SCIENTIFIC AMERICAN

CARTHAGINIAN HEAD

$1.25

February 1975

You'll find fuel injection on some of the fastest cars in the world

On some of the most expensive cars in the world.

And now, through a miracle of modern technology, you'll find it on one of the most inexpensive cars in the world.
The new, advanced '75 Beetle. '75 VW $2895*

This is a uranium mine. Today you see it. Someday you won't.

This is an Exxon mine near Casper, Wyoming. It holds one of the largest deposits of uranium discovered in America in recent years.

The mine processes about 14,000 tons of ore a week. The valuable uranium material in the ore is the base ingredient Exxon uses to fabricate fuel elements. The fuel elements are supplied to nuclear power plants that generate electricity to meet today's energy need.

Within six to ten years, most of the uranium will be extracted, and we'll be finished here But when that day comes, we won't just pull up stakes and pull out.

Almost all the area you see above will be filled in. The land will be contoured to match the adjacent countryside. Then it will be planted.

And someday, if you stand on the spot where our picture was taken, about all you'll see are gently rolling hills, a lake, and perhaps a flock of sheep, or an occasional deer or antelope.

This isn't just something we're going to do. Our reclamation program started before we took the first shovelful of earth out of the mine.

The material we removed to get down to the uranium was spread nearby, contoured and planted. And in 1971, the first grass, shrubs and wildflowers appeared.

Now don't get the impression that we think surface mining is beautiful. It's far from it—as you can see. The beautiful thing is that while today you may see it, Exxon's making sure that someday you won't.

EXXON

Established 1845 **SCIENTIFIC AMERICAN** February 1975 Volume 232 Number 2

ARTICLES

14 **THE FOREIGN MEDICAL GRADUATE, by Stephen S. Mick**
A third of the physicians starting their career in the U.S. were educated abroad.

22 **COMPUTER-MANAGED PARTS MANUFACTURE, by Nathan H. Cook**
Systems for this purpose are a major step on the way to fully automatic factories.

30 **THE MOST PRIMITIVE OBJECTS IN THE SOLAR SYSTEM, by Lawrence Grossman** They are found in the meteorites called carbonaceous chondrites.

46 **CHROMOSOMAL PROTEINS AND GENE REGULATION, by Gary S. Stein, Janet Swinehart Stein and Lewis J. Kleinsmith** What turns genes on?

61 **DUAL-RESONANCE MODELS OF ELEMENTARY PARTICLES, by John H. Schwarz** The hadrons are regarded as strings in multidimensional space.

70 **BIOLOGICAL CLOCKS OF THE TIDAL ZONE, by John D. Palmer**
Many organisms that live along the shore have both a solar clock and a lunar one.

80 **A CARTHAGINIAN FORTRESS IN SARDINIA, by Sabatino Moscati**
Before Carthage fought Rome she dominated this island off the Italian peninsula.

88 **ALFRED WEGENER AND THE HYPOTHESIS OF CONTINENTAL DRIFT, by A. Hallam** How is it that drift was rejected when Wegener proposed it?

DEPARTMENTS

6 LETTERS
8 50 AND 100 YEARS AGO
12 THE AUTHORS
40 SCIENCE AND THE CITIZEN
98 MATHEMATICAL GAMES
104 THE AMATEUR SCIENTIST
109 BOOKS
114 BIBLIOGRAPHY

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: $12 PER YEAR, U.S., POSSESSIONS AND CANADA; $15 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 shows a stylized representation of a woman's head from a recently discovered Carthaginian site in Sardinia. The role of the Carthaginians and their predecessors, the Phoenicians, as traders in the western Mediterranean has long been appreciated, but the Carthaginian exercise of military control over Sardinia in the centuries before its wars with Rome is only now becoming apparent (see "A Carthaginian Fortress in Sardinia," by Sabatino Moscati, page 80). The sculptured head had been stolen from a tomb chamber at the site, Monte Sirai, just before the Department of Antiquities of the Italian government began an investigation of the discovery in collaboration with the Institute of Near Eastern Studies at the University of Rome. An inquiry by the investigators resulted in its recovery. The head still bears traces of the red pigment that originally colored it.

THE ILLUSTRATIONS

Cover photograph courtesy of Sabatino Moscati

Page	Source
14B–21	Graphic Presentation Services, Inc.
22	Ralph Morse
24	Ralph Morse
25	Cincinnati Milacron Inc.
26–29	Ilil Arbel
31	R. S. Lewis, University of Chicago
32	Allen Beechel
33	E. A. King, Jr., University of Houston
34	Allen Beechel
35	L. H. Fuchs, Argonne National Laboratory (*top*); Allen Beechel (*bottom*)
36	Allen Beechel
37	Lawrence Grossman, University of Chicago (*top*); R. H. Beauchamp, Battelle Pacific Northwest Laboratories (*bottom left*); E. Olsen, Field Museum of Natural History, Chicago (*bottom right*)
38	Allen Beechel
42	Kitt Peak National Observatory
46	David E. Comings and Tadashi Okada, City of Hope Medical Center
48–49	George V. Kelvin
50	Gary S. Stein, University of Florida College of Medicine
51–57	George V. Kelvin
62–66	Ilil Arbel
71	John D. Palmer, University of Massachusetts
72–78	Tom Prentiss
80	Sabatino Moscati, University of Rome
81–83	Alan D. Iselin
84–86	Sabatino Moscati, University of Rome
89–96	Lorelle M. Raboni
99	*Mind*
100	Ilil Arbel
101	University of Chicago
102	*The New Yorker*
105–107	Jerome Kuhl

As your introduction to membership in the BOOK-OF-THE-MONTH CLUB®

The most complete and most scholarly dictionary of the English language 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 meaning, origin, history and usage of words in the English language. Until recently, it had been available only as a thirteen-volume set, 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 two volumes of *The Compact Edition.*

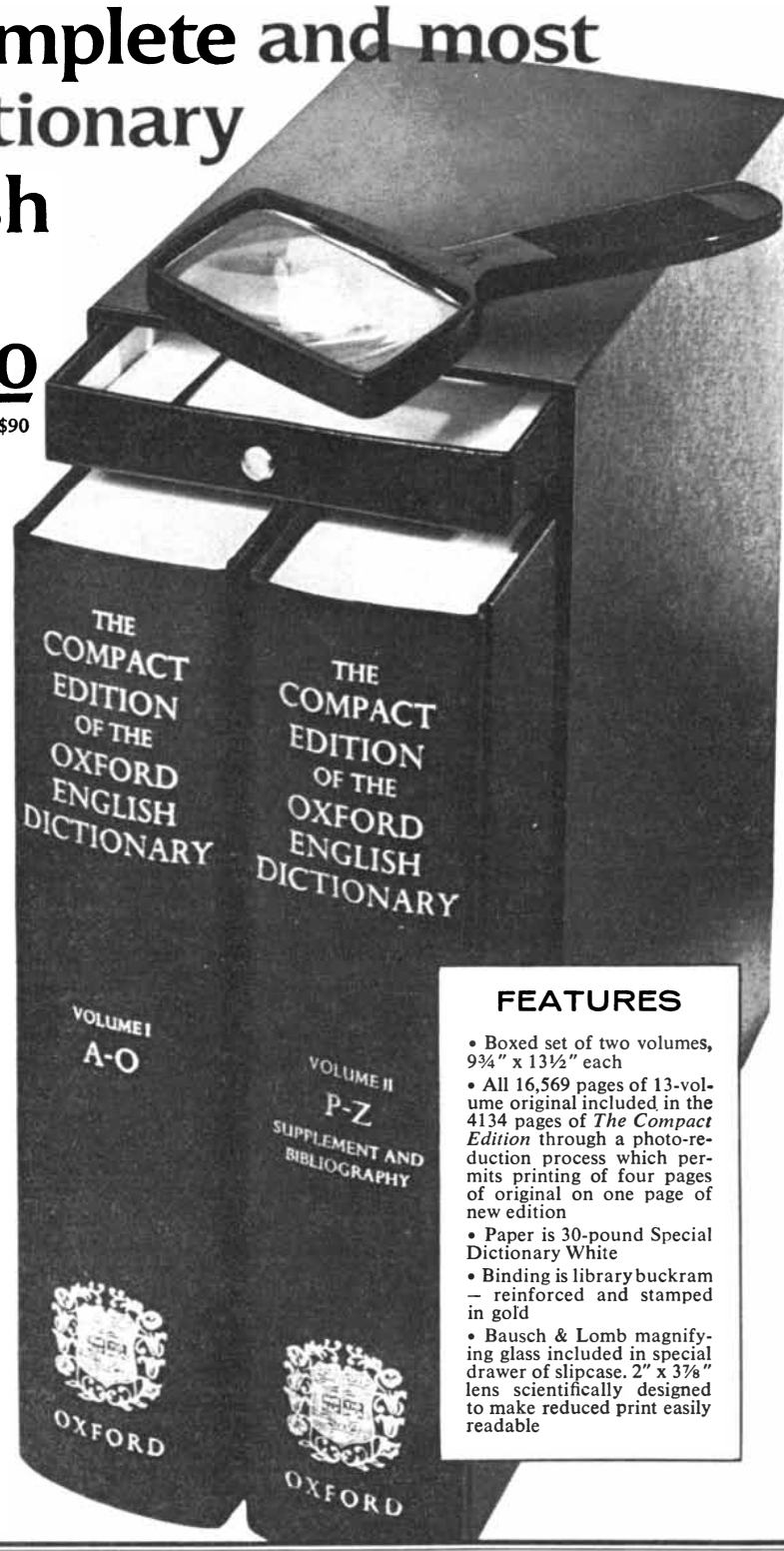
> The *New York Times* book critic Christopher Lehmann-Haupt has said of this edition: "It is something of a miracle. . . . *The Compact Edition* is easier to work with than the original with its 13 separate volumes."

Even more extraordinary, as a trial member of the Book-of-the-Month Club you may obtain *The Compact Edition of the Oxford English Dictionary* 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, most of which are available at substantial discounts – up to **40%** on more expensive volumes. All of these books **are identical to the publishers' editions in content, format, size and quality.**

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 entitle you to obtain a wide variety of books, called Book-Dividends, at astonishing savings – *at least 70% of publishers' list prices.* Under this unique system, more than 100 Book-Dividends will be available to choose from every year, enabling you to acquire a well-rounded library at a small fraction of what you would otherwise have to pay.

FACTS ABOUT MEMBERSHIP

- You will receive the Book-of-the-Month Club *News,* a literary magazine published by the Club fifteen times a year. The *News* describes the coming Selection and scores of Alternates, and will be sent to you approximately every three and a half weeks.
- If you wish to purchase the Selection, do nothing and it will be shipped to you automatically.
- If you do not want the Selection – or you would like one of the Alternates, or no book at all – simply indicate your decision on the reply form always enclosed with the *News* and mail it so we receive it by the date specified.
- If, because of late mail delivery of the *News,* you should receive a Selection without having had 10 days to decide whether you want it, that Selection may be returned at Club expense.



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 3⅞" lens scientifically designed to make reduced print easily readable

BOOK-OF-THE-MONTH CLUB, INC.
Camp Hill, Pennsylvania, 17012

5-A170-2

Please enroll me as a member of the Book-of-the-Month Club and send me *The Compact Edition of the Oxford English Dictionary,* billing me $17.50 (in Canada $19 – publisher's price $99). I agree to buy at least four Selections or Alternates during the first year I am a member, paying in most cases special members' prices. My membership is cancelable any time after I buy these four books. A shipping charge is added to all shipments.

MR. / MRS. / MISS .....
(Please print plainly)

Address.....53

City.....

State..... *Zip*.....

LETTERS

Sirs:

"Musical Dynamics," by Blake R. Patterson [SCIENTIFIC AMERICAN, November, 1974], seems a well-researched and documented treatise—on measuring devices of sound, yes, on music, no! Dr. Patterson, who says he is some kind of bassoonist, plays a helluva graph, and I hope he is happier doing it.

The alleged incapability of fine instrumentalists to observe and execute dynamic indications accurately is sheer drivel. First it must be understood that the science of music writing is inexact and arbitrary as to tempi, notation and dynamics, to say nothing of interpretation. Dynamic indications are to be observed only with respect to (and are a function of) orchestra or ensemble instrumental complement and size, ratio of instrumentation and acoustics of the auditorium or room (when empty, when full). The whims, artistry (or lack of it), idiosyncrasies and "stick" technique (or lack of it) of conductors, scoring weaknesses of particular and lesser composers, sublimation of supporting parts and projection and exposure of important passages, the vintage, period, style, school and ethnicity of music—all these factors are important to an experienced musician. He plays within his own artistic concepts and talents, his preoccupation with instrumental sound production and dexterity, his dedication to the music, but more important, he has to give and take with his immediate instrumental choir, interplayed with other ensembles, in a constant vigil and flux of balance.

A dynamic designation is merely a guide, to be dealt with artistically, with taste, musicality and musical intelligence. A good player can play mezzo-forte any number of ways for as many reasons (it may shock Dr. Patterson to learn), so that *while registering the same reading on a decibel scale* the sounds produced may be *projected*, or *sublimated*, the timbre controlled, the vibrato and harmonics varied! The listening public—civilians, we call them—its composers, critics and conductors are indeed fortunate that so many excellent instrumentalists spend so much time practicing and producing music, rather than talking or writing about it, their eloquence being reserved for performance.

MANNIE WEINSTOCK

Retired horn player
New York

Sirs:

Any new idea faces stern resistance in a discipline as tradition-bound as music. Instrumentalists probably reacted with indignation to introduction of the metronome, yet Beethoven welcomed it as a means to quantify a much abused aspect of performance. He might have welcomed greater control over dynamic levels had the technology been available.

Dynamic notation is surely imprecise. Yet it must represent more than a "guide." Witness its careful application in the Brahms excerpt in my article.

A minimum dynamic range of 30 decibels follows inevitably from the widely accepted hypothesis that different dynamics should be audibly distinguishable. Mr. Weinstock finds no error in this reasoning.

The many complications he cites require even greater dynamic flexibility. A 30-decibel range can delineate six levels only under the most ideal circumstances, such as an unaccompanied solo. Add more instruments or faulty acoustics, and

Scientific American, February, 1975; Vol. 232, No. 2. 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 Harry T. Morris, 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.

