the female principle, and the odd numbers with yang, the male principle. The central 5 represented the earth, around which, in evenly balanced yin and yang, were the other four elements: 4 and 9 symbolizing metal, 2 and 7 fire, 1 and 6 water, and 3 and 8 wood. For this and other historic interpretations of the *Lo shu*, and its connection with divination and the *I Ching*, see "The Magic Square of Three in Old Chinese Philosophy and Religion," by Schuyler Cammann, in *History of Religions*, Vol. 1 (Summer, 1961), pages 37–80, and "Old Chinese Magic Squares," by Schuyler Cammann, in *Sinologica*, Vol. 7 (1962), pages 14–53.

There are 880 magic squares (excluding rotations and reflections) of order 4. They were first given by Bernard Frénicle de Bessy in 1693. There are many ways to classify them. One of the best was devised by Henry Ernest Dudeney, who explains his system in an excellent article on magic squares in early printings of the 14th edition of the *Encyclopædia Britannica*. The last printing of that edition substitutes for Dudeney's article a superb historical article by Cammann. The current (15th) edition has a trivial microarticle on magic squares in the *Micropædia*.

How many magic squares are of order 5? The best estimate was given by Albert L. Candy in his *Construction, Classification and Census of Magic Squares of Order Five*, privately published in Lincoln, Neb., in 1938. Candy arrived at a total of 13,288,952. The exact number was not known until 1973, when the counting was completed by a computer program developed by Richard Schroeppel, a mathematician and computer programmer at Information International. The program, using a standard backtracking procedure, consists of about 3,500 "words" and took about 100 hours of running time on a PDP-10. A final report, written by Michael Beeler, was issued last October.

Candy's estimate was low by a wide margin. Not counting rotations and reflections, there are 275,305,224 magic squares of order 5. Schroeppel prefers to divide that number by 4 and give the total as 68,826,306. The reason is that in

addition to the eight variants obtained by rotation and reflection there are four other variants generated by the following two transformations, which also preserve magic:

1. Exchange the left and right border columns, then exchange the top and bottom border rows.

2. Exchange rows 1 and 2 and rows 4 and 5. Then exchange columns 1 and 2 and columns 4 and 5.

When these two transformations are combined with the two reflections and four rotations, the result is $2 \times 4 \times 2 \times 2 = 32$ forms that can be called isomorphic. With this definition of isomorphic the count becomes 68,826,306.

That number can be lowered even more by considering another well-known transformation. If every number in a magic square is subtracted from $N^2 + 1$ (in this case 26), the result, called the complement, is also magic. When the center of an order-5 square is 13, the complement is isomorphic with the original. If it is not 13, a different square results. If we broaden the term isomorphic to include complements, the count of order-5 squares drops to about 35 million.

The task of classifying order-5 squares in meaningful ways is staggering. Dudeney once wrote that certain ways of dividing magic squares into types seemed to him about as useful as dividing people into those who take snuff and those who do not. Nevertheless, certain divisions yield unexpected results. Consider, for example, the total number of order-5 squares with centers that are numbers 1 through 13:

1. 1,091,448
2. 1,366,179
3. 1,914,984
4. 1,958,837
5. 2,431,806
6. 2,600,879
7. 3,016,881
8. 3,112,161
9. 3,472,540
10. 3,344,034
11. 3,933,818
12. 3,784,618
13. 4,769,936

Note that the totals steadily increase from 1 through 8 but that the totals for 9 and 10 and for 11 and 12 are reversed. That there are more squares with a center of 11 than squares with a center of 12, and more squares with a center of 9 than squares with a center of 10, came as a surprise. Of course, these same anomalies occur in the counts of squares with centers 14 through 25, since every square with a center that is not 13 has its complement. There are as many squares with 1 in the center as there are with 25, and the same is true for all numbers except 13. Readers interested in the details of Schroeppel's program may write to him at 835 Ashland Avenue, Santa Monica, Calif. 90405.

The illustration below shows an or-

1	15	24	8	17
23	7	16	5	14
20	4	13	22	6
12	21	10	19	3
9	18	2	11	25

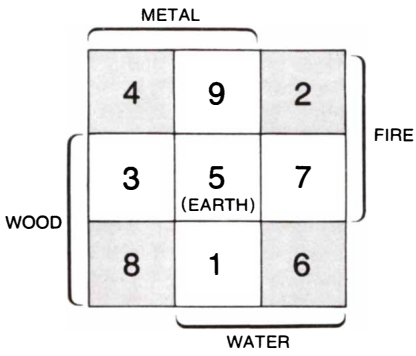
∞

1	15	24	8	17	1	15	24	8	17	1	15	24	8	17
23	7	16	5	14	23	7	16	5	14	23	7	16	5	14
20	4	13	22	6	20	4	13	22	6	20	4	13	22	6
12	21	10	19	3	12	21	10	19	3	12	21	10	19	3
9	18	2	11	25	9	18	2	11	25	9	18	2	11	25
1	15	24	8	17	1	15	24	8	17	1	15	24	8	17
23	7	16	5	14	23	7	16	5	14	23	7	16	5	14
20	4	13	22	6	20	4	13	22	6	20	4	13	22	6
12	21	10	19	3	12	21	10	19	3	12	21	10	19	3
9	18	2	11	25	9	18	2	11	25	9	18	2	11	25

∞ ∞

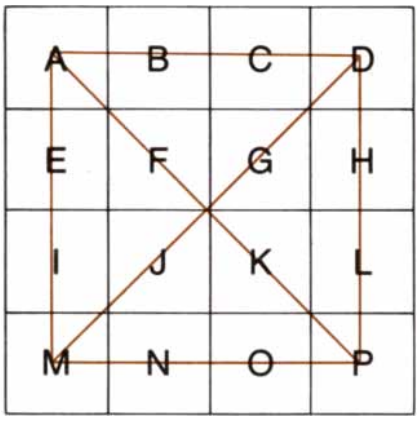
∞

An associative and pandiagonal magic square of order 5 (top) *and its cyclic permutations* (bottom)

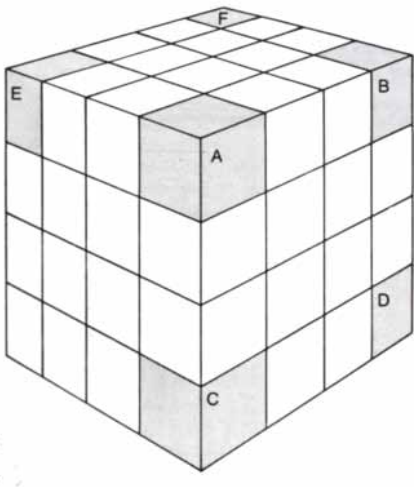


The Lo shu *magic square of ancient China*

der-5 square of a type that is more powerfully magic than any other. It is associative, which means that any pair of numbers symmetrically opposite the center add up to $N^2 + 1$. And it is pandiagonal (sometimes called Nasik or diabolic), which means that its broken diagonals add up to 65, the constant. To put it another way, if we tile the plane with this square, we can outline a 5-by-5 square anywhere on this infinite pattern and it will be magic, although not necessarily associative. To be associative too it must have 13 in the center.



Richard Schroeppel's proof, lemma 1



Schroeppel's proof, lemma 2

The *Lo shu* is associative but not pandiagonal. An order-4 square may be pandiagonal or associative but not both. The order-5 square is the smallest one that can have both properties. Excluding rotations and reflections, 3,600 order-5 squares are pandiagonal, or if we also exclude variants obtained by the cyclic permutation of rows and columns, 144 are pandiagonal. In other words, there are 144 infinite patterns of the type shown here, each containing 25 pandiagonal order-5 squares. Of the 144, just 16 contain a square that is also associative. All of this, by the way, was known before Schroeppel developed his computer program.

Of the 16 associative pandiagonal squares of order 5, four have 1 in the first cell, four have 1 in the third cell, four have 1 in the seventh and four have 1 in the eighth. The medieval Moslems were particularly intrigued by pandiagonal squares with 1 in the center. The patterns were not, of course, associative, but the Moslems thought of the central 1 as being symbolic of the unity of Allah. Indeed, they were so awed by that symbol that they often left blank the central cell on which 1 was supposed to go.

It is natural to extend the concept of magic squares to three dimensions and even higher ones. A perfect magic cube is a cubical array of positive integers from 1 to N^3 such that every straight line of N cells adds up to a constant. These lines include the orthogonals (the lines parallel to an edge), the two main diagonals of every orthogonal cross section and the four space diagonals. The constant is $(1 + 2 + 3 + \ldots + N^3)/N^2 = \frac{1}{2}(N^4 + N)$.

There is, of course, a unique perfect cube of order 1, and it is trivially true that there is none of order 2. Is there one of order 3? Unfortunately 3 does not quite make it. I do not know who first proved the impossibility, but Richard Lewis Myers, Jr., has a simple way of doing it. Consider any 3-by-3 cross section. Let A, B, C be the numbers of the first row, D, E, F the numbers of the third row and X the central number. Since the diagonals and the center column each must add up to 42, we can write:

$$3X + A + B + C + D + E + F = 3(42)\,.$$

From this we subtract $A + B + C + D + E + F = 2(42)$ to get $3X = 42$, and $X = 14$. Since 14 cannot be the center of every cross section, the cube is impossible.

Annoyed by the refusal of such a cube to exist, magic-cube buffs have relaxed the requirements to define a species of semiperfect cube that apparently does exist in all orders higher than 2. These are cubes where only the orthogonals and four space diagonals are magic. Let us call them Andrews cubes, since W. S. Andrews devotes two chapters to them in his work *Magic Squares and Cubes* (1917). (There is a Dover paperback reprint.) The order-3 Andrews cube must be associative, with 14 in its center. John R. Hendricks has proved (*Journal of Recreational Mathematics*, Vol. 5, January, 1972, pages 43–50) that there are four such cubes, not counting rotations and reflections. All are given by Andrews, although he seems not to have realized that they exhaust all basic types.

No perfect cube of order 4 exists. As far as I know the first proof was published by Schroeppel in a memorandum in 1972. The first step is to show that on any 4-by-4 section (orthogonal or diagonal) the four corners must add up to the constant. Let Q be the constant and label the 16 cells with other letters [*see middle illustration on this page*]. The colored lines indicate six quadruplets that catch all 16 cells. Since each corner cell is common to three lines, $3A + 3D + 3M + 3P$ plus each of the other cells taken once must equal $6Q$. If we subtract from this the values of the four rows, we are left with $2A + 2D + 2M + 2P = 2Q$, which reduces to $A + D + M + P = Q$, our first lemma.

Now consider the cube's eight corners. We prove that any two corners connected by an edge must have a sum of $Q/2$. Call the corners A and B. Let C, D and E, F be the corners of any two edges parallel to A, B [*see bottom illustration on this page*]. $ABDC$, $EFBA$ and $EFDC$ are each the corners of a 4-by-4 cross section, so that their total is $3Q$. Gather the like terms:

$$2A + 2B + 2C + 2D + 2E + 2F = 3Q\,.$$

Divide each side by 2:

$$A + B + C + D + E + F = 3Q/2\,.$$

From this we subtract $C + D + E + F = Q$ to obtain $A + B = Q/2$, our second lemma.