Change of address (or other subscription correspondence) should be addressed to Circulation Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017. Please notify us four weeks in advance of change. If possible, kindly furnish an address imprint from a recent issue. Send both old and new addresses or complete and mail Post Office Form 3578.

☐ Please enter change of address.

☐ Please enter subscription.
U.S. and Canada:
☐ 1 year, $12 ☐ 2 years, $22 ☐ 3 years, $30
All other countries:
☐ 1 year, $15 ☐ 2 years, $27 ☐ 3 years, $37

NAME

NEW ADDRESS

OLD ADDRESS

louder fortissimos and quieter pianissimos are needed.

Players are *capable* of adequate contrasts. Their skills remain dormant because conductors fail to encourage them. I regretfully quit after four unchallenging seasons with professional orchestras, confident that only in chamber music would I find dynamics effectively realized. Symphonic standards may have been higher during Mr. Weinstock's career, before dynamically compressed recordings and broadcasts diminished our expectations.

The popular, but unproved, notion that a musician can project or sublimate his sound while maintaining constant loudness is one of many music-psychoacoustic subjects deserving of careful research.

BLAKE R. PATTERSON

Bell Laboratories
Murray Hill, N.J.

Sirs:

Martin Gardner's article on the paradoxical situations that arise from nontransitive relations ["Mathematical Games," SCIENTIFIC AMERICAN, October, 1974] may have helped me win a bet in Rome on the outcome of the Ali v. Foreman world heavyweight boxing title match in Zaïre on October 30.

Ali, though slower than in former years, and a 4–1 betting underdog, may have had a psychological and motivational advantage for that particular fight. But in addition, Gardner's mathematics might be relevant. Even though Foreman beat Frazier, who beat Ali, Ali could still beat Foreman because there may be a nontransitive relation between the three.

I ranked the three fighters against the criteria of speed, power and technique (including psychological technique) as reported in the press, and spotted a nontransitive relation worth betting on:

	Ali	Frazier	Foreman
Speed	2	1	3
Power	3	2	1
Technique	1	3	2

Foreman's power and technique beat Frazier, but Ali's technique and speed beat Foreman. It was worth the bet. The future implications are, however, that Frazier can *still* beat Ali!

ANTHONY PIEL

Vaud, Switzerland

50 AND 100 YEARS AGO

SCIENTIFIC AMERICAN

FEBRUARY, 1925: "Sir Ernest Rutherford recently presented results he has obtained in studying the structure of the atom. Sir Ernest showed how he has been able to smash up the atoms of many elements by bombarding them with powerful particles projected from radium atoms as these break up spontaneously. The projectiles are known as alpha particles and weigh about four times as much as an atom of hydrogen. They carry a double charge of positive electricity, but whenever they can steal two electrons from the atoms that they drive through in their swift flight they settle down into inert atoms of helium gas. By means of a new and more delicate apparatus Professor Rutherford has succeeded in disintegrating the atoms of many of the natural elements, namely nitrogen, aluminum, sodium, potassium, boron, phosphorus, fluorine, magnesium, silicon, sulphur, chlorine and argon. Some of the elements, notably oxygen, have so far stubbornly resisted his best efforts. This exceptional stability may account for the fact that oxygen forms half of the substance of the rocks, air and water taken together. When the elements are arranged in the order of their atomic weights, it is found that the even-numbered elements form 86 per cent of the earth's crust; oxygen is one of the even-numbered elements."

"F. T. Courtney, in a paper before the Royal Aeronautical Society, has made some pertinent observations on the practical difficulties of commercial aviation. The main drawback is irregularity of service, caused by the inability to fly in foggy or very cloudy weather. The airplanes and engines are good enough, Mr. Courtney points out, and lighthouses and beacons may be a great help. Mr. Courtney tells us, however, that it will be necessary for the ship to fly with no visibility at all, day or night, by instruments alone. Navigational instruments are already good enough theoretically, and Mr. Courtney believes that under conditions of continuous cloud or darkness a pilot will be able to find his way from one airdrome to another without colliding with other airplanes and to descend to the ground without striking obstructions."

"The great Western observatories were well prepared for the favorable situation of Mars last summer, and extensive observations were made. Drawings and photographs were made by Slipher at the Lowell Observatory and by Trumpler at the Lick Observatory. A series of Slipher's photographs show in a very beautiful and convincing way how the dark regions grow larger and deeper in hue as the polar caps wane and the Martian spring advances. On one of the photographs the dark markings on the opposite hemisphere, where autumn is changing to winter, are seen to fade. Slipher says, with reason, that the seasonal changes in the dark markings all obey the law of change that we should expect of vegetation. Whether or not vegetation actually does cover much of the planet's surface is harder to determine. If the chlorophyll of the leaves of Martian vegetation is similar to that of terrestrial plants, it should reflect deep red light strongly and the dark markings should show bright when photographed through a suitable screen. Nothing of this kind has been observed, although the failure of the test is not conclusive."

SCIENTIFIC AMERICAN

FEBRUARY, 1875: "Spiritualism, the latest epidemic of superstition, broke out about 25 years ago, and the manifestations were popularly known as spirit rappings. The first mediums were three sisters; their name was Fox. They invented the raps, the rap language and a good part of the spiritualist lingo. They originated the *séance* and drove a lively business. Spiritualism speedily became a recognized institution; there was no lack of mediums; notoriety and money were the substantial incentives; people, it is said, are fond of humbug and pay more liberally for it than for the necessities of life. The majority of people, as at the present day, looked upon spiritualism as a supremely silly thing. The scientific world also treated it with ridicule or with a silence inspired by disgust and contempt. There were investigations, however, and the rapping trick was fairly exposed. The raps were traced to the persons of the Fox girls; the mechanism of the raps was concealed and protected by the defenses of womanhood. To the modest investigators the girls' skirts were

barriers more formidable than stone walls. Of the devices employed the most elaborate and successful was that of a bar of lead suspended by an elastic cord attached to and operated by the leg."

"The library of Asshurbanipal, king of Assyria, found during excavations at Nineveh, shows that science had made no little progress in Asia 2,500 years ago. The library consists of flat, square tablets of baked clay, having on each side a page of closely written cuneiform cursive letters. The great majority of these tablets are now in the British Museum, and they have been found to contain the remains of a grammatical encyclopedia. There are also fragments of many mathematical and astronomical treatises."

"The Hudson River ice crop for 1875 has now been harvested and is one of the largest and finest ever gathered. The blocks average 14 inches in thickness, and the total quantity secured is about two million tuns, or 70 million cubic feet. This enormous supply of ice will be chiefly consumed in the city of New York. It is brought down the Hudson River from the great ice houses, which are located at the water's edge, in large barges towed by steam. It is delivered directly from the barges to the ice carts, and in them is conveyed to the doors of private dwellings. From a quarter of a tun to a half a tun a month is a common supply for a small family. The price charged is from $15 to $30 a tun."

"We have recently perused a very interesting paper by Dr. Barnard of Columbia College, in which the writer, in his charming style, discourses on 'Aerial Navigation.' As many of our readers are devising plans for sailing in the air, we think it well to give a brief *résumé* of Dr. Barnard's article: As birds fly with wings, it has occurred to man to employ the same device, only to meet with failure. The reason is obvious: a bird has sufficient strength to fly and a man has not. Hence the conclusion that if a man wishes to fly, he must use some artificial motor to drive the necessary mechanism. In regard to this mechanism, it appears that a revolving wheel, such as a propeller, is better than a pair of wings. The method of moving the aerial vessel, however, does not present so many difficulties as the means to be provided for keeping it in the air and enabling it to rise or descend at the pleasure of the navigator. The simplest mechanism seems to be a fixed plane surface in an inclined position, which will sustain the vessel when it is put in motion."

Again
Mercedes-Benz quietly astonishes the engineering world.
Announcing the new Mercedes-Benz 300D. The world's only 5-cylinder automobile.

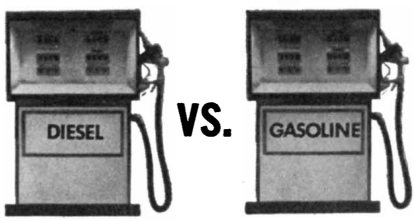
Internal workings of the new 300D, the world's only 5-cylinder automobile.

Raising the eyebrows of the engineering world is not new to Mercedes-Benz. We did it when we invented the automobile in 1886. We did it when we introduced the world's first production Diesel passenger car in 1936. Now, with the introduction of a 5-cylinder automobile – the new 300D – we've done it again.

The 5-cylinder 300D clearly flies in the face of engineering convention. It establishes a whole new category of Diesel – a whole new category of automobile.

Why a 5-cylinder?

Mercedes-Benz wanted to make a performance Diesel. A new car that had much greater horsepower and torque, but did not sacrifice the economy and reliability of previous Mercedes-Benz Diesels.



Countrywide, Diesel fuel averages about 3¢ less than gasoline. So the new 300D goes more miles on cheaper gallons.

Nearly four decades of Diesel experience told the Mercedes-Benz engineers that a major power increase would have to come from an entirely different engine. Not one with bigger cylinders – one with *more* cylinders.

If engineering convention were to be served, the next step up from our standard 4-cylinder Diesel should have been a 6-cylinder engine. But because of the strength necessities particular to all Diesels, it follows that a 6-cylinder Diesel would have to be substantially larger and heavier than a 4-cylinder.

The Mercedes-Benz solution was characteristically unconventional: A 5-cylinder Diesel engine.

The most advanced Diesel

Even for the Mercedes-Benz engineers, the 5-cylinder Diesel engine was quite a challenge. But their efforts were worthwhile. They made it work.

The new engine's size was set at 3 liters. Torque was raised nearly 20%. Horsepower was up a full 25% over any previous Mercedes-Benz Diesel. The result is the new 300D, the most advanced Diesel passenger car the world has ever seen.

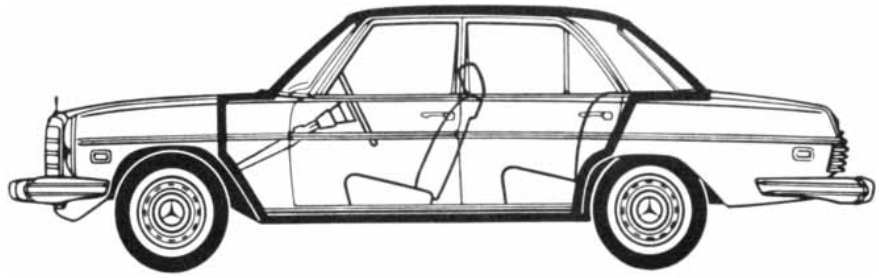
24 mpg in town, 31 on the highway

Increased performance does not mean decreased economy. According to the most recent Federal E.P.A. test report, the new 300D can get up to 24 miles in city driving, and up to 31 miles per gallon on the open highway. And since Diesel fuel is usually cheaper than gasoline, the 300D goes more miles on cheaper gallons.

Performance, economy, luxury

While the new 300D acts like an economy car, it doesn't look or feel like one. In every major appointment, it is equal to the Mercedes-Benz 280. It blends performance and economy in a luxury package.

A final thought. No manufacturer anywhere in the world has a greater safety involvement than Mercedes-Benz. Evidence of the decades of Mercedes-Benz safety engineering is everywhere in the 300 Diesel. Before the 300D, nobody had heard of a 5-cylinder Diesel car, much less seen one. So, it's only natural that the 300D should cause some very educated, engineering heads to be scratched. But now you can examine one, sit in one, drive one. It's waiting for you. Arrange a test drive with your local Mercedes-Benz dealer.



Safety first. A rigid, steel safety cell encloses the entire passenger compartment.

Mercedes-Benz

Engineered like no other car in the world.

THE AUTHORS

STEPHEN S. MICK ("The Foreign Medical Graduate") does research on various aspects of public health at the Yale University School of Medicine. After graduating from Stanford University in 1965 Mick spent two years with the Stanford Overseas Campuses Program in France. He returned to the U.S. in 1967 to study sociology at Yale, where he received his Ph.D. in 1973. Before taking up research full-time he taught sociology at Middlebury College and at Yale. In addition to studying the migration of physicians to the U.S. Mick reports that his other research interests "center on the sociology of occupations, labor markets and mobility. I have engaged in some research on the effects of plant shutdowns as a specific example of the larger issue of the forces affecting migration.... I am also contemplating doing a research project on the medical effects of unemployment. As for extracurricular activities, my hobbies include following the New England Patriots professional football team and doing a lot of cross-country running."

NATHAN H. COOK ("Computer-managed Parts Manufacture") is professor in charge of the materials-processing laboratory at the Massachusetts Institute of Technology. After being graduated from high school at the age of 21 (he had joined the Navy at 17 and served for three years on a destroyer in the Pacific), Cook proceeded to obtain "an all-M.I.T. education in mechanical engineering: S.B., 1950; S.M., 1951; M.E., 1954; Sc.D., 1955." He has been a member of the M.I.T. faculty ever since, specializing in "materials, materials processing, metal removal, machine tools, controls, vibration, applied mechanics, friction, wear and instrumentation." He spent most of 1968–1969 in India, where he is an adviser to the Birla Institute of Technology and Science in Rajasthan; he performs a similar service for the University of Benin in Nigeria. Since 1970 Cook and his family have lived in a new dormitory at M.I.T. where he serves as housemaster. Of his four children, aged 23, 20, 17 and 11, the eldest, a daughter, received her S.B. (in mechanical engineering) from M.I.T. in 1973 and his second eldest, a son, is a senior (also studying mechanical engineering) at M.I.T.

LAWRENCE GROSSMAN ("The Most Primitive Objects in the Solar System") is assistant professor of geochemistry at the University of Chicago. A native of Toronto, Grossman was graduated summa cum laude with honors in chemistry and geology from McMaster University in 1968. He then went to Yale University, where he obtained his doctorate in 1972. Since then his name has appeared as author or coauthor of more than 30 scientific papers dealing with the composition of the moon, meteorites and planets and with the early chemical history of the solar system. His work won him the W. E. Ford prize in 1971 and the F. W. Clarke medal in 1974. Among his outside interests he lists "mineral collecting, fishing in northern Ontario, hiking in the Alps, eating and professional football and hockey."

GARY S. STEIN, JANET SWINEHART STEIN and LEWIS J. KLEINSMITH ("Chromosomal Proteins and Gene Regulation") have collaborated for several years in spite of a considerable geographic separation. The Steins work together at the University of Florida College of Medicine. Gary Stein received his Ph.D. from the University of Vermont in 1969, and spent three years in the pathology department at the Temple University School of Medicine before moving to Florida in 1972. Janet Swinehart Stein received her B.S. in chemistry from Elizabethtown College in Pennsylvania in 1968 and is currently a candidate for a Ph.D. from Princeton University. Gary Stein lists skiing and other outdoor sports high among his nonprofessional interests, whereas Janet Stein writes that she enjoys music and reading. She adds: "I am currently trying to determine whether or not I shall be successful at skiing." Kleinsmith, professor of zoology at the University of Michigan, obtained his Ph.D. from Rockefeller University in 1968 after spending two years in medical school. "My outside interests," he writes, "include elucidating the mechanism of action of marihuana and attaining an understanding of altered states of consciousness."

JOHN H. SCHWARZ ("Dual-Resonance Models of Elementary Particles") is a research associate in theoretical physics at the California Institute of Technology. After majoring in mathematics as an undergraduate at Harvard University, Schwarz switched fields in graduate school, acquiring a Ph.D. in physics from the University of California at Berkeley in 1966. He taught at Princeton University before going to Cal Tech in 1972. He writes: "I have worked almost exclusively on dual-resonance theory since its inception six years ago and have published numerous papers dealing with it." He also enjoys swimming ("about a mile each day"), tennis, hiking and playing the flute ("preferably in small chamber groups").

JOHN D. PALMER ("Biological Clocks of the Tidal Zone") is professor and chairman of the department of zoology at the University of Massachusetts. A graduate of Lake Forest College, Palmer obtained his Ph.D. from Northwestern University in 1962. After a postdoctoral year at the University of Bristol, he joined the faculty of New York University, where he taught and did research for 10 years until taking up his present posts last semester. Since 1960 he has spent his summers as an instructor and investigator of "all aspects of clock-controlled biorhythms" at the Marine Biological Laboratory at Woods Hole. An "avid" saltwater fisherman, he is in the process of building a self-designed house on Cape Cod.

SABATINO MOSCATI ("A Carthaginian Fortress in Sardinia") is professor of philology at the University of Rome, where he also directs the Institute of Near Eastern Studies. The organizer of numerous archaeological expeditions in the Mediterranean region, Moscati has also lectured extensively at universities in Europe and the U.S. He currently serves as president of the National Academic Union of Italy and is a member of the Academy of Lynxes, Italy's leading scientific society. A number of his books have been published in English, including *Ancient Semitic Civilizations* (1957), *The Semites in Ancient History* (1959), *The Face of the Ancient Orient* (1960), *Historical Art in the Ancient Near East* (1963) and *The World of the Phoenicians* (1968).

A. HALLAM ("Alfred Wegener and the Hypothesis of Continental Drift") teaches geology and mineralogy at the University of Oxford. His bachelor's degree and his Ph.D. are from the University of Cambridge. He has taught at the University of Edinburgh, Stanford University and McMaster University in Ontario. This spring he will be a visiting professor at Harvard University. The author of an earlier article in SCIENTIFIC AMERICAN ("Continental Drift and the Fossil Record," November, 1972), Hallam writes that his current interests "are in the field of Jurassic geology and palaeontology, with my interest in the fossils being (a) patterns of evolution, (b) ecology and (c) biogeography."

Audition free for 10 days Karl Böhm's superb Deutsche Grammophon recording of

Mozart's Six Greatest Symphonies

SPECIAL BONUS: Mozart's Piano Concertos Nos. 12 and 26

All 4 International Award-Winning Records For Less Than The Price Of 1!

Your Introduction To The Unique Money-Saving Program That Offers Free Auditions Of Prize-Winning Recordings By The Leading Concert Artists Of Our Time!

Mozart—music'sgreatest natural genius! Wolfgang Amadeus Mozart — divinely gifted beyond any other musician who ever lived! And into his six greatest symphonies he poured a multitude of his most astonishingly beautiful, incredibly moving inspirations!

Symphony No. 41, "Jupiter"
Symphony No. 40 in G-minor
Symphony No. 39 in E-flat
Symphony No. 38, "Prague"
Symphony No. 36, "Linz"
Symphony No. 35, "Haffner"
Extra! Symphony No. 32

Now you are invited to hear these miraculous works in their finest recording . . . interpreted with extraordinary empathy by Karl Böhm . . . played to perfection by the Berlin Philharmonic Orchestra . . . captured in unsurpassed stereo realism by Deutsche Grammophon. So outstanding is this recording that it has won *three* of the music world's most eagerly sought honors: the Grand Prix International du Disque, Edison Award and Deutsche Schallplatten Prize! In addition, you will also receive Mozart's Piano Concertos Nos. 12 and 26, "Coronation," superbly performed by soloist Geza Anda with the Salzburg Camerata Academica—winner of the coveted Grand Prix des Discophiles.

Now enjoy and keep these 4 superb albums for less than the price you'd pay for 1!

Because these magnificent recordings have met with almost unprecedented acclaim, they have been chosen to introduce you to The Great Awards Collection, a totally new concept in home listening. You may enjoy all eight masterpieces, on four superb-quality, imported records, for 10 days absolutely free. Then keep all four if you wish, for only $6.98 *(that's less than the price you'd pay for just one record)!* Simply mail the attached card or coupon today.

Collect the world's greatest music—only in award-winning albums!

As a member of The Great Awards Collection you will receive only the finest recordings of prize-winning concert performances by top orchestras, conductors and soloists. Distinguished jurors each year select, from the hundreds of classical releases, the "golden few" that will be offered to members of The Great Awards Collection. Every month you'll enjoy free 10-day audition privileges on each award-winning single LP. Of course, you have the option of keeping or returning each one so you'll never waste a penny on disappointing purchases!

Special half-price bonus offer saves you money!

In addition to great music, you'll enjoy great savings with our half-price bonus plan. For every record you buy at our low members' price (always well below suggested retail), you may choose another one from a list of award-winners and other critically-acclaimed LP's and pay just half that price! A truly economical way to build your library of superb classical recordings.

How many records are you committing yourself to buy when you return the attached coupon? *None at all!* Even your introductory set of *Mozart's Six Greatest Symphonies* and *Piano Concertos 12 and 26* comes to you *on approval.* Audition it free for ten days, then either return it—or keep it for only $6.98 for all four records (plus a small postage/handling charge).

Here at last is the ideal way to acquire, in easy stages, a connoisseur's record library without wasting a penny on disappointing purchases. You listen at home to every award-winning selection before deciding whether to buy it! Please act today! Begin your money-saving, no-obligation membership in The Great Awards Collection by mailing the attached card or coupon for your free trial, with four-for-less-than-the-price-of-one purchase option of this essential collection of Mozart's Six Greatest Symphonies and exquisite Piano Concertos Nos. 12 and 26.

Eight ways The Great Awards Collection helps you enjoy fine music more than ever!

1. The greatest music by the world's immortal composers.
2. A major, award-winning recording each month.
3. Superior imported LPs—silent surfaces.
4. Free ten-day trials of all selections.
5. All selections always far less than retail price.
6. Extra savings with Half-Price Bonus Plan.
7. No obligation to buy—no minimum purchase.
8. Money-saving introductory offer.

Send no money now—Mail coupon TODAY!

Q90

Listen for 10 days free. Keep all 4 for less than the price of 1!

The Great Awards Collection • 175 Community Drive • Great Neck, N.Y. 11025

Please send me, for my free audition, the triple-award-winning recording of Mozart's Six Greatest Symphonies by Karl Böhm, together with Geza Anda's award-winning performance of Mozart's Piano Concertos Nos. 12 and 26. I may return the set after 10 days and owe nothing, or keep it and pay only $6.98* for all four Deutsche Grammophon records, plus a small postage/handling charge. This is all four records for less than the regular price of one!

I will also receive a free ten-day audition each month of another award-winning recording of great music. For every one of these records I buy at your members' price (always below suggested retail), I may also choose one record at half that price from your special list. I am not obligated to buy any minimum number of records. I may cancel my membership at any time.

*Sales tax added for New York residents.

NAME (please print)

ADDRESS APT. NO.

CITY STATE ZIP

Offer limited to the connecting 48 states. Offer expires 9/1/75. Limit: one membership per household. Only new members eligible. NOTE: All applications are subject to review, and we reserve the right to reject any application

29807/Q90

Established 1845 SCIENTIFIC AMERICAN February 1975 Volume 232 Number 2

The Foreign Medical Graduate

A third of the physicians now beginning their career in the U.S. received their education abroad. This substantial proportion reflects underlying flaws in the American system of delivering health care

by Stephen S. Mick

As recently as 1950 most of the physicians beginning their career in the U.S., as measured by the number of internships and residencies filled, were graduates of medical schools in the U.S. and Canada. The ratio was almost 10:1, that is, 10 graduates of North American medical schools for each graduate of a foreign medical school. Today the ratio is 2:1. In five of the seven years through 1973 the number of foreign medical graduates obtaining American visas was larger than the number of graduates of all American medical schools. By the end of 1973 more than a fifth of the physicians practicing medicine in the U.S. and Canada (77,660 out of 366,379) had been to medical schools elsewhere.

Clearly the graduates of foreign medical schools have had a strong and increasing numerical impact on the practice of medicine in the U.S. Have they also had a qualitative effect? The answer to that question is less clear; indeed, the question is the subject of considerable debate among people who pay close attention to the system of health care in the U.S.

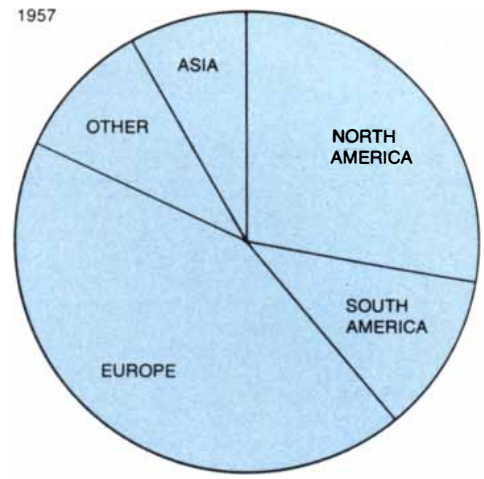
Although the U.S. is by no means the only country receiving foreign physicians, it outdoes all other countries in the number of physicians and in the number of originating countries involved. Before 1965 most of the foreign medical graduates arriving in the U.S. came from other countries of the Western Hemisphere and from Europe. Since 1965 Asian sources have predominated, chiefly as a result of changes in the immigration law under the Immigration Act of 1965.

Another aspect of the migration is that in general the countries contributing the largest numbers of physicians to the U.S. have the fewest physicians available for their own population. (The major exceptions are Canada and the United Kingdom.) In other words, the underdeveloped nations are educating thousands of physicians who ultimately end up practicing in what is perhaps the most developed nation of all.

The migration raises several other questions that are deceptively simple. How is it that a nation with the immense medical resources of the U.S., including the high ratio of one physician for every 650 people, is also the largest importer of foreign physicians? What is the explanation for the fact that a fifth of the nation's physicians, a third of the interns and residents, a fourth of the psychiatrists, a third of the full-time hospital physicians and almost half of the physicians newly licensed in 1973 are the product of foreign medical schools? How can this large and increasing inflow be occurring at a time when American medical schools are turning away more applicants than ever before, even though both the number of schools and their enrollment have increased in recent years?

The search for answers to these questions begins with an examination of changes in migration patterns in the 20th century and with a look at a series of postwar events affecting both immigration policy and the staffing of hospitals. At the outset one should view the flow of foreign medical graduates in the larger context of professional migration over the past 70 years. Between 1901 and 1910 only 1 percent of the immigrants to the U.S. were classified as "professional, technical and kindred workers," a category that includes physicians. By the decade from 1961 through 1970 the category constituted 10.2 percent of the total.

A factor contributing to the increasing



REGIONAL DISTRIBUTION of countries contributing foreign medical graduates to

migration of physicians is the unceasing demand by hospitals for house staff, meaning interns and residents [*see top illustration on page 19*]. Although the supply of physicians from both American and foreign medical schools has risen steadily, it has yet to fill all the positions the hospitals offer. For years the number of unfilled positions has ranged between 4,000 and 9,000 per year.

In part this unfilled demand and the rising flow of foreign medical graduates to help meet it reflect various efforts by the U.S. since World War II to encourage educational exchange programs. For example, the exchange-visitor program enables any foreign student who is in the U.S. under an approved program to remain until his studies are completed. Although it was explicitly stated that such programs were not intended to help hospitals meet their needs for staff, the spirit of the law has been ignored. By 1961 nearly half of the some 2,600 exchange-visitor programs then in effect were sponsored by hospitals.

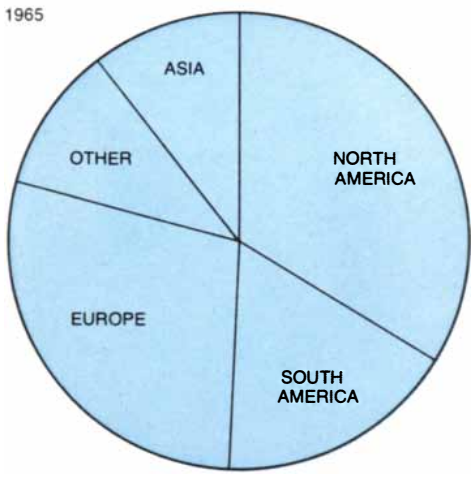
Because of concern in certain countries about the loss of medical graduates the U.S. Department of State instituted in 1972 a "skills list," which was designed to enable interested countries to specify what occupations they thought were short of people. Anyone in such an occupation who has come to the U.S with financial aid from his government or on an exchange-visitor visa must return to his home country for two years after he finishes his American studies. For reasons that are not altogether clear, many nations have taken little or no advantage of this option to curb the outflow of physicians and other health-care workers.