Now consider corner B. It is joined to corners A, D, F. Since $A + B = F + B = D + B$, we can take B from each equality

1

19	497	255	285	432	78	324	162
303	205	451	33	148	370	128	414
336	174	420	66	243	273	31	509
116	402	160	382	463	45	291	193
486	8	266	236	89	443	181	343
218	316	54	472	357	135	393	107
185	347	85	439	262	232	490	12
389	103	361	139	58	476	214	312

2

134	360	106	396	313	219	469	55
442	92	342	184	5	487	233	267
473	59	309	215	102	392	138	364
229	263	9	491	346	188	438	88
371	145	415	125	208	302	36	450
79	429	163	321	500	18	288	254
48	462	196	290	403	113	383	157
276	242	512	30	175	333	67	417

3

306	212	478	64	141	367	97	387
14	496	226	260	433	83	349	191
109	399	129	355	466	52	318	224
337	179	445	95	238	272	2	484
199	293	43	457	380	154	408	118
507	25	279	245	72	422	172	330
412	122	376	150	39	453	203	297
168	326	76	426	283	249	503	21

4

423	69	331	169	28	506	248	278
155	377	119	405	296	198	460	42
252	282	24	502	327	165	427	73
456	38	300	202	123	409	151	373
82	436	190	352	493	15	257	227
366	144	386	100	209	307	61	479
269	239	481	3	178	340	94	448
49	467	221	319	398	112	354	132

5

381	159	401	115	194	292	46	464
65	419	173	335	510	32	274	244
34	452	206	304	413	127	369	147
286	256	498	20	161	323	77	431
140	362	104	390	311	213	475	57
440	86	348	186	11	489	231	261
471	53	315	217	108	394	136	358
235	265	7	485	344	182	444	90

6

492	10	264	230	87	437	187	345
216	310	60	474	363	137	391	101
183	341	91	441	268	234	488	6
395	105	359	133	56	470	220	314
29	511	241	275	418	68	334	176
289	195	461	47	158	384	114	404
322	164	430	80	253	287	17	499
126	416	146	372	449	35	301	207

7

96	446	180	338	483	1	271	237
356	130	400	110	223	317	51	465
259	225	495	13	192	350	84	434
63	477	211	305	388	98	368	142
425	75	325	167	22	504	250	284
149	375	121	411	298	204	454	40
246	280	26	508	329	171	421	71
458	44	294	200	117	407	153	379

8

201	299	37	455	374	152	410	124
501	23	281	251	74	428	166	328
406	120	378	156	41	459	197	295
170	332	70	424	277	247	505	27
320	222	468	50	131	353	111	397
4	482	240	270	447	93	339	177
99	385	143	365	480	62	308	210
351	189	435	81	228	258	16	494

Cross section of Richard Lewis Myers' magic cube of order 8. (Computer printout courtesy of William Gosper.)

to prove that $A = F = D$. That is impossible, and so our proof is complete.

Is there a perfect magic cube of order 5? No one knows. Schroeppel has made a beginning by proving (using algebra and combinatorial thinking) that if such a cube exists, its center must be 63. It is not known if there are magic cubes of orders 6 and 7.

There are perfect magic cubes of order 8. A method of constructing them by the millions was discovered in the spring of 1970 by Myers, when he was 16 and a student at William Tennant High School in Johnsville, Pa. He sent me a short note about it, saying that he had obtained his first cube "after three months, seven theories and 31 sheets of graph paper." I am embarrassed to confess that I did not appreciate the significance of his claim. He did not send an actual cube, and I replied by suggesting a mathematical journal that could evaluate his work.

I next heard about Myers' cubes in December, 1972, from John H. Staib, a mathematician at Drexel University in Philadelphia, where Myers had enrolled as a freshman. Staib sent an order-8 cube [*see illustration on preceding page*], and although he extolled its symmetries and provided hints about how Myers had constructed it (by superposing three Latin cubes and applying base-8 notation), I still failed to comprehend the cube's importance. I had glanced too quickly at Andrews' book, noting references to magic cubes of order 3 and higher, without realizing that those cubes were only semiperfect. It was not until I started researching this report that I awoke to a full realization of what Myers had done.

Myers dropped out of college after his freshman year and is now working as a computer programmer for the International Computaprint Corporation in Fort Washington, Pa. He hopes to save enough money to enable him to return to a college where he can continue his education in mathematics.

Every orthogonal and diagonal line of eight numbers on the Myers cube shown on the preceding page, including the four space diagonals, add up to 2,052. The cube is associative: any two numbers symmetrically opposite the center add up to 513. It follows that not only do the eight corner cells of the cube total 2,052 but also (as Staib pointed out) the corners of every rectangular solid centered in the cube add up to the same number. If that is not enough, the cube can be sliced into 64 order-2 cubes, and the eight numbers in each of these cubes add up to the constant!

These remarkable symmetries make possible an enormous number of rearrangements of the cube, all in a sense isomorphic, and of course every arrangement can be rotated and reflected 48 ways. Imagine this cube with each of its 512 cells replaced by the same cube in any of its thousands of rearrangements or orientations. In cell 1 we put a cube that starts with 1. In cell 2 we put a cube that starts with $8^3 + 1 = 513$, in cell 3 we put a cube that starts with $(2 \times 8^3) + 1 = 1{,}025$, and so on for the other cells. The result is a perfect magic cube of order 64. Order-64 cubes in turn will build perfect magic cubes of order 512, and the same is true for all higher orders that are powers of 8.

How many order-8 magic cubes are there? By choosing different Latin cubes to superpose, Myers can construct millions different from one another and from the one given here, although not all will be associative. The number of Latin squares of order 8 have been counted (there are billions), but the number of Latin cubes has not, so that the problem of counting just the order-8 cubes that can be generated by Myers' procedure alone is horrendous.

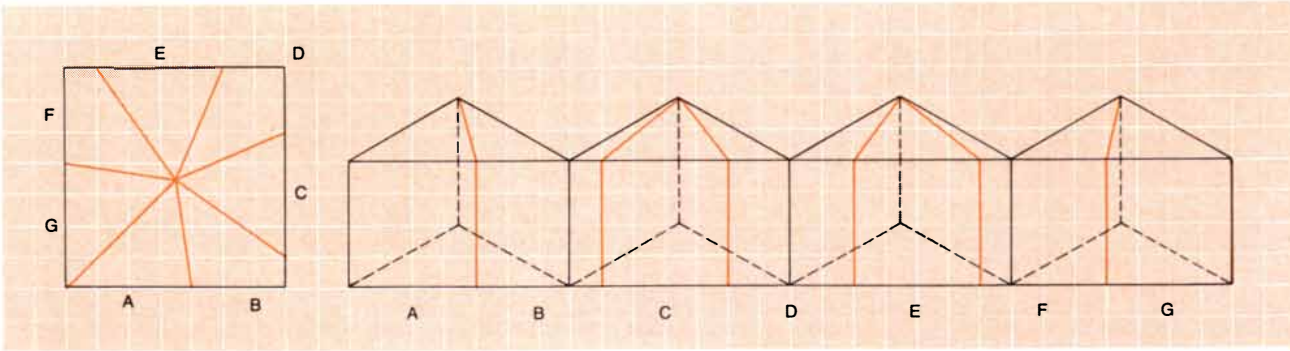
Is 8 the lowest order a perfect magic cube can have? Are there magic cubes of orders not in the 8-power series? Both are open questions.

Here are the solutions for the random assortment of puzzles presented in this department last month.

1. The first symbol of Tom Ransom's I.Q. test differs from the other four in having a colored border. The second symbol differs in not being shaded. The third differs in not being square and the fourth in not having a colored spot. The fifth is the only symbol that is not unique with respect to some property, and therefore it is the one symbol "most different" from the others. To put it differently, the first four are each different in the same general way, whereas the fifth is different in a different way, which makes it radically different.

One is reminded of the story about the report of a psychiatrist on the most re-

Solution to N. J. A. Sloane's series of figures



Solution to the square-cake problem

markable case in his experience. The patient was a person who showed no trace of any kind of neurosis. In my opinion Ransom has found a simple, elegant model of a semantic paradox that is frequently encountered but seldom explicitly recognized.

The sixth figure in N. J. A. Sloane's puzzle is given in the upper illustration on the opposite page, with the vertical lines of symmetry shown in color. On the right of each colored line are the numerals 1 through 6; on the left are their mirror reflections.

2. To cut a square cake frosted on the top and four sides into n pieces of equal volume and equal frosting area we need only mark the perimeter into n equal parts and cut the cake in the usual manner [*see lower illustration on opposite page*]. To understand why this is so we adopt a stratagem given by Norman N. Nelson and Forest N. Fisch in their article "The Classical Cake Problem," *The Mathematics Teacher,* Vol. 66, November, 1973, pages 659–661. Imagine the cake cut along its diagonals into four congruent triangular prisms, and the prisms arranged in a row as shown at the right in the illustration. The base line is the cake's perimeter. Divide the cake into seven equal portions, *A, B, C, D, E, F, G,* as shown by the seven colored lines. Those lines correspond to the seven cutting planes in the solution.

It is easy to see that the area of icing on each portion is the same. The sums of the one or two rectangles that form the sides of each portion clearly are equal. The tops of each portion (composed of one or two triangles) are also equal in area because the sums of the bases of these triangles are equal, and all triangles have the same altitude. Finally, the volumes of the seven portions are the same because their heights are equal and their tops are equal in area.

The solution obviously generalizes to n portions. We cannot, however, apply the procedure to a cake on which the frosting has a finite thickness because the four corners of such a cake introduce complexities.

3. The French cryptarithm is solved with $94{,}851 + 6{,}483 + 6{,}483 = 107{,}817$, and the German one with $821 + 821 + 821 + 821 = 3{,}284$.

4. The first published attempt to solve Lewis Carroll's anagram poem was made by Sidney H. Williams and Falconer Madan in their *Handbook of the Literature of the Rev. C. L. Dodgson.* Some of their errors were corrected in 1962 in a revised edition, the *Lewis Carroll Handbook,* by Roger L. Green. For further criticism and speculation see Philip S. Benham, "Sonnet Illuminate," in *Jabberwocky,* the journal of the Lewis Carroll Society, Summer, 1974. (The society's address is 55 Heath Cottages, Chislehurst, Kent, England.)

Solution to the third-man theme

Drawing on the above sources, and with help from Spencer D. Brown, here is how matters stand:

1. As to: *oats* (not *oast, stoa*). 2. The war: *wreath* (not *thawer*). 3. Try elm: *myrtle*. 4. I tried: *tidier*. 5. The wig: *weight*. 6. Cast in: *antics* (not *sciant*). 7. I went: *twine*. 8. To ride: *rioted*. 9. Ring yes: *syringe*. 10. We rang: *gnawer*. 11. Let's rap: *plaster* (not *stapler, persalt*). 12. We don't: *wonted*. 13. O shew: *whose*. 14. Her wit: *writhe* (not *wither, whiter*). 15. As yet: *yeast*. 16. She won't: *snoweth*. 17. Saw eel: *weasel*. 18. In Rome: *merino* (not *moiren, minore*). 19. Dry one: *yonder*. 20. He's wet: *seweth* (not *thewes, hewest*). 21. I am dry: *myriad*. 22. O forge: *forego*. 23. Th' rogue: *tougher* (not *rougeth*). 24. Why a net: *yawneth*.

Benham notes that words 10, 16, 20 and 24 are the most questionable. Carroll's letter containing the puzzle poem first appeared in *Six Letters by Lewis Carroll,* privately printed in 1924. It is reprinted as Letter XLV in *A Selection from the Letters of Lewis Carroll to His Child-Friends,* edited by Evelyn M. Hatch (Folcroft, 1973).

5. The only solution to the third-man chess problem is to place a white bishop on the board as shown in the illustration above. Black obviously cannot move, and White cannot move because it is not his move. To prove it must be Black's move (therefore Black is stalemated) requires such elementary retrograde analysis that I leave it to the reader. The problem appeared in *The Problemist* (September, 1974, page 471, Problem R16).

THE AMATEUR SCIENTIST

How an amateur can construct a model of an enzyme molecule at modest cost

Conducted by C. L. Stong

Until recently only a full-time scientific worker would have been likely to undertake to build a model of an enzyme, partly because these giant molecules that catalyze the chemical reactions of the living cell are highly complex and partly because the task is quite expensive with most model-building kits. Frank H. Clarke (14 Long Pond Road, Armonk, N.Y. 10504) decided to see if the project could be brought within the reach of amateurs in terms of both skill and cost. He found that it could. Clarke, a medicinal research chemist, is deputy director of the chemistry division of Ciba-Geigy Corporation. He recently received a patent for the novel features of the model-building components he describes as follows.

"There are two major kinds of model for illustrating molecular architecture. The skeletal model indicates the centers of the atoms and the bonds that join them; it resembles the kind of line drawing an organic chemist would make to define a molecular structure. The space-filling model shows both the shape of the molecule and the volume it occupies. It is particularly useful for suggesting how one molecule interacts with another.

"Up to now it has not been very practical even for the professional scientist to build a model that is both skeletal and space-filling. In the method I have devised it is possible to make such a model, and with inexpensive components. The finished model not only is instructive as a guide for understanding the details of modern biochemistry but also is a handsome object to have in the home.

"The method can be applied to any biological molecule for which the atomic architecture is fully known. The enzyme lysozyme is a good example. A model I have made portrays the catalytically active site of alpha chymotrypsin, a proteolytic (protein-splitting) enzyme [*see illustration on opposite page*]. The first step in learning to make such a model is to consider the structure of peptides and the bonds that give rise to it.