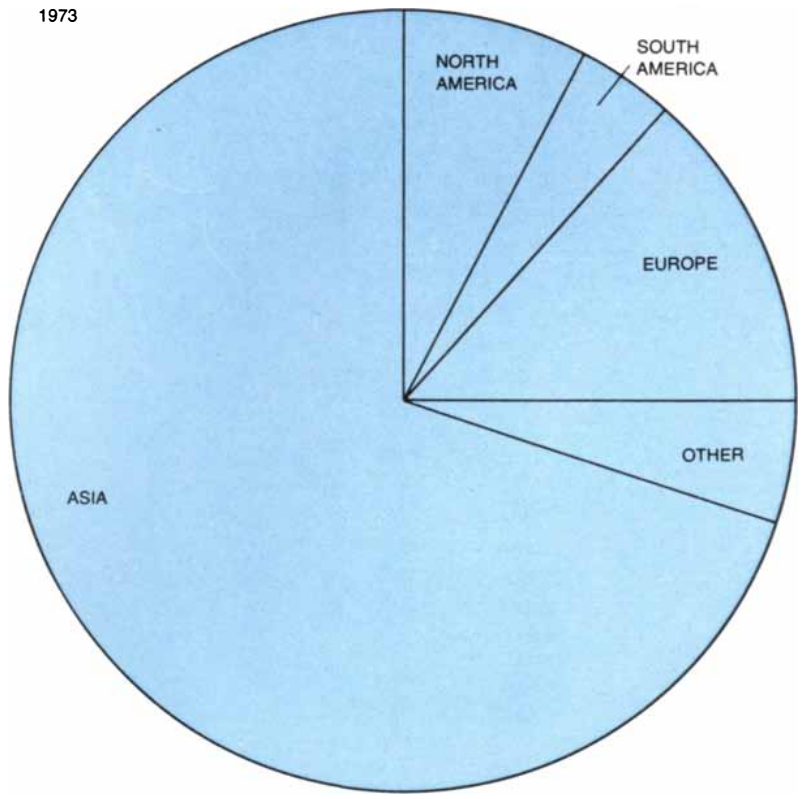
Nonetheless, the number of foreign medical graduates coming to the U.S. on exchange-visitor visas has declined in recent years. During the same period, however, the number entering on standard immigrant visas has increased. The reason is the Immigration Act of 1965, which abolished the long-standing system of quotas based on national origin. It sets a ceiling of 120,000 immigrants per year from countries of the Western Hemisphere, with no country quotas and acceptance on a first-come, first-served basis. The ceiling for immigrants from outside the Western Hemisphere is 170,000, with no more than 20,000 from one country. Physicians and surgeons receive preference over certain other kinds of immigrants, but in most cases the country quotas are so generous that physicians can immigrate without invoking the preference. The effect of the law has been twofold: more physicians and many more Asian physicians are immigrating.

So far I have described what could be called "pull" factors, which originate in the U.S. and tend to attract physicians. One can also cite certain "push" factors, which originate in the home countries and tend to encourage physicians to leave. A general push factor is the tendency of medical schools in underdeveloped countries to pattern themselves after the medical schools in the developed countries of Europe and North America. It is a science-oriented and specialty-oriented training, and a graduate of such a school in an underdeveloped country is therefore predisposed to seek advanced education in places where this kind of training is the norm.

Another type of push factor is specific to the country from which the physician comes. In a survey of foreign medical graduates in the U.S. that my colleagues and I at the Yale University School of Medicine have been conducting for several years the two reasons most often given for emigrating are "Better training in the U.S." (69 percent) and "Political factors" (8 percent). A closely related reason appears to explain why the physician who has left his native country usually does not return, namely his judgment that the possibility of developing a satisfactory practice there is severely limited compared with his prospects in the U.S.

Here one encounters a general problem that has been cited by a number of observers. Although the need for physicians in underdeveloped nations may be substantial, those nations have little capacity for employing physicians. Such a



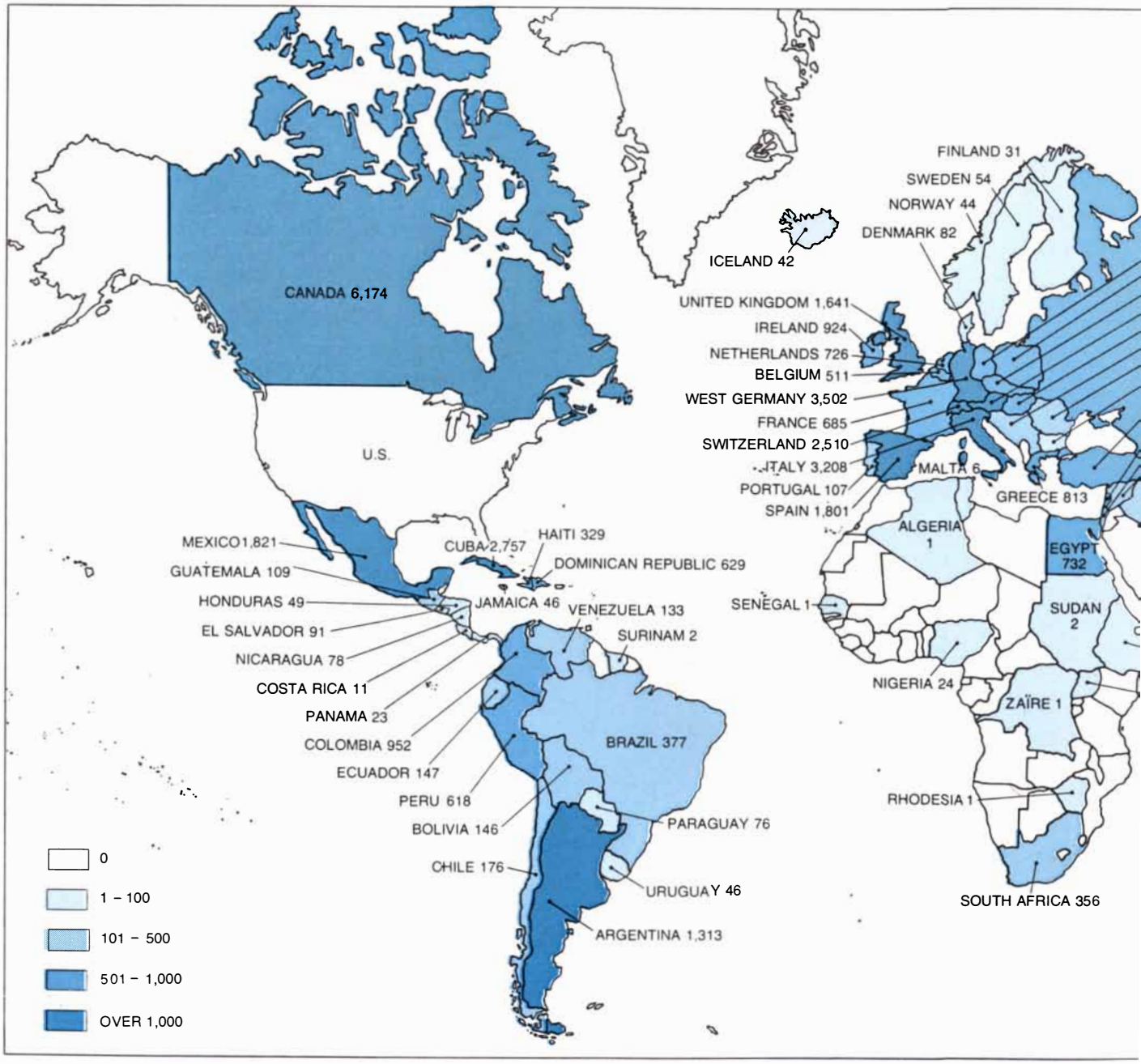


the U.S. is shown for three years. The charts reflect the region of last permanent residence of the physicians admitted to the U.S. as immigrants in each year. They also reflect by their relative size the total immigration each year, from 1,990 in 1957 to 7,119 in 1973.

capacity includes several things that the underdeveloped countries have in short supply: an adequate system for delivering health care, sufficient opportunity to advance within the system and attractive economic prospects. Moreover, most physicians would rather work in large cities than in rural areas, and the underdeveloped countries offer too few opportunities for urban practice.

As the flow of incoming physicians began to increase after World War II, demands arose in the U.S. for the establishment of standards of quality to assist the difficult process of evaluating foreign medical graduates from many nations and from an even larger number of medical schools, which ranged over the spectrum from poor training to excellent training. The first effort in this direction was a list of 39 foreign medical schools, mostly in Europe, drawn up by the Council on Medical Education of the American Medical Association and the executive council of the Association of American Medical Colleges. The criterion for inclusion on the list was that in the judgment of the two groups the schools provided a quality of training equal to the quality found in most American medical schools.

The people involved in licensing or employing foreign physicians soon found the list to be inadequate, particularly since it consisted mainly of European schools at a time when increasing numbers of foreign medical graduates were from schools outside Europe. As a replacement for the list the Educational Council for Foreign Medical Graduates was established in 1954 by a committee made up of representatives of the American Medical Association, the Association of American Medical Colleges, the



PLACES OF EDUCATION of the foreign medical graduates practicing in the U.S. in 1970 are indicated. At the time some 58,000 physicians who received their medical education in other countries were at work in the U.S. The figure associated with the name of

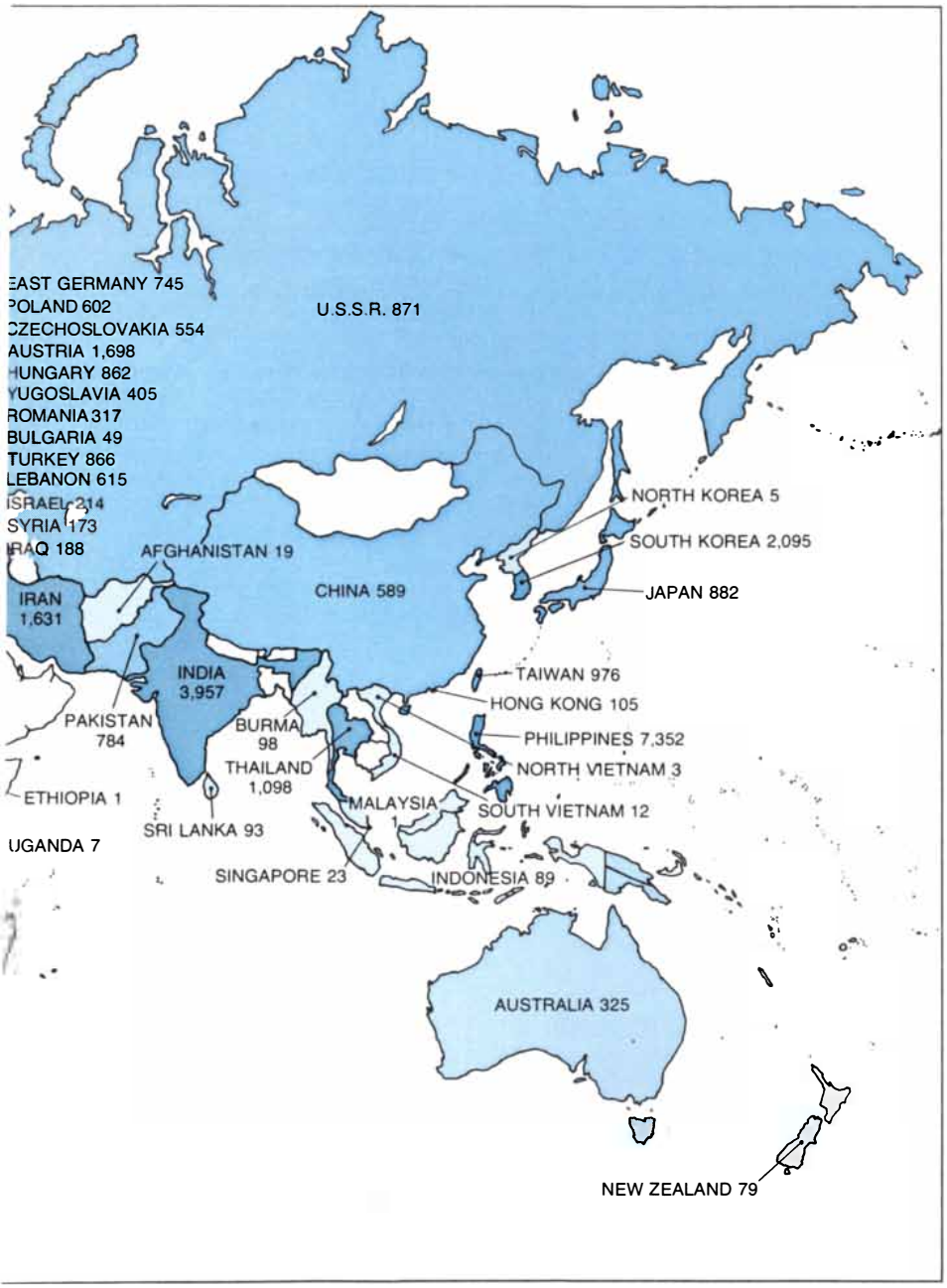
American Hospital Association and the Federation of State Medical Boards. In 1958 the council began to administer semiannually a standardized examination for all foreign-trained physicians, including the growing number of Americans receiving their medical training abroad.

The examination is in two parts. The first part tests general medical knowledge; its questions are taken from an examination drawn up by the National Board of Medical Examiners and normally given to American students in their fourth year of medical school. The second part tests knowledge of the English language.

Since the medical questions come from previous National Board examinations and the success of American students in answering each question is therefore known, one can compute the rate at which American students could be expected to pass the council's examination and compare it with the performance of foreign medical graduates. The comparison has always been unfavorable to the foreign graduates. From 35 to 40 percent of them will achieve a passing score on the first try, whereas the "expected" rate for American medical graduates is more than 90 percent. This comparison is frequently employed to call into question the competence of foreign medical graduates—a point to which I shall return.

As matters now stand, a foreign medical graduate cannot engage in anything more than a limited type of practice in the U.S. without passing the examination administered by the Educational Council for Foreign Medical Graduates. The convention of obtaining the council's certificate is now established, and it represents the major attempt by the medical profession to establish standards for foreign medical graduates. Nonetheless, the council's examination has its critics, who argue that it provides too easy a path for foreign medical graduates and promotes a system of dual standards circumventing the entire process of quality control that is assumed to exist in American programs. As a result, it is argued, the U.S. has a "two-class" medical profession, with American medical graduates constituting the upper class and foreign medical graduates the lower class.