"A peptide bond is constructed by the linking of two amino acid molecules through the carboxyl (COOH) group of one and the amino (NH_2) group of another. A polypeptide is a long-chain polymer made in this way from the 20 or so natural amino acids. The chain is arranged so that the carboxyl group of the first amino acid forms part of an amide (peptide) bond with the amino group of the second and so on. The final amino acid in the chain has a free carboxyl group. It is convenient to build a model in the same way that nature builds a protein. Each chain begins with a free amino group of the first amino acid and ends with the free carboxyl group of the last amino acid.

"As the polypeptide chain grows, the successive amino acids become fixed in a particular configuration that can be described in two ways. One is to give the spatial coordinates of each atom. The other, which is possible because the distances between successive atoms are known, is to work in terms of the dihedral angles around the bonds that join the atoms in the chain.

"For the molecular hobbyist it is convenient to use components cut to predetermined lengths and to fix the angles around the successive bonds. Nearly all peptide bonds are trans and planar, meaning that hydrogen and carbonyl oxygen (CO) are on opposite sides of the bond. Hence it is necessary only to fix the two dihedral angles around each alpha carbon atom and then to link each amino acid successively with its predecessor in the chain to form trans and planar amide bonds.

"As the chain is extended small errors are often multiplied. It is therefore necessary to have additional means of determining the shape of the chain. Skeletal drawings based on the horizontal coordinates of the chain segments are helpful. Moreover, since all proteins have a secondary structure determined partly by hydrogen bonds, these bonds can also help the model builder to work out the conformation of the molecule. When the main chain has been made, it is mounted on a base and the side chains are added. Finally, space fillers are attached to round out the molecule.

"I make models with components I obtain commercially and then modify. Two of the basic components are made by Prentice-Hall, Inc. (Englewood Cliffs, N.J. 07632), namely Framework Molecular Model tubes and straight aluminum connecting pins. Two others are available from Science Related Materials, Inc. (P.O. Box 1422, Janesville, Wis. 53545), namely color-coded Minit atom centers and expanded-polystyrene balls.

"The modified components I make are unique in two ways. First, they are color-coded, so that not only the plastic atom centers but also the connecting tubes are colored to identify the atoms that are joined. Second, the tubes have tapered cavities so that they can be pressed onto the arms of the atom centers to fix the dihedral angles. Each atom and bond of the completed model is readily identified by its color: blue for nitrogen, red for oxygen, black for carbon, yellow for sulfur and white for hydrogen. The polystyrene space fillers are also color-coded to represent the individual atoms.

"I work at a scale of 12.5 millimeters per angstrom. It is the same as the scale of the well-known Corey-Pauling-Koltun (CPK) space-filling models. The two sets of models can therefore be used interchangeably.

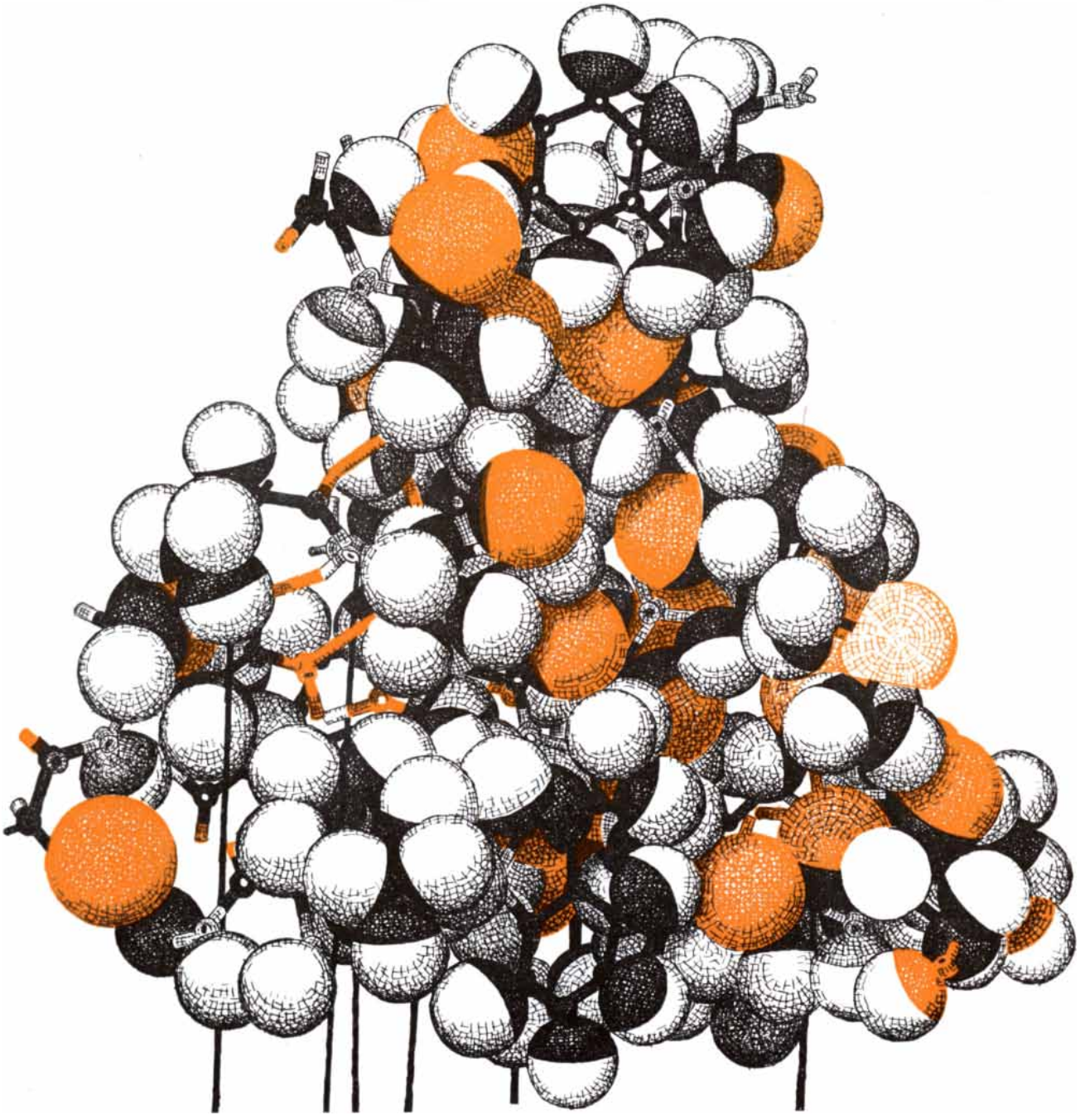
"For the model builder the most important consideration is the overall form of the molecule, particularly when the molecule is a complex bio-organic compound. It is not necessary to locate the center of each hydrogen atom. Moreover, the distance between carbon-hydrogen bonds can be made longer than the scale distance between the atom centers. The distance between carbon-carbon bonds, however, must be exact because the length and shape of the chain are important.

"Even so, chains are frequently not stretched out; indeed, polypeptide

chains are normally bent back on themselves. The angles of the bends are important, but often the individual distances between atoms need not be exact to build a good representation of a complete peptide. Thus the connecting tubes of the color-coded skeletal models have only three different lengths (12.5, 14.5 and 21.7 millimeters). Those lengths suffice for a wide variety of models of bio-organic compounds. When the tubes are connected to the Minit atom centers, they provide interatomic distances representing 1.36 angstroms (for short bonds, such as those in aromatic rings), 1.54 angstroms (for bonds of medium length, such as those in aliphatic carbon compounds) and 1.81 angstroms (for longer bonds, such as those involving sulfur atoms). The color combinations of the various lengths are indicated in the accompanying illustration [*upper illustration on next page*]. Yellow Testors Pla enamel paint is used to color black and white tubes for S—C and S—H bonds.

"The color-coded tubes with tapered cavities are not yet available commercially. It is easy to make them, however, by cutting to appropriate lengths the Framework Molecular Model tubes. Hold a tube in a suitable hand vise and drill the tapered cavities with a hand power drill fitted with an eighth-inch bit that has been ground to a taper with an emery stone. Alternatively a tapered cavity can be made by pushing the tube onto the hot No. 5 conical tip of a Weller soldering iron.

"The color-coded Minit atom centers are most convenient for use with these tubes. The atom centers are inexpensive and are supplied in a variety of colors with tetrahedral and planar-trigonal arrangements of the valence arms. Special trigonal carbon and nitrogen centers are



Frank H. Clarke's model of the active site of alpha chymotrypsin

supplied for building the amide bonds of peptides.

"In addition the short, straight aluminum connecting rods supplied by Prentice-Hall are convenient for forming the many different angles involved in the hydrogen bonds of polypeptides. They are easily bent to the proper angle with two short lengths of brass tubing, such as those that hold the ink in certain ballpoint pens. A sturdy base and stiff wire rods are needed to support a complex model. The rods can be made from wire coat hangers.

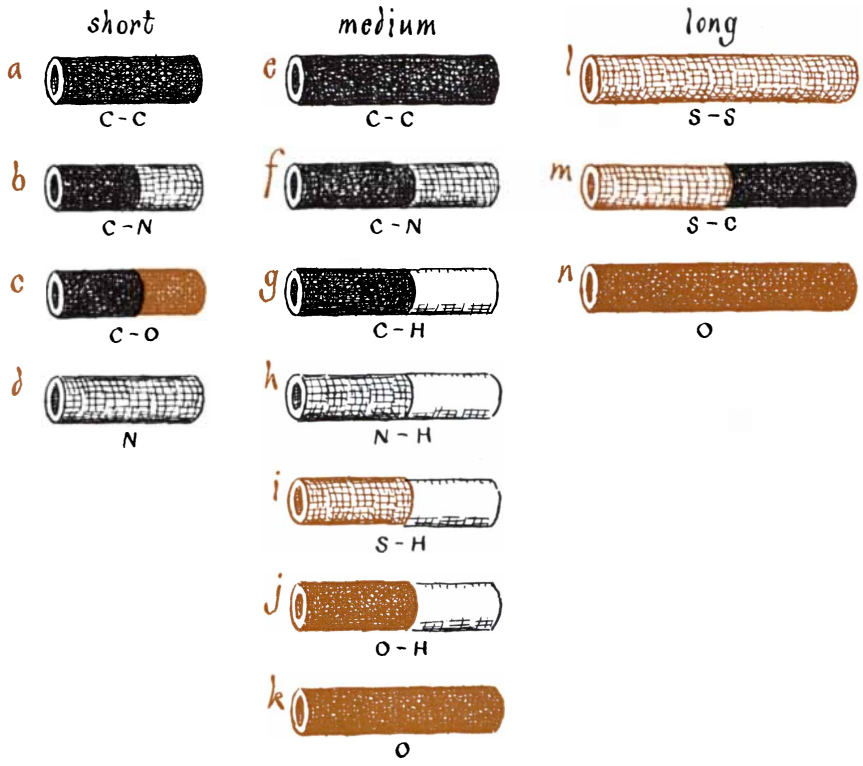
"The space-filling polystyrene balls come in two diameters, one inch (25.4 millimeters) for hydrogen atoms and 1.37 inches (34.8 millimeters) for oxygen and sulfur atoms. The hydrogen atom is exactly to scale. The oxygen atom is somewhat larger than the scale calls for and the sulfur atom is a little smaller, but the discrepancies are not large.

"The bottom third of the hydrogen atom is painted the color of the atom to which it is joined: black for carbon, blue for nitrogen and red for oxygen. Carbonyl-oxygen atoms are painted red with a black base. The sulfur atoms are yellow, with black for the carbon portion of a carbon-sulfur bond. Since the sulfur atoms are divalent, the spheres must be shaped. One must also make holes in the base of each hydrogen and carbonyl-oxygen atom so that it will fit neatly over the skeletal tube.

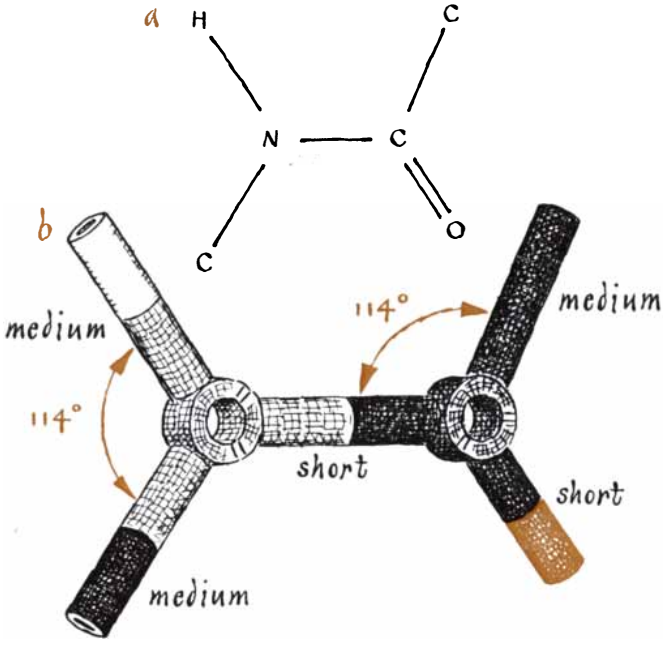
"In sum, the components of a model-building kit that is color-coded and both skeletal and space-filling are simple and either available or easily prepared. They are so inexpensive and the assembled model is so attractive that it is preferable to make a new model from fresh components rather than to disassemble an existing model. Entire collections of molecules can be easily assembled and employed in comparative studies of molecular architecture.

"The points that must be taken into account in assembling a model of a polypeptide are the construction of the amide bonds, the arrangement of atoms around the asymmetric alpha carbon atom, the measurement of dihedral angles and the construction of the hydrogen bonds. As I have mentioned, each amino acid of the peptide chain is joined to its neighbor by an amide bond. The N–H is trans and planar to the carbonyl oxygen. The amide bond is formed between two Minit atom centers with two double bars [*see bottom illustration at left*]. The angle opposite the pair of double bars is 114 degrees. Short connecting tubes serve for the N–C bond and for the carbonyl group. The other tubes are of medium length.

"To make a peptide chain one first prepares the amino acid units and then joins them in sequence with the amide bonds. An amino acid unit is depicted in the accompanying illustration [*bottom of opposite page*]. It has three components: the N–H, or amino, component [*"c" in illustration*], the alpha carbon atom [*"d"*] and the carbonyl component [*"e"*]. The alpha carbon atom is shown attached to its alpha hydrogen atom. It is advisable to put on the tube for the hydrogen atom a label with a number to indicate the particular amino acid that is to be represented. Avery self-adhesive labels are convenient for this purpose, provided that you cover each one with a



Color-coded scheme for connecting tubes



Amide bond to join amino acids in the peptide chain

small piece of transparent tape to keep it from unwrapping. The number is particularly important because in model building the distinguishing side chain *R* is left off initially, so that only the vacant arm indicates the stereochemistry.

"To assemble the amino and carbonyl components insert the appropriate tubes in vertical holes you have drilled in a board so that one end of each tube projects above the board. Then coat the tip of the appropriate arm of a Minit center with '5 Minute' epoxy cement (freshly mixed) and insert it in a tube. Repeat the process until all the components have been assembled.

"Finally, two of the arms of the alpha carbon are coated with epoxy cement and inserted in the black tubes of the amino component and the carbonyl component respectively. Take care to maintain the alpha carbon in the orientation shown. The completed unit [*"b" in illustration*] has the N–H on the left and the CO on the right; the hydrogen of the alpha carbon points down and the vacant arm points up. The unit now has the correct absolute stereochemistry.

"The next step is to determine and fix the dihedral angles around the alpha carbon atom. The angles are called phi and psi [*see top illustration on next page*]. Phi is the angle between successive carbonyl carbon atoms around the nitrogen–alpha carbon (N–CA) bond. Psi is the angle between successive nitrogen atoms around the CO–CA bond.

"To measure phi one sights along the N–CA bond with the carbonyl carbon pointed to the right. Phi is the angle the vacant arm on the nitrogen (blue) atom center makes with the CO–CA bond. Since the vacant arm is so short, more accuracy can be obtained by putting a thin drinking straw over the arm to lengthen it. The psi angle is measured by sighting along the CO–CA bond with the nitrogen having the vacant arm pointing to the right. Psi is the angle the CO–N bond makes with the CA–N bond.

"Dihedral angles of proteins are recorded in various ways. It is important to check which angle is measured around the N–CA bond and which one is measured around the CO–CA bond, whether an angle of zero degrees is cis or trans and whether a positive angle is measured clockwise or counterclockwise.

"When you have measured the dihedral angles, press the tubes firmly onto the arms to fix the atoms in position. The angles are easily adjusted while the cement is soft, but after it has set each bond soon becomes permanent.

"Approximately 10 amino acid units are first joined to form segments of a polypeptide chain, and later the segments are joined. Make the amide bond by coating the vacant arm of amino acid unit No. 2 with epoxy cement and inserting the arm into the nitrogen-bond cavity of the C–N tube attached to the CO carbon center of amino acid No. 1. Do it so that the amide bond is trans and planar. The amine of unit 3 is then attached to the carbonyl of unit 2 and so on until the segment is complete.

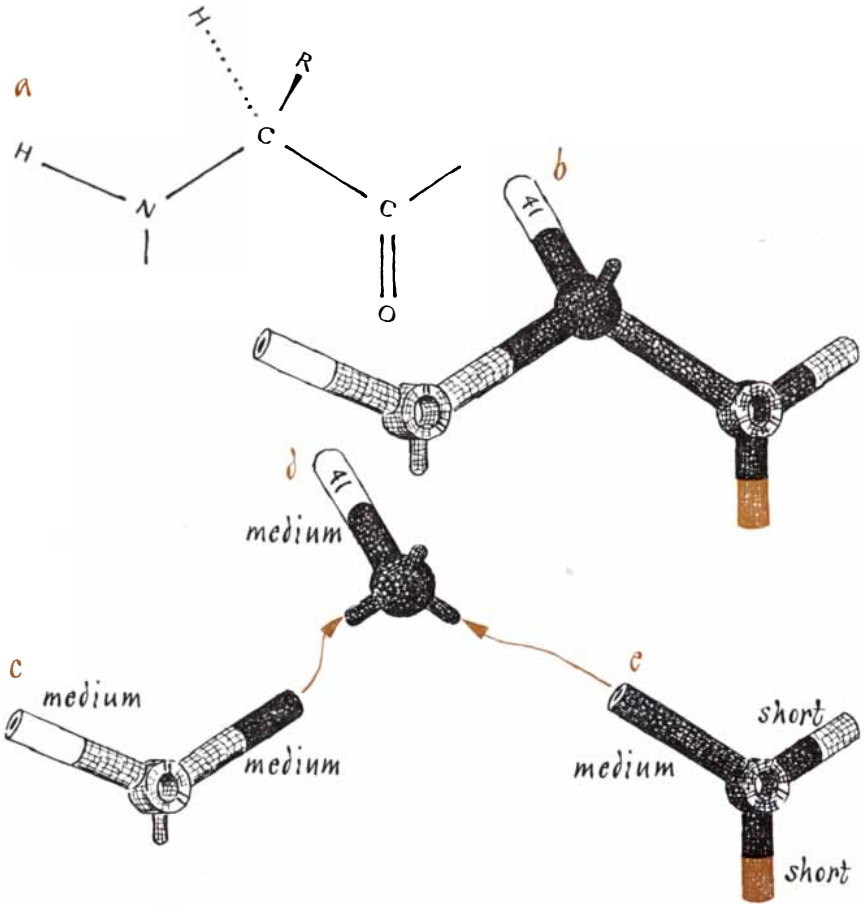
"Often it develops that when the chain folds back on itself, hydrogen bonds are formed between amido hydrogens and carbonyl oxygens of the same chain. Hydrogen bonds are also formed between adjacent chains, and they provide the secondary structure that keeps the polypeptide in a particular shape. For the model builder hydrogen bonds provide structural strength, so that fewer outside supports have to be supplied.

"In a polypeptide molecule many of the hydrogen bonds are between an amido hydrogen and the carbonyl of a neighboring amide. Often the line joining the carbonyl, the hydrogen atom and the nitrogen atom follows a zigzag course. When the hydrogen bond is linear, the distance between the center

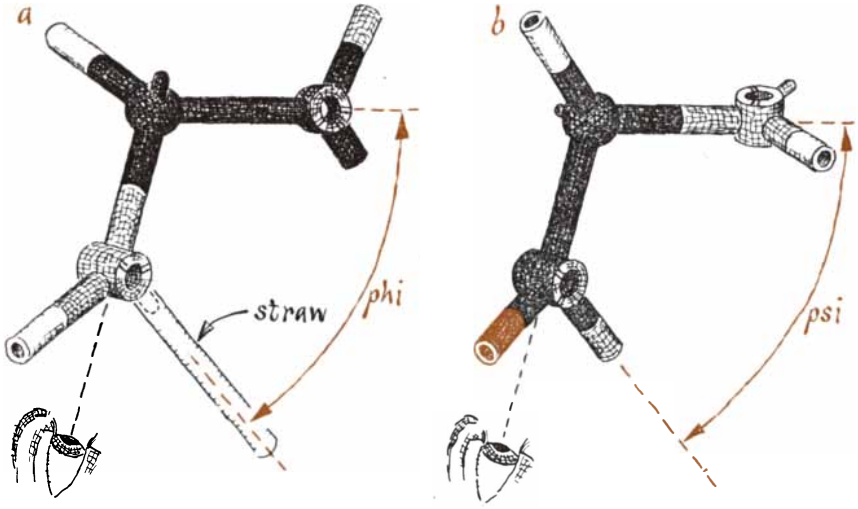
A space-filling hydrogen atom

of the nitrogen atom and the center of the oxygen atom is three angstroms. On the scale of 12.5 millimeters per angstrom the distance is represented in the model by a distance of 37.5 millimeters [*see middle illustration on next page*].

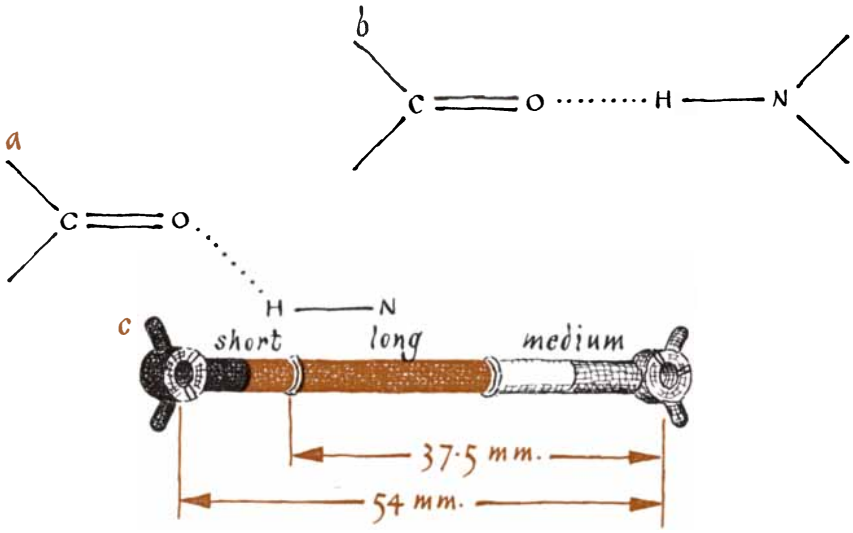
"To make such an arrangement the van der Waals radius of oxygen is represented by a long tube, the CO double bond by a short tube and the NH by a tube of medium length. The long tube is joined to the CO and the NH by Framework Molecular Model linear aluminum



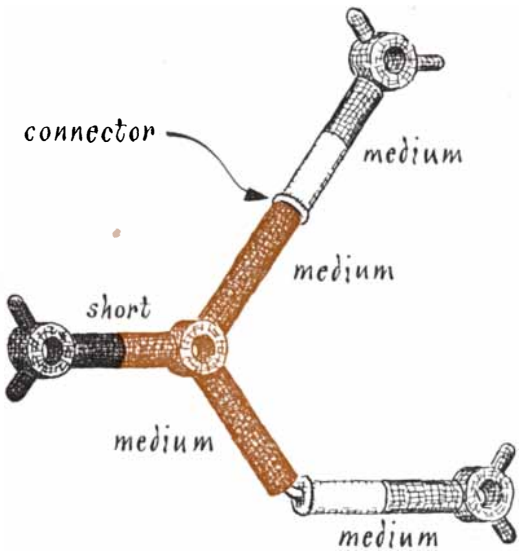
Forming an amino acid unit



Determining the dihedral angles around the alpha carbon atom



Scaling of a linear hydrogen bond



Planar, trigonal arrangement for an oxygen atom involved in two hydrogen bonds

pins. In the assembled linear hydrogen bond the distance between the center of the carbon of the carbonyl and the center of the nitrogen of the amide group is 54 millimeters. It is a relatively simple matter to extend this construction technique to form hydrogen bonds that are not linear. For this purpose you bend the aluminum connectors to the proper angles before joining them to the plastic tubes.