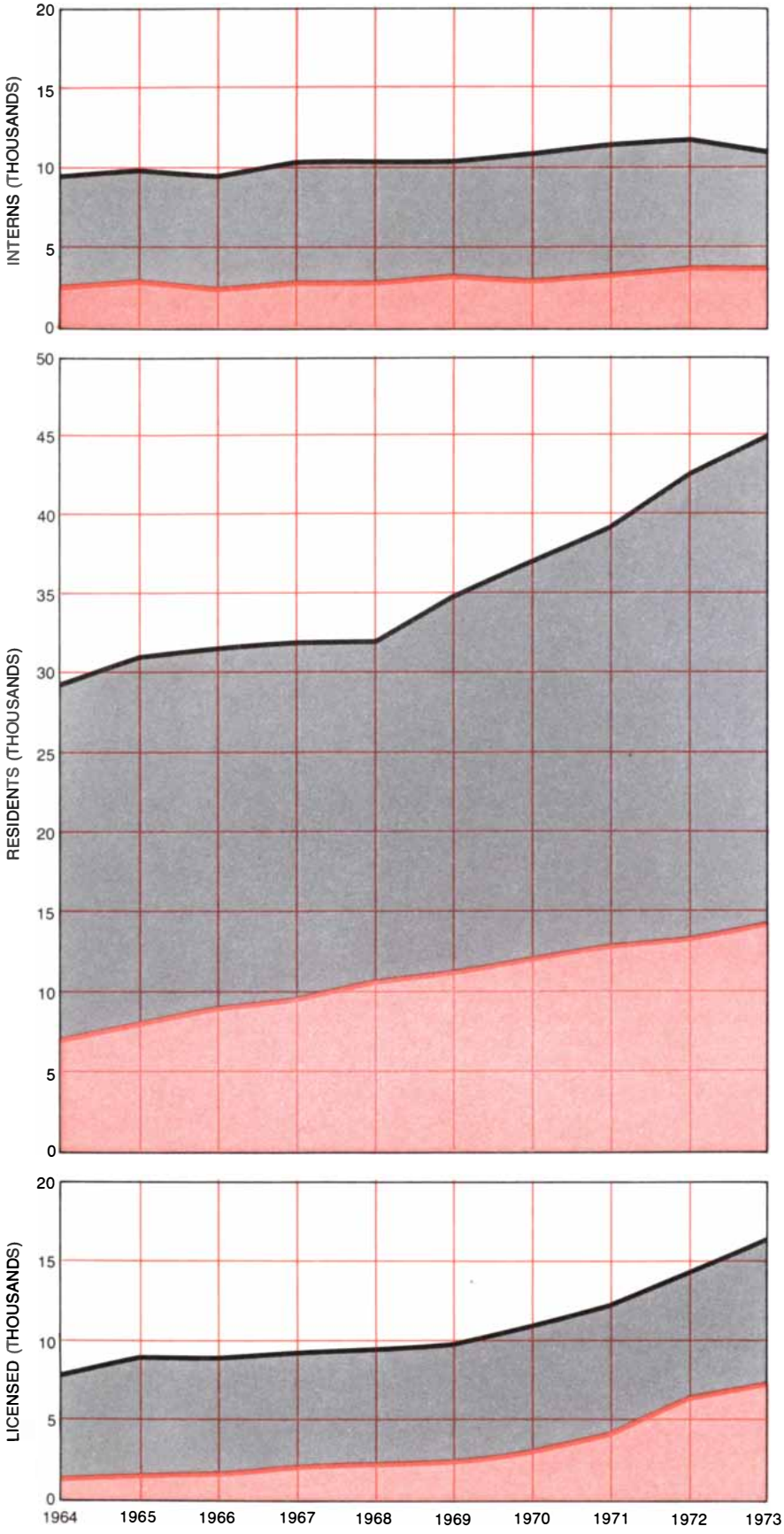
It is certainly true that foreign medical graduates tend to differ from American medical graduates in the roles they fill. They are heavily represented in house-staff positions and among full-time hospital-based physicians. They are underrepresented in private practice and in administrative jobs. Geographically they are heavily concentrated in the Northeast and the Great Lakes states. They fill definite shortages in urban environments and in nonaffiliated hospitals, that is, hospitals without ties to medical teaching programs. With certain exceptions the patterns seem to support the two-class hypothesis.



each country shows how many of those physicians received their medical degree in that country. Major sources can be identified at a glance by means of the shading on the map.

All the foregoing remarks constitute what I have intended to be a descriptive section of this article. I now turn to more argumentative matters concerning the current debate over foreign medical graduates and the various assertions that they constitute a problem. It is not entirely rhetorical to ask if there is, in fact, a problem. As sociologists have pointed out, what may constitute a problem for society and what is defined as a problem are not necessarily the same.

A statement that summarizes the views of the people who see foreign medical graduates as a problem was made at the 1972 meeting of the American Psychiatric Association by E. Fuller Torrey and Robert L. Taylor, who are



TEN-YEAR TREND in the number of graduates of medical schools in the U.S. (*gray*) and of foreign medical schools (*color*) beginning practice in the U.S. is charted. The top chart reflects the number of physicians employed as interns, the middle chart the number employed as hospital residents and the bottom chart the number obtaining a license from a state to practice medicine. The two curves in each chart are cumulative, so that by reading them together one can arrive at the total number of physicians in each category each year.

American-educated physicians. Their title was "Cheap Labor from Poor Nations." The argument, which relates to psychiatrists but can by extension be applied to all types of foreign medical graduates, runs as follows.

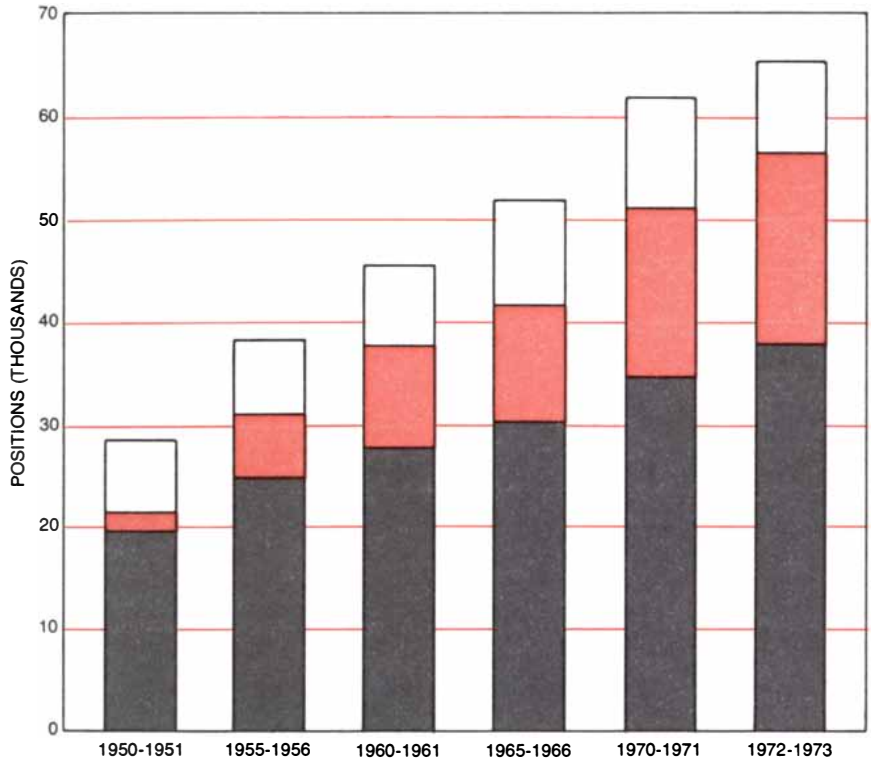
Foreign medical graduates practice in institutions (such as city hospitals, prison hospitals, state hospitals and institutions for the mentally retarded) that American psychiatrists regard as undesirable. The reasons are inaction and lack of planning by Federal agencies and professional associations. Federal agencies spend money on medical training with too much emphasis on increasing numbers of physicians and too little on the problem of how they are distributed. Moreover, the foreign medical graduates are deficient in English and in acclimation to American culture, and many of them are not up to American standards of medical proficiency. Finally, the professional associations have rigidly upheld out-of-date standards that restrict professionals other than physicians from providing psychiatric care.

The final comment appears to be at odds with the suggestion that foreign medical graduates are often poorly qualified to deliver psychiatric care. It reveals, however, what we believe is the reason for the agitation expressed by Torrey and Taylor and for the wider criticisms of the nation's heavy reliance on foreign medical graduates of all types. The reason is the realization that the growing presence of foreign-trained physicians is symptomatic of wider problems in the American system of delivering health care. Associated with this realization is the recognition that the medical profession as it has been structured since the reorganization of medical education following the Flexner report of 1910 (*Medical Education in the United States and Canada*) is under severe pressure to change. Allegations of inequities in the delivery of care, of discriminatory admissions policies by medical schools against women and racial minorities, of a disjuncture between academic medicine and the effective delivery of medical care, of self-interest of physicians and their professional associations and of an overriding failure to plan systematically have all taken their toll. The evidence of these shortcomings is the foreign medical graduate, who has helped to maintain a faltering system by disproportionately filling the roles that American-trained physicians regard as least desirable.

Two international organizations have been studying the migration of health professionals. The World Health Orga-

nization recently began to examine the causes and consequences of the migration and to assess its true magnitude. The Pan American Health Organization has sponsored several conferences on the migration as it relates to Latin America. No definite results or policy recommendations have yet emerged from these activities.

The situation within the U.S. is best described as bizarre. Although the Federal Government gives the appearance of activity, its movements seem contradictory. A year ago Charles C. Edwards, assistant secretary for health in the Department of Health, Education, and Welfare, asserted that $3.5 billion of Federal spending over the past decade had eased the physician shortage and that the U.S. "may well be facing a doctor surplus." On the other hand, the Department of Labor operates as if a physician shortage existed. As the agency charged under the Immigration Act of 1965 with declaring what occupations are understaffed in the U.S., the department has unfailingly kept physicians and other health professionals on the list. As a result they encounter minimal problems in gaining entry to the U.S.

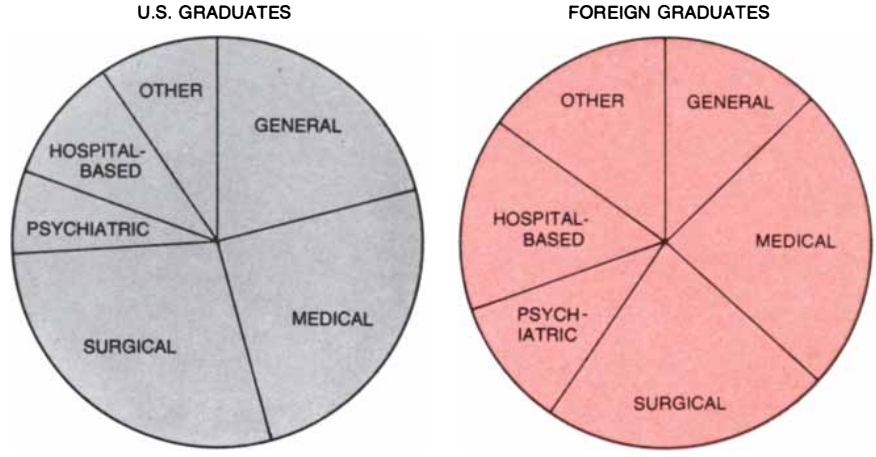


STAFFING OF HOSPITALS has relied increasingly on foreign medical graduates in recent years. Each whole bar on the chart shows the number of interns and residents that hospitals in the U.S. sought to employ in each of the years indicated. Shown within that total are the number of those house-staff positions filled by American medical graduates (*gray*) and by foreign medical graduates (*color*) and also the number of positions unfilled each year.

The major professional organizations have also had difficulty articulating a position on foreign medical graduates. In a report published last August the Association of American Medical Colleges called for a reduction in the number of foreign-trained physicians (including Americans) and for an end to dual testing and evaluation standards, meaning that foreign medical graduates should take the same route (presumably a more difficult one) that American medical graduates take to qualify for internships, residencies and practice. The report also urged maintenance of the present standards of delivering health care, although how that could be done while the number of foreign medical graduates was being drastically reduced was not explained.

The crucial question in this report, as in other reports and in Congressional hearings and journal articles on foreign medical graduates, is the competence of foreign medical graduates and to what degree it differs, if at all, from that of American medical graduates. An examination of this issue reveals certain less obvious concerns behind the stated concern over competence. The examination also brings into sharper focus the role of foreign medical graduates in the broader picture of changes in the American system of health care.

A statistic often cited to question the competence of foreign medical graduates is their record in obtaining state licenses to practice medicine. For example, calculations we made on the basis of a survey by the American Medical Association (designed to ascertain what all the people who had been interns and residents in 1963 were doing in 1971) showed that 93 percent of the American medical graduates and 66 percent of the foreign medical graduates had obtained licenses. A close look at data of this kind, however, suggests that too many factors other than competence are associated with licensure to warrant its use as an unbiased measure of competence. For example, one finds a strong correlation between the type of visa a foreign medical graduate holds and his ability to obtain a license. People who had become



CHOICE OF SPECIALTY is depicted for graduates of American medical schools (*left*) and for foreign medical graduates in the U.S. (*right*). The data are for 1970. Among "medical" specialties are cardiology, dermatology, internal medicine, pediatrics and work with allergies. Among "hospital-based" specialties are anesthesiology, radiology and pathology.