"Occasionally a single carbonyl oxygen is involved in two hydrogen bonds. Make them with a planar trigonal oxygen [*see bottom illustration at left*]. In this instance the van der Waals radius of oxygen is represented by a tube of medium length. The aluminum connectors can be either straight or bent as depicted in the upper and lower parts of the illustration respectively. The same principles are easily adapted to form hydrogen bonds between hydroxyl groups or between a hydroxyl group and the nitrogen atom of an imidazole group of histidine. For the latter a short blue tube represents the van der Waals radius of nitrogen.

"The arms of the aluminum connectors are smaller in diameter than the arms of the Minit atom centers. The pins fit the Framework Molecular Model tubes without a tapered cavity. Accordingly the white ends of the tubes representing OH and NH are not drilled, and neither are the ends of the long red tubes. One end of each medium-length red tube is drilled to fit the arm of a Minit center. Carbonyl tubes for amide bonds are drilled only at the black end, but CO tubes for hydroxyl groups are drilled at both ends.

"When the main chain, with its hydrogen bonds, is assembled, the side chains are constructed and added. For this task it is most convenient to prepare plans of the horizontal coordinates of the nonhydrogen atoms of small segments of the main chain. The plans become quite crowded if the segments are longer than about 10 amino acid units.

"After the skeletal model is complete it can be partly or completely converted to a space-filling model by adding the painted polystyrene balls. The illustration [*top of preceding page*] shows the space-filling hydrogen atom with its painted bottom third.

"The use of models of enzymes enables the medicinal chemist to study the interaction of small molecules with complex ones, to design enzyme inhibitors and to gain understanding of the intricate mechanisms of complex bio-organic processes. Eventually studies such as these will aid directly in the discovery of new medicinal agents."

"The most complete and most scholarly dictionary of the English language"

—The Christian Science Monitor

As an introduction to membership in the

BOOK-OF-THE-MONTH CLUB®

yours for only

$17.50

[Publisher's list price $90]

THE SUGGESTED TRIAL: You simply agree to buy four Club choices within a year at substantial savings on most books you choose.

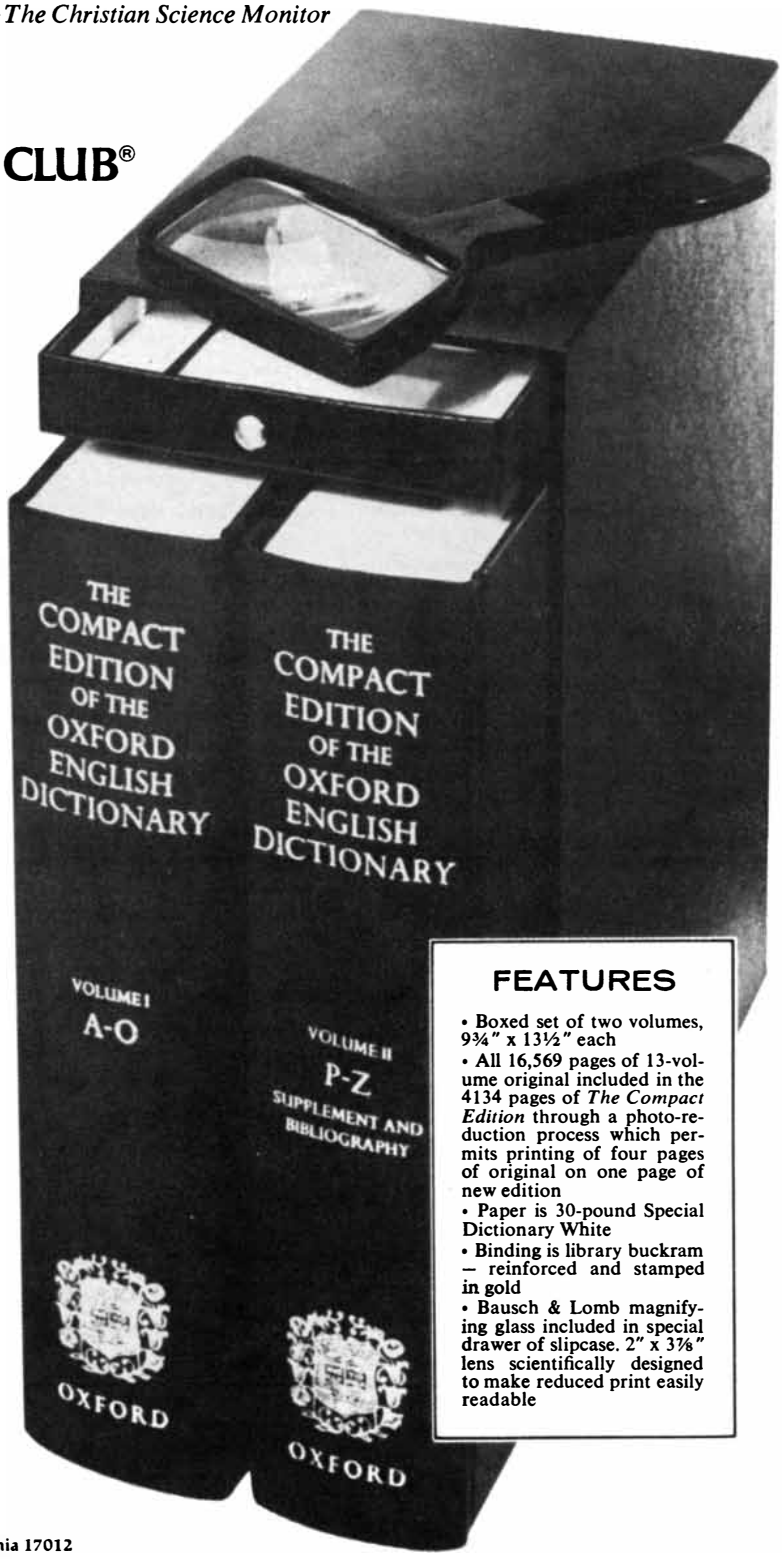
THE OXFORD ENGLISH DICTIONARY is generally regarded as the final arbiter of the English language. Until recently it had been available only as a thirteen-volume set, currently priced at $350. Now, through the combination of an ingenious method of micrographic reproduction and a fine Bausch & Lomb optical lens, every single one of its 16,569 pages, fifty million words and close to two million illustrative quotations appears, in easily readable form, in *The Compact Edition.*

> The *New York Times* book critic Christopher Lehmann-Haupt has said of this edition: "It is something of a miracle. . . . *The Compact Edition* is easier to work with than the original with its 13 separate volumes."

Even more extraordinary, as a trial member of the Club you may obtain the two-volume set for only $17.50. And as long as you remain a member, you will receive the Book-of-the-Month Club *News,* a literary magazine announcing the coming Selection and describing other important books, known as Alternates, most of which are available at substantial discounts – up to **40%** on more expensive volumes.

If you continue after your trial membership, you will earn at least one Book-Dividend® Credit for every Selection or Alternate you buy. These Credits enable you to acquire for your library a wide variety of books, called Book-Dividends, at astonishing savings—*at least 70% of publishers' list prices.*

All Book-of-the-Month Club books, including Book-Dividends, are identical to the publishers' editions in content, format, size and quality.



FEATURES

- Boxed set of two volumes, 9¾" x 13½" each
- All 16,569 pages of 13-volume original included in the 4134 pages of *The Compact Edition* through a photo-reduction process which permits printing of four pages of original on one page of new edition
- Paper is 30-pound Special Dictionary White
- Binding is library buckram — reinforced and stamped in gold
- Bausch & Lomb magnifying glass included in special drawer of slipcase. 2" x 3⅛" lens scientifically designed to make reduced print easily readable

BOOK-OF-THE-MONTH CLUB, INC., Camp Hill, Pennsylvania 17012

BOOKS

Precision production instruments, Gypsies and natural catastrophes

by Philip Morrison

DIVIDING, RULING AND MASK-MAKING, by D. F. Horne. Crane, Russak & Company, Inc. ($57.50). "He had great faith in instruments, and I do not think it naturally occurred to him to doubt the accuracy of a scale.... He was astonished when we found that one of his micrometers differed from the other." So wrote Francis Darwin of his father, Charles Darwin. This volume is devoted to the justification of our faith in the elegant and exacting industry that forms and calibrates the linear and circular scales that fix the settings of machine tools, theodolites and levels. These in their turn determine the fit, size and shape of most of our engineering products, large and small. Today's best production practice routinely reaches an accuracy of a quarter of a micrometer (formerly a micron). This volume surveys the techniques by embedding within a clear and expert general account rather detailed excerpts from actual manufacturer's directions for the use of a wide variety of special instruments and procedures. It is well illustrated with many diagrams and photographs of real setups.

The classical products are mainly sets of lines on stainless steel or on glass, either ruled directly with a sharp diamond edge or etched through a "resist" that has been made by ruling or by photochemistry. Evaporated metal can also be deposited through the resist. The widest contemporary extension of these arts is the one covered in the third term in the title: the remarkable technology of making integrated circuits. There calibration is of little importance, but high precision is nonetheless needed for the proper and precise registry of half a dozen layers (resistive, conductive and carrier-diffusing) that are placed on a silicon semiconductor base to yield the complex arrays of minute circuit elements—smaller and denser each year—that are the works of pocket calculators and guided missiles. Such patterns come to resemble molecular biology more than they do the uniform crystalline divisions of setting circles and micrometers (the instrument, not the unit).

Calibration occupies one chapter, fascinating to the general reader. Slip (or block) gauges are fundamental in the precision machine shop. They must be finished well enough so that they can be "wrung" together, or to special flat steel surfaces, with allowance made for wringing thickness, which amounts to a quarter of a microinch. One clever instrument uses four different wavelengths of the mercury-vapor lamp to give a kind of optical vernier reading of gauge thickness, without counting many interference fringes. Air pressure, temperature and humidity must be corrected for, and the gauges can be checked to a microinch. A standard meter bar (made of a steel alloy less noble than the platinum-iridium prototype in Paris) is still the basis for calibration of steel tapes and scales. The fiducial lines are engraved on the crossbar of the H-section standard, cleverly arranged to be in the neutral axis of the beam, where length does not change with small bending. The firm of V.E.B. (formerly Carl Zeiss) Jena has found a neat extension of that idea. Transparent glass scales have advantages (they give strong signals when they are scanned photoelectrically for fast calibration at many scale points, and the polished surfaces are consistently good), but it is not practical to form a glass H section in one piece. The optical wits at Jena have found a cross section, a carefully computed but simple trapezoidal shape, such that rulings made on the lower glass surface but viewed from above through the thickness of glass are seen in virtual image at the unbending mechanical neutral axis inside the solid glass.

The future seems to lie closer to laser interferometry. Electronic counting of fringes (with heterodyne techniques) and stable wavelengths give such instruments remarkable powers. One is described here: a Hewlett-Packard device based on a helium-neon laser, which can accommodate itself to calibrating 200-foot lengths with one-microinch resolution. Beam alignment might be a limit in use, but the narrow beam makes such errors minimal in the hands of an alert operator. There are even correcting thumbwheels for feeding in ambient air conditions in order to get standardized results on the digital display; they affect only the fourth digit and beyond. With a cube-corner reflector and a beam-bender one can fully survey a near-flat area. Here is displayed one such map, with half a dozen profiles running across a yard-long granite surface plate, accuracy in the microinch range. This is virtuoso work in the tool shop.

Here is the contemporary standard of circle division too. No official bar need be consulted for it; every circle has its own million-plus seconds of arc. The trick is equal division. The Moore Special Tool Company of Bridgeport, Conn., will supply you with a small-angle divider to calibrate circles to 1/10 second of arc at any position. (A test certificate shown here gives corrections to the nominal at the level of a few hundredths of a second of arc.) The heart of this classically inspired instrument (developed about 1960) is a pair of identical steel "face gears," each of which is about eight inches in diameter and has 1,440 identical *V*-shaped teeth. The two gears mesh well, of course. Once the teeth are engaged they are mutually locked. Then they are disengaged, the gears are rotated a finite number of teeth and are once again brought in contact. Thus a definite angle is stepped off, a quarter of a degree per tooth. High precision arises from the averaging effect of all those teeth in mesh at once; each gear has been mutually lapped against its mate by such engaging and disengaging motions, with random rotations, through long hours of lapping with increasingly fine abrasives. Finally the gear teeth are mutually run in for still more hours of random repetition while they are clean and dry. Interpolation between quarter-degree steps is made by a fine-pitch micrometer screw with a dial and a vernier, its flat end carefully held in contact with the main index. It need work linearly over only a

fortieth of an inch, and it is adjusted to agree with the result of a single-tooth increment.

Auxiliary arts are treated too: the making of the wide variety of optical graticules, with their lettering and scales, that we so often look through or see projected; the preparation of diamond cutting edges; the cleaning of glass; the evaporation of metal layers, and more. The arcana of diffraction gratings are laid out in a long chapter. Some gratings are ruled directly by great interferometer-controlled ruling engines; some are made from a fine master screw thread by casting cylindrical plastic replicas and splitting them, a surprisingly accurate technique developed at the National Physical Laboratory in England. The design of clean rooms is sketched, with examples. The usual ceiling-to-floor clean airflow was not good enough for the Moore Special Tool Company because the temperature of the air in such a flow pattern tends to stratify horizontally: the room is dust-free but is thermally unsuited for precision work in steel. The company added air intakes at a number of levels and temperature-control of the floor.

Horne is no stranger to this industry. He is the developer of a new device—now called the Nikon I C (integrated circuit) Printer—that produces microimages optically, without making contact with the base semiconductor slice. (The device is aligned automatically to a few micrometers and its sharp images are projected across a gap of up to tens of micrometers, preventing contamination and damage to the mask or the slice itself.) A wide-ranging chapter on metrological history, from Pharaoh's forearm to Maudslay and Johansson, has been supplied by a colleague of Horne's at Hilger and Watts, Ltd, Mrs. C. E. Arregger.

The volume was first to be called simply *Dividing and Ruling*. That sounded too political, and so the third noun, *Mask-making*, was added. It does not help: modern states behave in ways that pun on the matter-of-fact activities of that remarkable tradition of manufacture, a fusion of the graphic arts with the precision of the optical laboratory.

GYPSIES: THE HIDDEN AMERICANS, by Anne Sutherland. The Free Press ($13.95). GYPSY ON 18 WHEELS: A TRUCKER'S TALE, by Robert Krueger. Edited by Sam Yanes. Praeger Publishers ($6.95). Real Gypsies are the subject of the Sutherland book, "an ethnographic description of a group of American Rom" as they lived about 1970 in a "depressed town of substandard housing" in northern California, "its most noticeable feature being the maze of railway tracks that form natural boundaries for the various ethnic groups."

The Gypsies call themselves O Rom. They belong to the quasi-nomadic folk, now spread all over the world, who speak a Sanskrit-derived language they call Romanes. It is 1,000 years, more or less, since their ancestors left the north of India and began their wanderings. In all that time, over four continents, through profound economic and political changes in the host lands across which they have nearly always moved as a tiny, outcast minority, they have remained unified, distinct, recognizably as they were described in medieval Europe. The question that motivates this acute and technical field study by a young Oxford-trained American social anthropologist on the faculty of the University of Durham is: How have they done it?

The author is pretty convincing, even though her answer is trendy. The Rom live in a world of social experience cleft into purity (*wuzho*) and pollution (*marime*). All that is non-Rom (*gaje*) is morally wrong. The *gaje* are polluted. They do not even know to keep the upper and the lower body strictly separated, as the Rom do. Let the living room be strewn with discarded food to be swept out into the yard every day. No matter. The house stays pure if the six classes of laundry (male and female, tops and bottoms, food and children) are kept separate, and if food is never handled by *gaje* or by the temporarily unclean. To eat, to sleep or to work among the careless outsiders is manifestly to risk the danger of *marime* and of eventual ejection from the Rom. The Rom always try to cooperate; they work together and not side by side with *gaje;* they must be free to travel widely to fulfill their social duties centered on marriage rites and death rites. The Rom families that are studied here, about 300 people in 33 households, spent almost half of their time away from home, traveling.

The economic life of the Rom, constant through superficial change, remains based on the extraction of support from the *gaje*—shameless people who can command no respect yet are the only possible source of a livelihood based on mobility and separation. Horse trading gives way to coppersmithing; the tinker's art becomes fender-fixing; there is always itinerant farm labor—and now there is welfare, the chief income source for this sample. For 1,000 years the Rom have been persuading *gaje* to give them money on one slender pretext or another: tea-reading or palmistry, doctoring lame horses or fiddling in the street. They have become fully adapted to the way of the hustle. Welfare is only the latest hustle, further proof of the deep gullibility of the *gaje*. Read the I.Q. estimates made of certain Rom who played retarded to avoid having to accept employment. These clever people elicited results such as: "An I.Q. score of 64... a mental age of about ten years. Thinking was markedly concrete... motor performance was also poor.... Retraining does not seem feasible." But the subject of that report told Sutherland: "The trick is never to protest anything but act like you are doing everything right.... I just give some wrong answer... there was a picture of this doll, and I was supposed to connect the arms and legs. Well, I put the legs in the armholes and the arms below. She kept trying to help me but I stuck with that like I was sure it must be right." That is the subtle art of pretense, which informs the petty deceits and pilfering long associated with the Gypsies. They are almost never violent and generally engage in a fully temperate parasitism that avoids deep damage to the host. It is a social adaptation.

And the Rom steadily adapt. They remain generally illiterate, although the need to move within a bureaucratized society has begun to persuade them that writing is useful. The most prolonged contact with the Rom the author reports was her time as the principal of a church-supported, Gypsy-controlled "Romany school," which was intended to allow their children to acquire literacy without losing the values of the Rom. Costumes change. The men dress colorfully, with Stetsons, gold badges and flashy colors, but fundamentally in *gaje* modes; children wear normal clothes. Only the women hold to distinctive dress: the 10-yard skirt and a low-cut blouse, echoes of the basis of *wuzho* and *marime*.

The book tells much more, particularly about the complicated kinship system and the elaborate social groupings of the Rom, from the level of "the tent" to the lineage, the community and the tribe. There are documents of Gypsy trials (transcripts of recorded narratives), a glossary of Romanes terms and a critical bibliography. (The reader is warned not to generalize much beyond this one examined community.)

It is not likely that the Rom will be assimilated, although the possibility is constant. Official recognition is the newest hazard they have encountered (they were named an official U. S. minority in 1972), and it may increase their vulnerability. Soon there will be nowhere to go, and territoriality may become their

What's new in physics... applied mathematics... the lives of bees... our galaxy...chemistry... early civilizations...

To keep up with everything that's happening in every scientific area that interests you: Take any 3 books (values to $64.95) all for only $3.95— as a member of

The Library of Science.

if you will join now for the trial period and agree to accept only 3 more books—at members' discount prices—over the next 12 months.

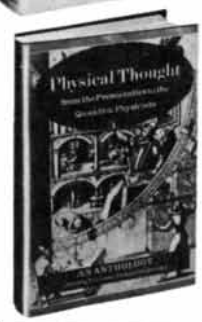
(Publishers' Prices shown)

70170. PRINCIPLES OF HOLOGRAPHY. Second Edition. *Howard M. Smith.* Completely revised and expanded edition of classic work covers every aspect of holography from fundamentals to rigorous theory, from chemical formulas to applications. **$16.50**

70190. PRINCIPLES OF OPERATIONS RESEARCH. Second Edition. *Harvey M. Wagner.* Based on the author's twenty years of experience applying operations research to a wide variety of problems, this 1039-page book offers a complete overview of all the fundamental concepts. **$17.95**

52190. HANDBOOK OF APPLIED MATHEMATICS. *Edited by Carl Pearson.* This joint effort by over 20 contributing author-specialists provides over 1,000 pages of vital and practical techniques for scientists and engineers who use mathematics as a tool. *Counts as 2 of your 3 books.* **$37.50**

87855. THE WORLD OF MEASUREMENTS. *H. Arthur Klein.* Abounding in enlightening insights and illuminating observations, this delightfully wide-ranging overview of meteorology gives a fascinating history of the development of weights and measures. **$14.95**



68750. PHYSICAL THOUGHT FROM THE PRE-SOCRATICS TO THE QUANTUM PHYSICISTS. *Samuel Sambursky.* Skillfully edited by one of England's preeminent science writers, this more than 500-page anthology is an impressive history of the most important events in the development of the physical sciences. Includes excerpts of great scientists from Aristotle to Pauli. **$20.00**

How the Club Operates • The Book Club **News**, describing the coming Main Selection and Alternate Selections, will be sent to you 15 times a year at 3-4 week intervals. • If you wish to purchase the Main Selection, do nothing and it will be shipped to you automatically. • If you prefer one of the Alternates, or no books at all, simply indicate your decision on the reply form always enclosed with the **News** and mail it so that we receive it by the date specified. • The **News** is mailed in time to allow you at least 10 days to decide if you want the coming Main Selection. If, because of late mail delivery of the **News**, you should ever receive a Main Selection without having had the 10 day consideration period, that Selection may be returned at Club expense. • Continued membership past the trial period brings you the added benefits of our Bonus Book Plan—an important way to save even more, at least 70% off publishers' prices.

The Library of Science 2-4AD
Riverside, New Jersey 08075

Please accept my application for membership and send me the three volumes indicated, billing me only $3.95. I agree to purchase at least three additional Selections or Alternates during the first year I am a member, under the club plan described in this ad. Savings range up to 30% and occasionally even more. My membership is cancelable any time after I buy these three books. A shipping and handling charge is added to all shipments. **Send no money. Members are billed when books arrive.**

3 books for $3.95
Indicate by number the 3 books you want.

A few expensive books (noted in book descriptions) count as 2 choices.

Name _____

Address _____

City _____ State _____ Zip _____

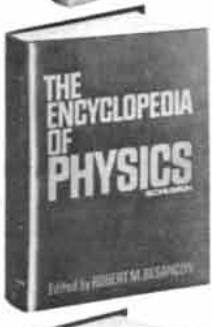
(Offer good in Continental U.S. and Canada only. Prices slightly higher in Canada.) Book selections purchased for professional purposes may be a tax deductible expense.

68735. PHYSICS OF THE EARTH AND PLANETS. *Alan H. Cook.* What forces shaped the earth? What is the mechanism that prompts continental drift? These questions—and dozens more—are answered by one of the world's leading authorities on physical meteorology. *Counts as 2 of your 3 books.* **$24.75**

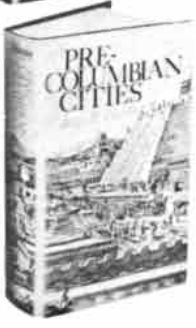
62630. THE MILKY WAY. Fourth Edition. *Bart Bok and Priscilla Bok.* A thoroughly updated and expanded edition of the most popular book on our galaxy. This outsized volume, filled with over 125 pictures, charts and diagrams, has been revised to include the latest scientific research while retaining the same elegant, precise style of its predecessors. **$15.00**

48630. THE FIRST GREAT CIVILIZATIONS. *Jacquetta Hawkes.* A distinguished English archaeologist has written an extraordinary history of the great civilizations of Mesopotamia, the Indus Valley, and Egypt. Big and lavishly illustrated. **$12.50**

52650. THE HERITAGE OF COPERNICUS. *Edited by Jerzy Neyman.* Sponsored by the National Academy of Sciences, this volume commemorates the 500th birthday of Copernicus by offering 25 highly readable essays on the quasi-Copernican revolutions that have occurred in the past five centuries. Contributions by Bart J. Bok, Stanley L. Miller, Glenn T. Seaborg, and two dozen other scientists. *Counts as 2 of your 3 books.* **$30.00**

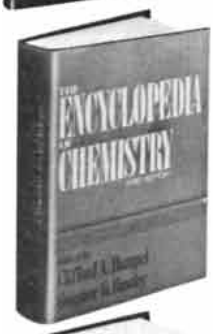


44941. THE ENCYCLOPEDIA OF PHYSICS. Second Edition. *Edited by Robert M. Besançon.* In this updated and expanded edition of a modern classic, over 300 distinguished authorities in 344 articles offer incisive, authoritative information on every major topic in physics today. Outsized volume contains more than 1000 pages. *Counts as 2 of your 3 books.* **$37.50**