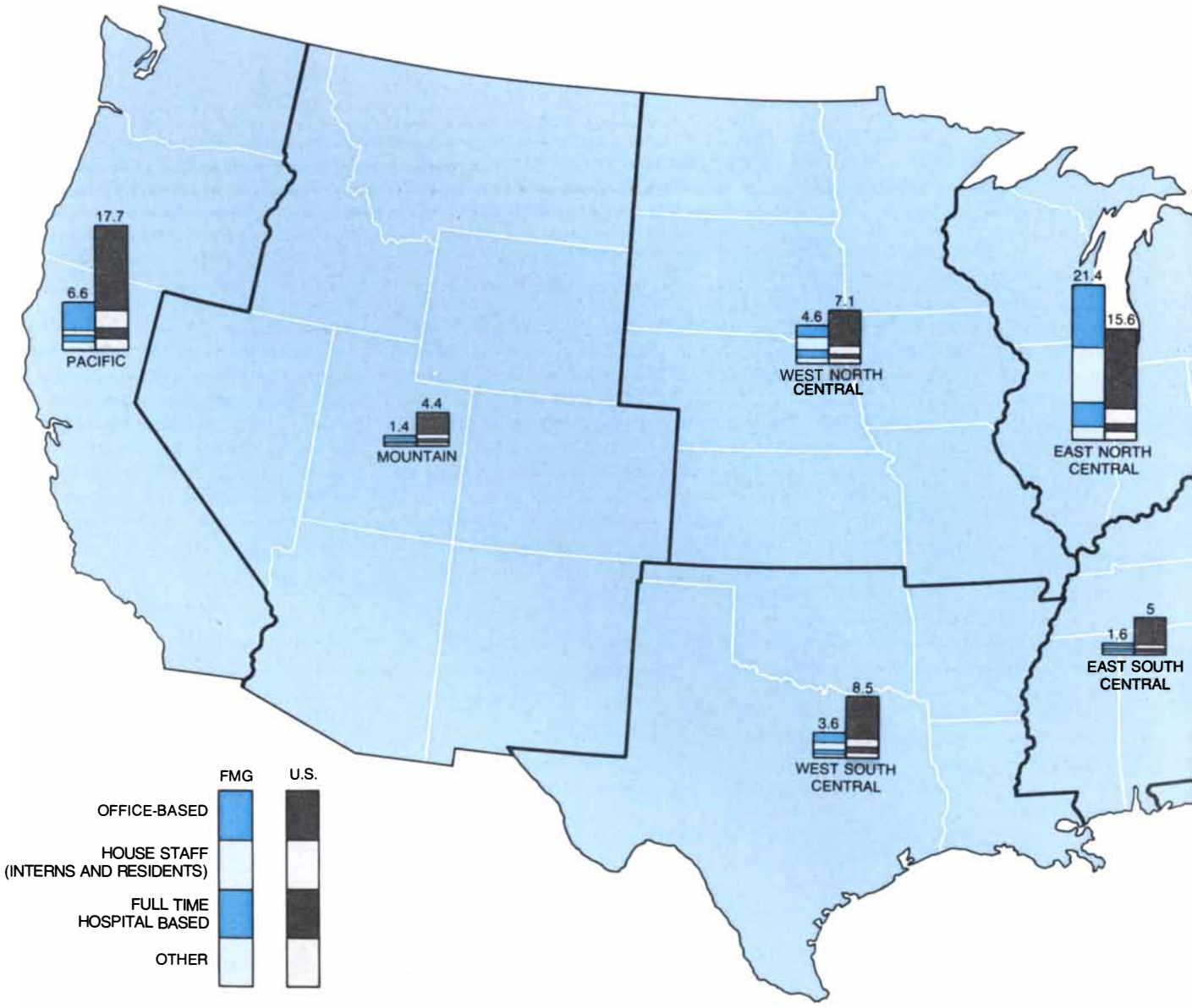
U.S. citizens through the naturalization process were licensed at the same rate as American medical graduates, immigrant physicians at a slightly lower rate and exchange visitors at the lowest rate. There is no a priori reason related to competence why these differentials should exist.

It is instructive to see how native Americans who took their medical training abroad fared in licensure. Examining the 10 countries that educate the greatest number of Americans, one finds that (with two exceptions) the graduate's ability to obtain a license in the U.S. is higher for the Americans than it is for native students being graduated from the same school. Here the medical-school background is equivalent for each type of applicant.

One can turn the picture around and examine the few foreign-born physicians who have been trained in American medical schools. Again one finds that the pattern of licensure varies according to the type of visa held by the applicant. Evidently administrative factors have as much bearing on the ability of a foreign medical graduate to obtain a license as the quality of his medical degree does.

What of the statistics showing that foreign medical graduates score more poorly on the examination given by the Educational Council for Foreign Medical Graduates than American medical students could be expected to score, based on their performance on the examination given by the National Board of Medical Examiners? It is noteworthy that Americans who study medicine abroad are even less successful than foreign medical graduates in passing the council's examination but are as successful as American medical graduates in obtaining licenses.

In addition to their record on the council's examination foreign medical graduates do not fare as well as American medical graduates on the examinations administered by the various medical boards that certify physicians for such specialized types of practice as orthopedic surgery and ophthalmology. A conclusion usually drawn from such statistics is that foreign medical schools are by and large of lower quality than Amer-



DISTRIBUTION OF PHYSICIANS in the U.S. is portrayed according to whether they were educated in U.S. (*gray*) or foreign medical schools (*color*) and according to type of practice. The figure at the top of each bar is the percentage of each type of graduate

ican schools. Another conclusion is that one cannot cite language as a barrier for foreign medical graduates, since Americans who studied abroad fare even more poorly on the council's examination than foreign medical students.

It would take further study, however, to determine whether or not these conclusions are justified. Perhaps Americans who go abroad to study medicine represent a significantly less qualified pool of physicians than foreign medical graduates. An even more plausible supposition is that, although foreign medical graduates have language problems that impair their performance on an examination such as the one the council gives, Americans who study abroad suffer from language difficulties while they are learning medicine, resulting in deficiencies that may never be overcome and that manifest themselves in examination scores. As for the poorer record of foreign medical graduates in obtaining certification from specialty boards, no one has disproved the hypothesis that a contributing factor is the large number of foreign medical graduates who receive graduate training in American hospitals that are not affiliated with teaching centers.

I am not arguing here that all the data on foreign medical graduates are false or that no real differences exist between the foreign graduates and American medical graduates. My point is that quality of care and being a foreign-trained physician may be correlated only under certain specific conditions. In fact, I am suggesting that no correlation can be demonstrated because no uniform procedures for measuring quality of care have yet been proposed, let alone applied to foreign medical graduates.

Analysis of our group's National House Staff Survey shows that foreign medical graduates occupying house-staff positions have had significantly more clinical practice (before arriving in the U.S.) than their American counterparts, most of whom go directly from medical school to a house staff. It may be that in terms of experience the foreign medical graduate is able to deliver better care to the patient than the American graduate, although language, cultural differences and hospital experiences may detract from the foreign graduate's performance. Indeed, if there is a "problem" involving foreign medical graduates, the problem may be that a valuable manpower resource is being underutilized in the nation's health-care system.

If the issue of competence is as clouded as I have indicated, what other factors might account for the current climate of concern about foreign medical graduates? The roots, I believe, are to be found in economic and political considerations. The Nixon and Ford administrations have made clear that an accounting of the large Federal investment in medical education is expected. Various officials of the Department of Health, Education, and Welfare have strongly implied that there will be cuts in spending.

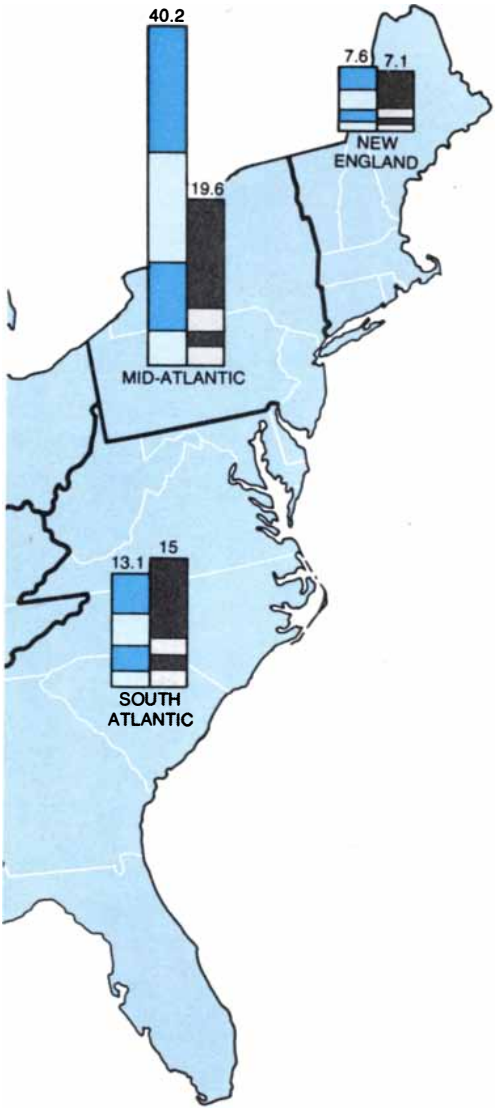
The continuing presence of foreign medical graduates may be a powerful weapon for the Government against arguments that Federal funding should not be cut and should even be increased. When the argument is raised, the Government can counter that the corps of foreign medical graduates reduces the need to educate more American physicians. The countervailing argument, which is advanced mostly by professional medical organizations, fits neatly with the assertions that the foreign graduates may be delivering substandard care. By calling into question the quality of the foreign graduates, the arguer can contend that more money rather than less is needed to educate more American physicians to take over the places now occupied by foreigners.

From the point of view of the foreign medical graduates it must be difficult to understand some of the attitudes they encounter. The foreign graduates have been welcomed to the U.S. through lenient regulations on visas. In some cases they were actually recruited by representatives of American hospitals. Yet once here they are attacked as being less competent than American physicians.

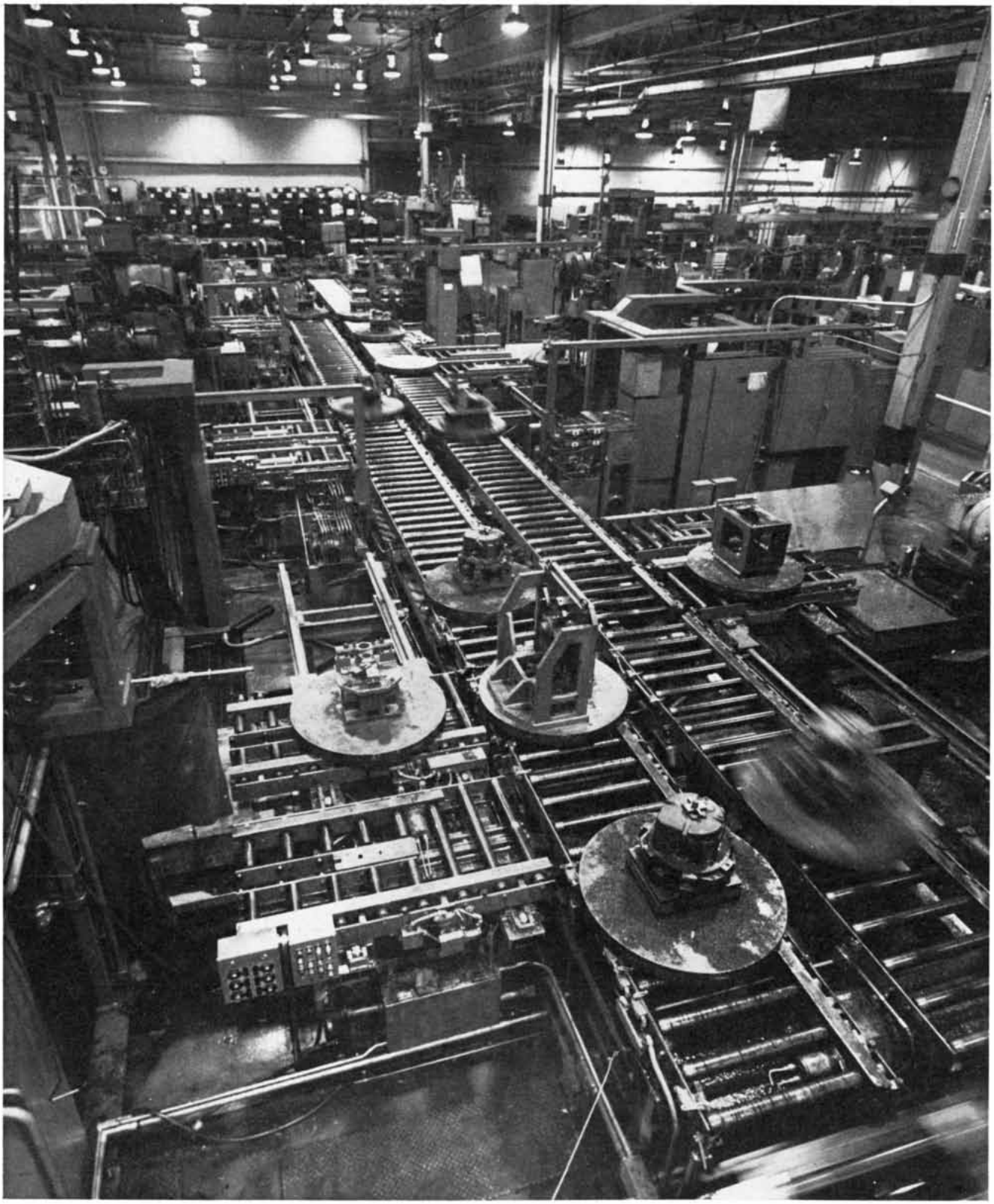
In this situation the foreign medical graduate has become a pawn among warring factions in a medical profession that is caught up in ambivalence over what to do about changing the system of health-care delivery in the U.S. Lacking a coherent policy regarding the numerous issues in medicine, the medical profession in the U.S. has allowed the foreign medical graduate to fill in the most obvious cracks in the system. Once here, however, the foreign graduates have their own interests to protect.

Meanwhile the problems of the countries from which the foreign medical graduates came remain unsolved. Indeed, it can be argued that these problems are being intensified as a result of a decline in the amount of foreign aid provided by the U.S. Foreign medical graduates understandably express reluctance to return to their home countries unless those countries develop better systems of health care. Improvement is unlikely to be rapid in any case, but with lower levels of foreign aid from the U.S. it may be even longer in coming.

The answers will not be found easily. Because of the large number of licensed foreign-educated physicians who have become naturalized citizens of the U.S., it is easy to predict that foreign medical graduates will continue to play an important role in the nation's health-care system. It is more difficult to predict whether this dependence on foreign medical graduates will continue to increase or whether it will decrease under the pressure of the complaints raised about it. It may not be possible for other nations to solve their problems regarding medical manpower until the U.S. has solved its own.



in each geographical region. Distribution by type of practice is indicated within the bars.