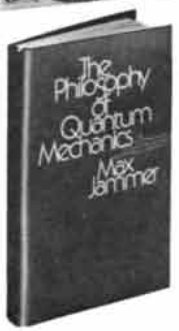
70240. PRE-COLUMBIAN CITIES. *Jorge E. Hardoy.* A highly acclaimed synthesis of the urban civilization of the New World before Columbus. "... first-class reading." —*Scientific American.* **$19.95**

54240. HUNTING FOR DINOSAURS. *Zofia Kielan-Jaworowska.* An absorbing account of a team of paleontologists' three-year effort to find and decipher the meaning of the ancient bones that litter the Gobi desert. Crammed with fascinating descriptions of dinosaurs and their evolution, the Gobi and its people, and the team's work and discoveries. **$12.00**



44861. THE ENCYCLOPEDIA OF CHEMISTRY. Third Edition. *Edited by Clifford A. Hampel and Gessner G. Hawley.* In this completely revised and greatly expanded new edition of a classic encyclopedia, 600 leading chemists thoroughly explore over 800 theoretical and applied aspects of chemistry and chemical technology. Over 1200-page, outsized volume offers hundreds of invaluable charts, diagrams, and illustrations. *Counts as 2 of your 3 books.* **$39.50**

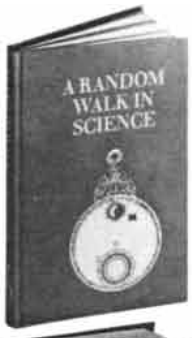
42325. THE DISCOVERY OF OUR GALAXY. *Charles A. Whitney.* "... a clear and exciting account of some of the history behind modern extra-galactic astronomy ... Charles Whitney's account is easily understandable, rapid moving, remarkably free of jargon, and requires no mathematical background."—*American Scientist.* **$10.00**



68690. THE PHILOSOPHY OF QUANTUM MECHANICS. *Max Jammer.* A distinguished scholar and physicist offers a truly impressive study of the interpretations of quantum mechanics. This 500-page volume explores virtually all of the major interpretations—from the early semi-classical interpretations, through the hidden variable theories, to the modern stochastic and statistical interpretations. **$19.95**

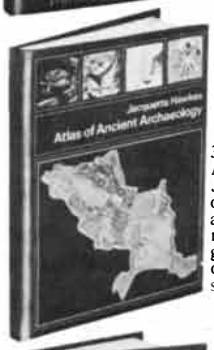
78820. THE SOCIAL BEHAVIOR OF THE BEES. *Charles D. Michener.* One of the world's foremost authorities on bees offers the first definitive work on group phenomena among these most amazing insects. A publishing classic. **$25.00**

46500. THE EVOLUTION OF RADIO ASTRONOMY. *J. S. Hey.* One of the pioneers of radio astronomy offers an absorbing history of this powerful technique's remarkable development. "... an important work in the history of astronomy."—M. L. Meeks, in *Sky and Telescope.* **$10.00**

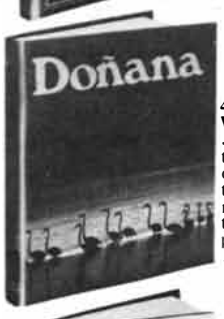


73330. A RANDOM WALK IN SCIENCE. *Compiled by Robert L. Weber and Eric Mendoza.* A collection of jokes, short essays and poems on humor and humanity in physics and related sciences. "... it is the best anthology of its kind I have seen. I can't imagine a science-minded reader who would not find it vastly entertaining."—Martin Gardner. **$12.50**

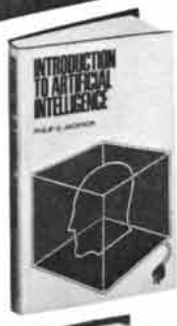
67190. THE PATH TO THE DOUBLE HELIX. *Robert Olby.* A captivating history of twentieth century experimental biochemistry and a highly readable account of the physical and chemical nature of the gene. "... The first full scholarly account of how the structure of DNA was discovered ... absorbing reading."—Francis Crick. *Counts as 2 of your 3 books.* **$23.50**



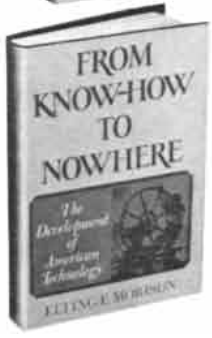
34720. ATLAS OF ANCIENT ARCHAEOLOGY. *Edited by Jacquetta Hawkes.* With the aid of over 350 maps, site plans, and drawings, two dozen internationally renowned archaeologists offer detailed descriptions of 170 major archaeological sites. **$19.50**



42680. DOÑANA: SPAIN'S WILDLIFE WILDERNESS. *Juan Antonio Fernandez.* Sumptuously illustrated portrait of one of the world's most important wildlife sanctuaries. Stunning 9-by-11 inch volume contains over 240 pages of full color photographs. *Counts as 2 of your 3 books.* **$29.95**



56190. INTRODUCTION TO ARTIFICIAL INTELLIGENCE. *Philip Jackson.* A comprehensive survey of the work in artifical intelligence as well as a broad coverage of other fields interested in the question of intelligence. **$18.50**



49080. FROM KNOW-HOW TO NOWHERE: The Development of American Technology. *Elting E. Morison.* One of the leading historians of technology offers a fascinating history of how America developed the astounding technological capacity to design and construct just about anything and go just about anywhere. **$10.00**

only means of preventing overexploitation of the *gaje*. In Europe a proposal for some Romany Zion has even arisen among them. Once long ago in Romania they were enserfed and settled; that influence is strong still, but it has not integrated the Rom. Their travail of hidden superiority preserves among us a lasting example of human wit and ingenuity, a touching, even romantic memorial to conscious solidarity, to the enduring power of the social recognition of "purity and danger."

The metaphorical gypsy—not the real one—is the theme of the book by Robert Krueger. No work of social science, it is an extraordinary evocation in fine photographs and an edited text: a set of first-person narratives by truckers themselves of the way of life of the owner-operator of a five-axle, 18-wheel diesel rig. These lonely blue-collar entrepreneurs, mostly men but also "a few brave women," roam the concrete ribbons of the Interstates as the merchantmen from Tyre roamed the wine-dark sea 3,000 years ago. They are profoundly integrated; indeed, theirs may be the very archetype of American life today. Consider the symbolic intensity of the highway system itself, the diesel fuel and its tax load, the overarching mortgage, the Teamsters Union and General Motors, the citizen-band two-way radio, Radar Alley (where lurks Smokey the Bear), the universal truckers' café and the waitress's high blue hairdo, the disc jockey and his country music, Benzedrine, five national trucking publications—and always one ironic dream: "The problem is that a truck is ... something that can be taxed, but it can't vote. ... The truck driver [is] Don Quixote all right, but ... he charges bureaucracies. He could be so free, like I always dreamed. Workin' hard but havin' fun." There is the dream made visible in the frontispiece rig pounding up the long smooth hill at Raton, N.M., nothing else moving under the wide sky.

I got ten forward gears and a sweet
Georgia overdrive,
I'm takin' little white pills and my
eyes are open wide.
I just passed a Jimmy and a White,
I'm passing everything in sight,
Six days on the road, and I'm going
to make it home tonight.

GEOLOGICAL HAZARDS: EARTHQUAKES — TSUNAMIS — VOLCANOES — AVALANCHES—LANDSLIDES—FLOODS, by B. A. Bolt, W. L. Horn, G. A. Macdonald and R. F. Scott. Springer-Verlag ($25.80). PHYSICAL ASPECTS OF NATURAL CATASTROPHES, by Adrian E. Scheidegger. Elsevier Scientific Publishing Company ($29.25). Each of these remarkably complementary books manages to convey an air of wonder and awe by entirely sober methods of exposition. The first book—the work of three Californians and one Hawaiian—is aimed at the general reader and college student. It is rich with information and is often explicit in discussing mechanisms, but it has not one equation (although there are tables, maps and graphs). It is built on particulars. Perhaps a fourth of the text is given over to case histories: accounts in a page or two of specific events in most of the categories of the subtitle. Dramatic illustrations accompany the text. One photograph taken by an unknown seaman shows a huge seismic sea wave striking the pier at Hilo in 1946. A man is 100 feet from the wave crest. ("He was never seen again.") There is an aerial photograph of Long Beach, Calif., with the contours showing the subsidence at the shore induced by the withdrawal of subsurface oil and gas. There is the great spine of rock protruding from Mont Pelée, funeral monument to the city of the dead below, asphyxiated St. Pierre.

It is odd to see a four-story concrete apartment house lying on its back, structurally intact but surely rather surprised, after the induced liquefaction of the saturated sands at its foundation during the Niigata earthquake of 1964. The worst disasters on record, the Kansu earthquake of 1920 and the Shensi one of 1556, seemed to have been largely the result of an analogous process in wind-deposited loess. A few pages are given to an explanation of this strange phenomenon. The main cause appears to be the increase in pore pressure in a water-soaked granular material. That pressure increases following the repeated alternating stress of earthquake shaking, as the loaded grains settle to a denser packing. The increased pore pressure takes the normal load off the grains, so that the material may become near-liquid and begin to flow under the changing shear stresses. The process is seen in virtually all earthquakes and is of major importance in some.

The second volume looks quite different. Its geophysicist author, trained in Zurich and Toronto and now a professor in Vienna, gives us plenty of equations but no photographs and little particularity. Scheidegger's maps of earthquake regions are postage-stamp drawings, not big, clear spreads. The aim is general understanding. He treats the same topics as the casebook does and adds for good measure a chapter on air catastrophes, from lightning to hurricanes. He sums up the empirical facts (the quantities measured, the range of variability), and then he seeks an analytic description, which for most of the events is mainly a study in the classical mechanics of instability. Both solid elastic theory and fluid dynamics are drawn on. Since the initial conditions are natural and complex and the results that are required are less than precise, he can seek models of some simplicity; elaborations are hardly worth while.

It is a delight to the reader with some knowledge of the mechanics of Newton to see how certain of these awesome phenomena fit cunningly employed mainstays of the high school physics laboratory. On a simple frictional model a landslide will reach out a distance that is merely the height of fall divided by the average coefficient of friction; the height, the speed reached and the rock mass involved are immaterial. Slides ranging up to a volume of about 100 houses fit the theory well with plausible constants. Introducing a coefficient of friction that varies smoothly and simply with slide volume (much smaller friction for much larger volumes), one fits the data very well up to catastrophic slides of an entire mountainside. Perhaps the cause is lubrication by trapped air; we do not know.

Both books pay a good deal of attention to the statistics of unusual events; the estimates of risk and reasonable efforts at alleviation and prevention require that kind of understanding. The reader who wants reliable and yet dramatic facts and the more mathematically trained reader who wants to see what models work, and perhaps to begin making models of his own, will find one of the books rewarding. There is much left to do; many questions of fact and theory remain unanswered. A minor disaster for the reader and his library is the high cost of these by no means very specialized books.

ASTRONOMICAL PHENOMENA FOR THE YEAR 1976, issued by the Nautical Almanac Office, United States Naval Observatory. U.S. Government Printing Office (80 cents). SKY CALENDAR, by Robert C. Victor. Abrams Planetarium, Michigan State University ($2 per year, monthly sheets mailed quarterly, three issues per mailing). Fountainhead of public information on the motions of the bodies of the solar system, the paradigm of Newtonian science, is the computing office of the United States Naval Observatory (with its partner in the United Kingdom). A big tabular hardback of its work, *The American Ephemeris and*

Nautical Almanac, appears yearly, bearing the coordinates of the sun, the moon, the planets and the satellites, and handling authoritatively many finicky questions of timekeeping, geocentric and heliocentric relations, eclipses and occultations. It is genuine prediction: the main tables are ready three years in advance, and so each January we get one of these inexpensive slender pamphlets of tables selected from the bigger volume for the third year ahead. (The 1976 issue first appeared in 1973.) The first 10 pages are mainly taken up with calendar tables, but they always include tables giving to the minute the times of the four phases of the moon all year long, a list of the morning and evening stars and a laconic half page, full of symbols, that details the "planetary configurations" for the year. That is the central source of every sky gazer's guide to the unaided-eye sights of the ecliptic. It is hard to unpack, but the news is all there. Then there are half a dozen pages on the eclipses of the year, including world maps of eclipse tracks. Detailed tables of the sun's and moon's rising and setting times, and the limited motions of Polaris, tabulated for precise latitude fixes, make up most of the other pages. Early and authoritative, the pamphlet has the austere beauty of unadorned numerical prediction.