HEAVY MACHINING CENTER in the Roanoke, Va., plant of the Ingersoll-Rand Company was probably the first major example in the U.S. of computer-managed parts manufacturing (CMPM). The center, which was designed and built by the Sundstrand Corporation, went into operation in 1972. Arranged around the system that transfers pieces from one machine tool to the next (*see flow diagram on opposite page*) are six numerically controlled tools: two five-axis milling machines, two four-axis milling machines and two four-axis drills. The six machines and the transfer system are all under the control of an IBM 360/30 computer. The three-second exposure needed for this picture has blurred the piece just starting down the transfer line. The center turns out a wide variety of parts used in Ingersoll-Rand products, chiefly parts for hoists and winches. The circular pallets hold parts whose maximum dimensions fit within a 36-inch cube. Machining operations include milling, turning, boring, tapping and drilling. At any one time some 200 tools are in automatic tool-changing carousels, available for selection by computer. On each 12-hour shift the center is run by three operators and a supervisor. To achieve the same output a conventional shop would need perhaps 30 machines and 30 operators.

COMPUTER-MANAGED PARTS MANUFACTURE

Although a completely automatic factory still lies in the future, important advances have been made in fully automatic systems for the batch manufacture of a variety of complex components

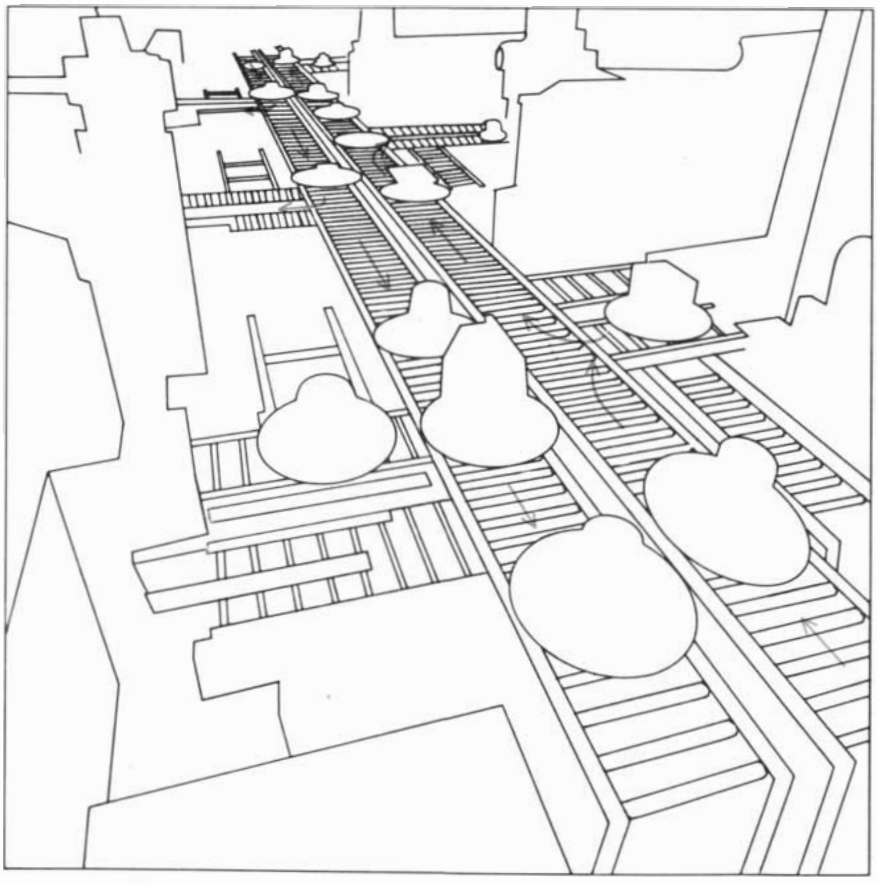
by Nathan H. Cook

Evolutionary developments in manufacturing technology are opening the door to a new kind of factory, a factory where, in comparison to present-day factories, the working environment is greatly improved, where labor requirements are vastly reduced, where raw materials are more effectively utilized and where, as a result, goods can be manufactured at a substantially lower cost. The developments responsible for these advances involve an intensified application of computers to the control and management of manufacturing machines and systems. Since the developments are recent, the terms are not generally familiar, and I shall introduce them as we proceed.

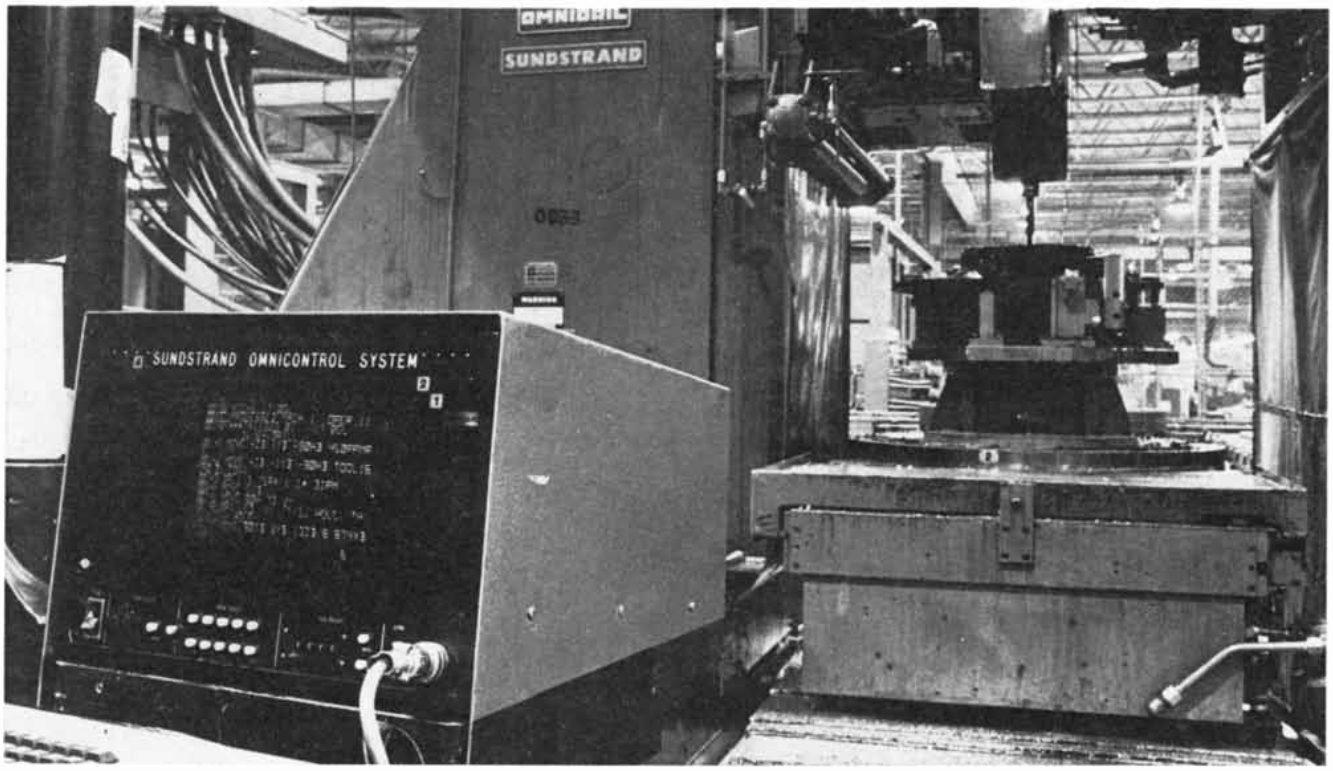
At the highest, most advanced level, there are (or eventually will be) computer-integrated manufacturing systems (CIMS). Such systems call for the coordinated participation of computers in all phases of a manufacturing enterprise: the design of the product, the planning of its manufacture, the automatic production of parts, automatic assembly, automatic testing and of course the computer-controlled flow of materials and parts through the plant. To the best of my knowledge there are no fully developed systems of this kind anywhere in the world. Various subsets, or components, of such systems are nonetheless emerging. Here I shall describe one subset that my colleagues at the Massachusetts Institute of Technology and I call computer-managed parts manufacture (CMPM).

In order to demonstrate the potential of computer-managed parts manufacture I shall begin by briefly describing the present-day manufacturing scene. For the sake of simplicity let us consider only the manufacture of metal parts such as we find in automobiles, dishwashers, refrigerators, can openers, pencil sharpeners and so on. Americans are accustomed to thinking of their factories as being highly productive; the efficiency of mass production is part of our national heritage. We hear about Henry Ford in grade school, and we are told that our productivity, in terms of output of goods per man-hour of input, is the highest in the world. We tend to think complacently that our productivity is probably about as high as it can be. Such is not the case.

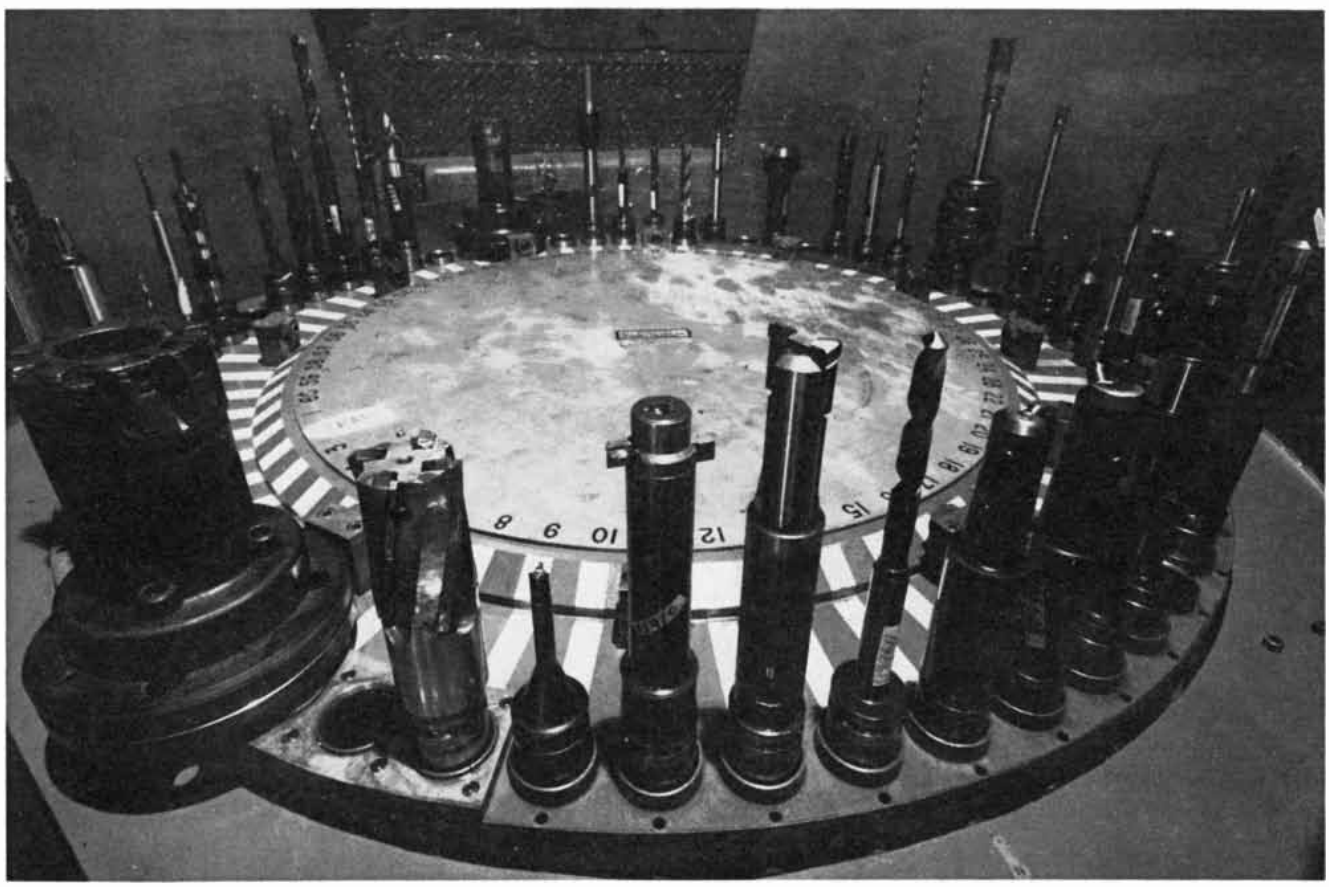
For those items that are mass-produced, such as automobile engines, our methods are indeed so highly productive that only marginal improvements can be expected in the foreseeable future. For items that are not mass-produced, how-



FLOW OF WORKPIECES in the Ingersoll-Rand Heavy Machining Center follows the path shown here. At any one time the computer can direct the machining of as many as 16 different kinds of parts. Depending on the complexity of the part, it may have to be directed to only two or three of the six machines, bypassing the others. A typical part has to be mounted on the pallet no more than twice in order to present all its facets to tools. In a conventional shop a part might have to be mounted in jigs five to 15 different times.



CATHODE RAY DISPLAY SCREEN next to each of the machine tools in the Ingersoll-Rand Heavy Machining Center describes the piece that is being worked on and how far the machine has progressed in carrying out its computer-assigned sequence of operations. This particular screen is associated with a four-axis drill. Machining instructions are entered in the computer by punched card.



AUTOMATIC TOOL-CHANGING CAROUSEL provided with each Sundstrand machine in the Ingersoll-Rand plant can hold as many as 60 different tools. The tools are selected in sequence by computer program and are affixed to the machine automatically in a matter of seconds. Tools are perpendicular to the carousel; the wide-angle lens needed for photograph makes them appear tipped.