Not many can enjoy a year of sky-watching guided only by the succinct half-page table of the Nautical Almanac Office. If that work supplies the first and the most data for a dollar, the *Sky Calendar* out of East Lansing is the serious beginner's best buy. For each month there is one sheet of paper. On the reverse a star chart generally appears, sensibly simple, bearing only the chief constellations in the month's evening sky. The other side is made up as a calendar; in each day box there is a bit of text or a neat, tiny drawing of noteworthy objects along the sunset or dawn horizon. (Michigan skylines show an unexpected number of contemplative horses in silhouette.) Text and drawing guide one to the planets, to the bright ecliptic stars as the moon passes them, even to a few sights that can be enjoyed only with binoculars. A special event—a lunar eclipse or bright apparitions of Venus, for example—may sometimes augment the month's star chart, bringing a detailed and helpful summary of the event for the amateur viewer. This work is unappealing typographically, dominated by crowded typewritten copy in grayish offset, but its low cost, explicitness and simplicity are outstanding; handsome is as handsome does.

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

THE NECESSITY OF FISSION POWER

THE NUCLEAR CONTROVERSY. Ralph E. Lapp. Fact Systems/Reddy Kilowatt, Inc., 1974.

PROJECT INDEPENDENCE REPORT. Federal Energy Administration. U.S. Government Printing Office, 1974.

WASH-1400, REACTOR SAFETY STUDY: AN ASSESSMENT OF ACCIDENT RISKS IN U.S. COMMERCIAL NUCLEAR POWER PLANTS. U.S. Atomic Energy Commission. U.S. Government Printing Office, 1975.

ERDA-48. Energy Research and Development Agency. U.S. Government Printing Office, 1975.

AMERICA'S ENERGY FUTURE. Ralph E. Lapp. Fact Systems/Reddy Kilowatt, Inc., in press.

ENVIRONMENTAL HAZARDS IN HIGH-LEVEL RADIOACTIVE WASTE DISPOSAL. Bernard Cohen in *Reviews of Modern Physics,* in press.

THE VOLCANOES OF MARS

VOLCANISM ON MARS. M. Carr in *Journal of Geophysical Research,* Vol. 78, No. 20, pages 4049–4062; July 10, 1973.

MARTIAN SURFACE AND CRUST: REVIEW AND SYNTHESIS. William K. Hartmann in *Icarus,* Vol. 19, No. 4, pages 550–575; August, 1973.

MARTIAN PLANETWIDE CRATER DISTRIBUTION: IMPLICATIONS FOR GEOLOGIC HISTORY AND SURFACE PROCESSES. L. A. Soderblom, R. A. West, B. M. Herman, T. J. Kreidler and C. D. Condit in *Icarus,* Vol. 22, No. 3, pages 239–263; July, 1974.

THE SEARCH FOR NEW FAMILIES OF ELEMENTARY PARTICLES

INTRODUCTION TO HIGH ENERGY PHYSICS. Donald H. Perkins. Addison-Wesley Publishing Co., Inc., 1972.

UNIFIED THEORIES OF ELEMENTARY-PARTICLE INTERACTION. Steven Weinberg in *Scientific American,* Vol. 231, No. 1, pages 50–59; July, 1974.

THE DETECTION OF NEUTRAL WEAK CURRENTS. David B. Cline, Alfred K. Mann and Carlo Rubbia in *Scientific American,* Vol. 231, No. 6, pages 108–119; December, 1974.

PROBING THE WEAK FORCE WITH NEUTRINOS. David B. Cline, Alfred K. Mann and Carlo Rubbia in *Physics Today,* Vol. 28, No. 3, pages 23–31; March, 1975.

A DNA OPERATOR-REPRESSOR SYSTEM

STRUCTURE OF THE λ OPERATORS. Tom Maniatis and Mark Ptashne in *Nature,* Vol. 246, No. 5429, pages 133–136; November 16, 1973.

THE NUCLEOTIDE SEQUENCE OF THE *LAC* OPERATOR. Walter Gilbert and Allan Maxam in *Proceedings of the National Academy of Sciences,* Vol. 70, No. 12, pages 3581–3584; December, 1973.

CONTROL ELEMENTS IN THE DNA OF BACTERIOPHAGE λ. Tom Maniatis, Mark Ptashne and Russell Maurer in *Cold Spring Harbor Symposia on Quantitative Biology,* Vol. XXXVIII, pages 857–869; 1974.

REPRESSOR, OPERATORS, AND PROMOTERS IN BACTERIOPHAGE LAMBDA. Mark Ptashne in *The Harvey Lectures: Series 69.* Academic Press, 1975.

STOMATOPODS

THE AGGRESSIVE AND TERRITORIAL BEHAVIOR OF THE MANTIS SHRIMP *GONODACTYLUS BREDINI* MANNING (CRUSTACEA: STOMATOPODA). Hugh Dingle and Roy L. Caldwell in *Behavior,* Vol. 33, Parts 1–2, pages 115–136; 1969.

INTERSPECIFIC AGGRESSIVE BEHAVIOR IN TROPICAL REEF STOMATOPODS AND ITS POSSIBLE ECOLOGICAL SIGNIFICANCE. Hugh Dingle, Raymond C. Highsmith, Kenneth E. Evans and Roy L. Caldwell in *Oecologia,* Vol. 13, No. 1, pages 55–64; 1973.

ECOLOGY AND EVOLUTION OF AGONISTIC BEHAVIOR IN STOMATOPODS. Roy L. Caldwell and Hugh Dingle in *Die Naturwissenschaften,* Vol. 62, No. 5, pages 214–222; May, 1975.

PALEONEUROLOGY AND THE EVOLUTION OF MIND

A STROLL THROUGH THE WORLDS OF ANIMALS AND MEN: A PICTURE BOOK OF INVISIBLE WORLDS. Jakob von Uexküll in *Instinctive Behavior: The Development of a Modern Concept,* edited and translated by Claire H. Schiller. International Universities Press, Inc., 1957.

FACING REALITY: PHILOSOPHICAL ADVENTURES BY A BRAIN SCIENTIST. John C. Eccles. Springer-Verlag New York Inc., 1970.

THE BRAIN IN HOMINID EVOLUTION. Phillip V. Tobias. Columbia University Press, 1971.

OBJECTIVE KNOWLEDGE: AN EVOLUTIONARY APPROACH. Karl R. Popper. Oxford University Press, 1972.

EVOLUTION OF THE BRAIN AND INTELLIGENCE. Harry J. Jerison. Academic Press, 1973.

MIRAGES

METEOROLOGISCHE OPTIK. J. M. Pernter. Wilhelm Braumüller, 1902.

THEOLOGICAL OPTICS. Alistair B. Fraser in *Applied Optics,* Vol. 14, No. 4, pages A92–A93; April, 1975.

THE PREVENTION OF SMALLPOX BEFORE JENNER

THE HISTORIC EVOLUTION OF VARIOLATION. Arnold C. Klebs in *Johns Hopkins Hospital Bulletin,* Vol. XXIV, No. 265, pages 69–83; March, 1913.

SMALLPOX. C. W. Dixon. J. & A. Churchill Ltd, 1962.

A HISTORY OF EPIDEMICS IN BRITAIN. Charles Creighton. Frank Cass & Co. Ltd., 1965.

THE MEDICAL WORLD OF THE EIGHTEENTH CENTURY. Lester S. King. Robert E. Kreiger Publishing Co., Inc., 1971.

INOCULATION: PRO AND CON. Jason S. Zielonski in *Journal of the History of Medicine,* Vol. XXVII, No. 4, page 447; October, 1972.

MATHEMATICAL GAMES

MAGIC SQUARE PROBLEMS. Henry Ernest Dudeney in *Amusements in Mathematics.* Dover Publications, Inc., 1958.

MAGIC SQUARES. Martin Gardner in *The Second Scientific American Book of Mathematical Puzzles and Diversions.* Simon and Schuster, 1961.

MAGIC SQUARES. John L. Fults. Open Court Publishing Company, 1974.

THE AMATEUR SCIENTIST

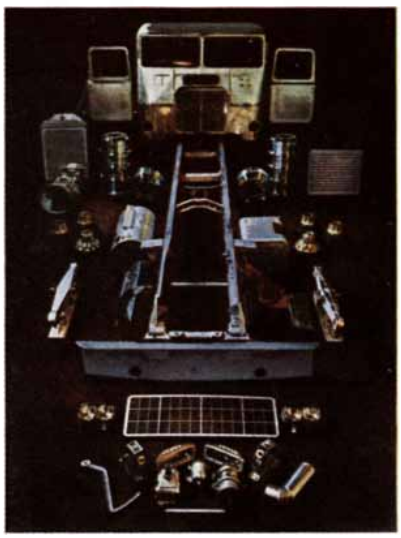
THE THREE-DIMENSIONAL STRUCTURE OF AN ENZYME MOLECULE. David C. Phillips in *Scientific American,* Vol. 215, No. 5, pages 78–90; November, 1966.

STRUCTURE OF CRYSTALLINE ALPHA-CHYMOTRYPSIN: V. THE ATOMIC STRUCTURE OF TOSYL-ALPHA-CHYMOTRYPSIN AT 2 Å RESOLUTION. J. J. Birktoft and D. M. Blow in *Journal of Molecular Biology,* Vol. 68, No. 2, pages 187–240; July 21, 1972.

Truck drivers have always known what motorists need to know about aluminum.

That aluminum can take it. That it can be engineered to perform as well as the material it replaces. Yet, because it's a weight-saver, aluminum gives you a lighter vehicle that saves gas, brake wear, tire wear, even registration costs in some states.

Aluminum has performed so well in trucks, which are built to last hundreds of thousands of miles, that automotive engineers are now evaluating additional aluminum applications for the automobile—specifically, hoods, wheels, trunk decks, bumpers and bumper reinforcements . . . even, in some cases, aluminum doors. They understand better than most of us that American cars must become less expensive to operate, without sacrificing strength. So, they're specifying strong, lightweight aluminum alloys specially designed for passenger cars.



Sturdy aluminum components that save weight in trucks.

The men who drive truck rigs equipped with aluminum cabs, frame rails, radiator parts, bumpers, fuel tanks, suspension parts and forged disc wheels know how well aluminum stands up over the long haul. Under rough treatment. In all kinds of weather. They also know that lightweight aluminum components pay off financially. Because a lighter rig can carry more payload, yet requires less fuel when rolling. Either loaded or empty.

If you'd like a more in-depth look at how aluminum in automobiles helps in the conservation of energy, send for our brochure, *Energy, Aluminum and the Automobile.* Write to Aluminum Company of America, 504-A Alcoa Building, Pittsburgh, PA 15219.

The reasons for using aluminum are found in aluminum itself.

ALCOA

Chevrolet

"You can get it fixed anywhere in the country."
Autoweek-Sept. 20, '75

"...extensive corrosion protection."
Automotive News-Sept. 15, '75

"Keeping in mind that it is a small car, the Chevette interior is marked by considerable roominess."
Road & Track-Oct. '75

"...in the hardcore measurements of the marketplace—price, gas, mileage, appearance (and) overall size ...the Chevette seems a winner."
Car & Driver-Oct. '75

"Should be a standout choice... perfectly suited to fill the U.S. buyer's need for high-fuel-economy (and) low cost transportation."
Popular Science-Oct. '75

"Chevette is a car of the times: lightweight and thrifty... a most pleasing car."
Road & Track-Oct. '75

"May be the most important new U.S. car of the next decade."
Motor Trend-Oct. '75

Prices start at $2899.*
2-seat Scooter $2899.*
Chevette Coupe (shown) $3098.*
The Sport $3175.*
The Rally $3349.*
The Woody $3404.*
*Manufacturer's Suggested Retail Price including dealer new vehicle preparation charge. Destination charge, available equipment, state and local taxes are additional.

40 mpg highway, 28 mpg city,
based on EPA test results with standard 1.4-litre engine, standard rear axle ratio, and standard 4-speed manual transmission. Remember: These mileage figures are estimates. The actual mileage you get will vary depending on the type of driving you do, your driving habits, your car's condition, and available equipment. In California, see your Chevy dealer for EPA mileage figures on California emission-equipped cars.

Chevette

Chevrolet's new kind of American car.

At all 6030 Chevy dealers.

GM
MARK OF EXCELLENCE