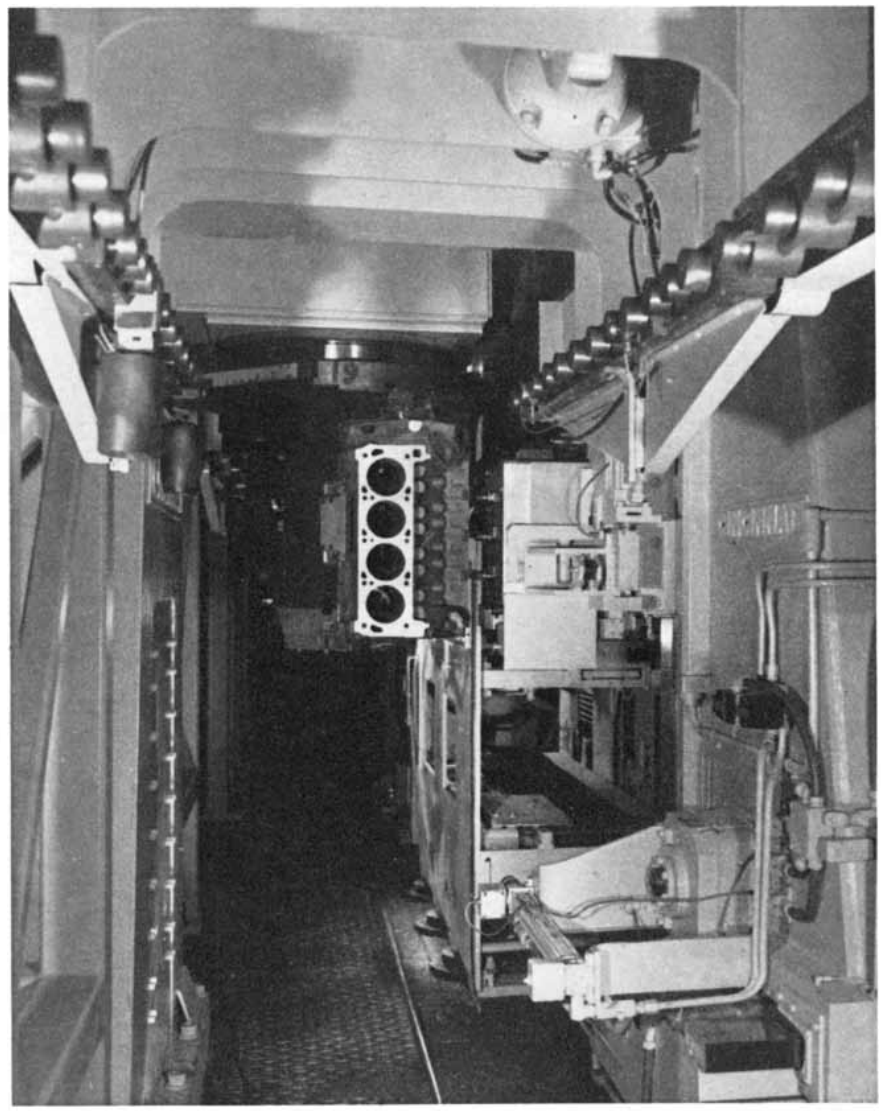
ever, such as airplane engines or the heavy and costly machines of an electric generating station, productivity is disappointingly low, and the potential for dramatic improvement is great.

Mass production, as I am using the term, calls for a special set of production machines that are usually designed to make repetitively only one part in the most efficient manner possible. If the cost of installing the special set of machines dedicated to the manufacture of that one part is to be recovered, the annual production of the part clearly must be large enough to keep the machines running almost continuously. For many manufactured items there simply is not sufficient demand to warrant the installation of special machines; hence one must call on general-purpose machines, machines that are capable of being adapted to making a wide variety of parts. Most of us are somewhat familiar with general-purpose machine tools such as lathes, drill presses, milling machines and the like. With suitable tooling and skilled workmen a great variety of parts can be made on a set of such machines. The cost of the parts, however, is high compared with the cost of parts made by mass-production.

A given part can be made in any one of several different ways, ranging from manufacture by general-purpose machines at one extreme to manufacture by mass-production methods at the other. The logical selection of a set of machines depends heavily on the number of parts to be made, ranging from a single unit to millions of units per year. The cost of machining normally varies with production quantity [*see illustration on next page*]. To machine a single unit with general-purpose tools may cost 100 times as much as to manufacture the same part by the most efficient mass-production methods. As an example, consider a complex mass-produced part with which almost everyone is familiar: the cylinder block for a typical V-8 automobile engine. Under mass-production conditions, where the engine block is conveyed automatically along a transfer line, with the various operations (drilling, tapping, boring, milling and so on) being executed in sequence at the different stations along the line, the complete machining cost (excluding the raw material but including labor, tooling, machine depreciation and interest on investment) would be of the order of $25. If, however, only a few special cylinder blocks were to be made with general-purpose machines and skilled labor, the machining cost per block could easily rise from $25 to $2,500 or more. At intermediate production rates, where parts are made in "lots," or batches, it is still necessary to use general-purpose machines (so that they can also be used to make other parts), but one can economically justify some degree of automation, with a corresponding reduction in the machining cost per part.

We can now examine the problem more closely and ask: Where are parts-manufacturing dollars spent? Although industry mass-produces parts in great variety and number, their individual cost is low. On the other hand, although "one of a kind" parts are extremely costly, there are not many of them. Between these extremes lie parts of many kinds that are batch-produced; both the costs and the quantities are moderately high. It has been estimated that between 50 and 75 percent of the total national outlay for parts manufacturing is accounted for by batch-production methods where the individual batch size is 50 or fewer. Parts made in this fashion may cost anywhere from 10 to 30 times more than they would if they were mass-produced. Thus there is considerable economic leverage: if some of the efficiencies of automated mass-production could be introduced into the batch-production area, the potential for savings would be great.

It is not surprising that the workers involved in the batch manufacturing of metal parts constitute about 40 percent of the total manufacturing labor force. It happens that it is becoming increasingly difficult to attract apprentices to



VARIABLE-MISSION MANUFACTURING SYSTEM, being developed by Cincinnati Milacron Inc., uses an unconventional overhead transfer system to move workpieces. Here a cylinder block has been conveyed to a numerically controlled, five-axis "manufacturing center" that automatically changes toolheads, some of which carry multiple drills, taps or other tools to perform simultaneous operations. The toolheads are stored in a carousel.

the metalworking trades. Young people seem to prefer other types of jobs, particularly jobs in the service sector of the economy. This is a worldwide development; many European countries are already forced to import large numbers of semiskilled and unskilled industrial workers. This, then, is the area on which I shall focus: that segment of the manufacturing industry now devoted to the batch manufacture of metal parts, which constitutes a substantial portion of the gross national product (perhaps 10 percent), which employs 40 percent of the manufacturing labor force in a time of decreasing labor availability and which is characterized by unit costs from 10 to 30 times higher than those in the mass-production sector of the economy.

How exactly does the introduction of computer control and computer management reduce labor and production costs? An answer can best be developed by considering how general-purpose machine tools have evolved over the past two decades through the introduction of computer technology.

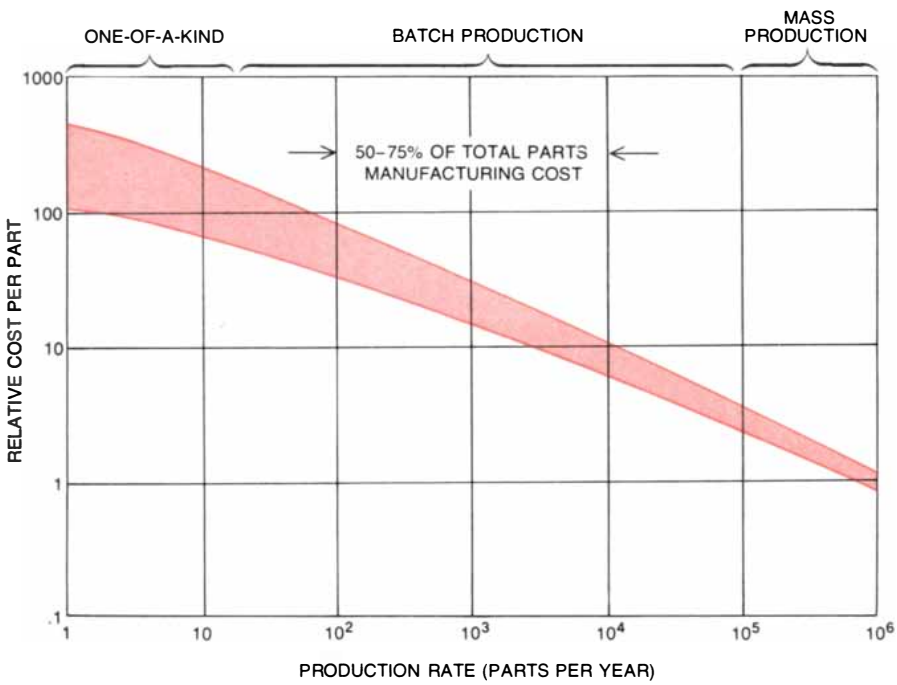
In the operation of a general-purpose machine there are a certain number of functions that must be performed either manually or automatically: (1) Move the proper workpiece to the machine; (2) load the workpiece onto the machine and affix it rigidly and accurately; (3) select the proper tool and insert it into the machine; (4) establish and set machine operating speeds and other conditions; (5) control machine motion, enabling the tool to execute the desired function; (6) sequence different tools, conditions and motions until all operations possible on that machine are complete; (7) unload the part from the machine. In the operation of the traditional general-purpose machine (the lathe, the milling machine, the drill press and so on) all seven functions are performed by the operator [*see illustration at upper left on opposite page*].

Two decades ago the first numerically controlled (N/C) machine tool was developed at M.I.T. In essence function No. 5, the control of machine motion, was taken from the operator and made automatic by means of information stored on a punched tape [*see illustration at upper right on opposite page*]. Numerically controlled machines now exist in various degrees of complexity, depending on how many axes (motions) are controlled and on whether only the final position is controlled (point-to-point operation) or a contour is prescribed. At the simple end of the scale is a "two-axis point-to-point drilling machine" in which the drill table can be automatically positioned only in the horizontal (x-y) plane. At the complex end of the scale would be a "five-axis milling machine" in which three linear motions (x-y-z) and two angular rotations are continuously and synchronously controlled to produce sculptured parts with complex contours. Such machines are widely used in the airframe industry.

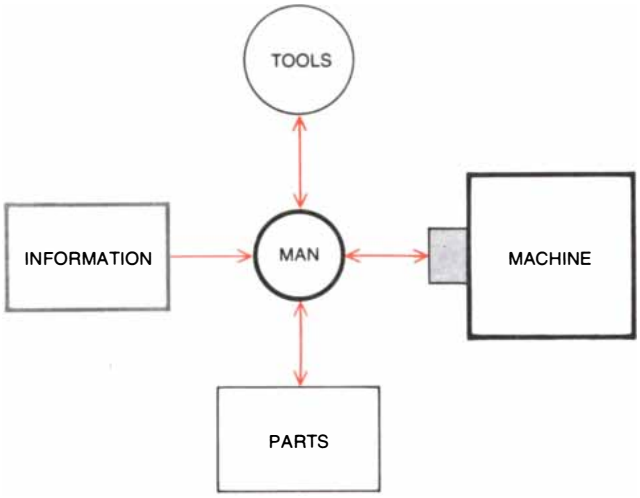
The development of numerically controlled machines constituted a major technical advance because in general the other functions call for far less operator skill and training. As more and more numerically controlled machines were installed, an entirely new, highly skilled job came into being: the job of the parts programmer, whose function was to translate the engineering drawing of a part into the punched tape that would cause the machine to generate the part. Over the years there have been numerous developments in parts-programming methods and in special computer-programming languages; they make it fairly easy, with computer assistance, to produce the required tapes. It is expected that current development work will lead to a much higher level of automation in parts programming over the next decade.

The next major development was the introduction of automatic tool-changing (ATC) systems, which relieved the operator of function No. 3 [*see illustration at lower left on opposite page*]. This kind of system endows an automatic machine tool, such as a three-axis numerically controlled milling machine, with the capability of storing, selecting and changing cutting tools, all under the control of a punched tape. These machines, which are often called machining centers, have been developed to the point where a single machine can incorporate a bank of as many as 60 tools that can be changed in a few seconds.

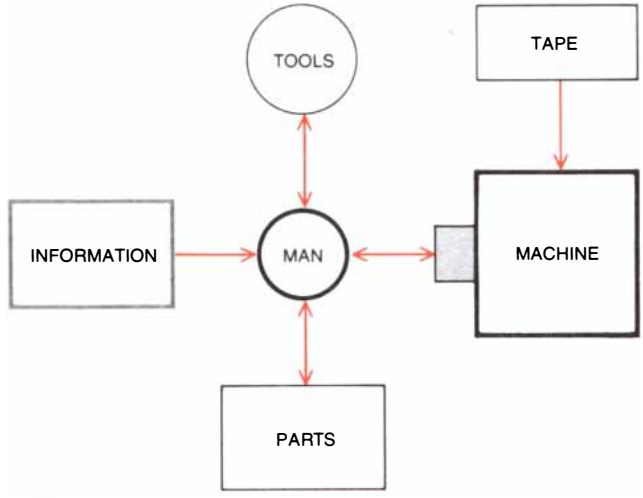
The next step in the development of numerical machine control was the replacement of punched-tape information storage by magnetic-disk storage or computer-memory storage, yielding a computer-numerical-control (CNC) system [*see illustration at lower right on opposite page*]. This change made it far easier to edit and alter the parts program; in addition it made it possible for the computer to take over many auxiliary functions, such as preselecting the next tool required and having it ready when it is needed. When a battery of machines is placed under the control of a single computer, the result is a system known as direct numerical control (DNC). This system obviously enhances the utilization of computers.



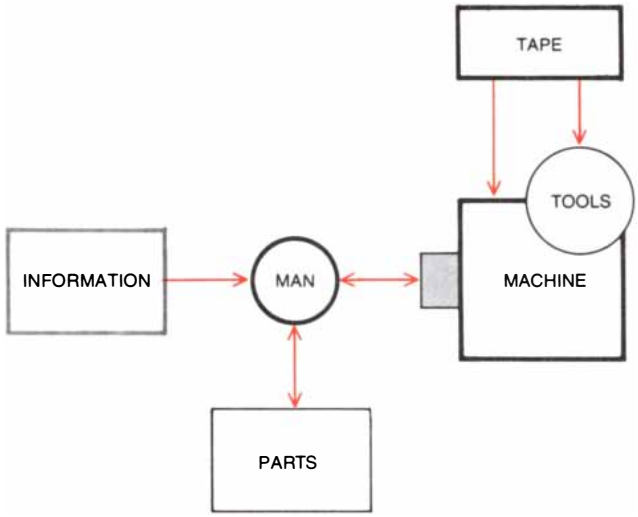
COST OF MACHINING typical industrial parts declines by a factor of perhaps 100 in going from "one of a kind" production to mass production. The term "batch production" applies to parts manufactured in lots ranging from several units to more than 50, for which the total annual demand is fewer than, say, 100,000 units. When the demand exceeds that volume, depending on the specific product, use of special-purpose machines can generally be justified. The author estimates that batch-production methods account for from 50 to 75 percent of national expenditure for manufactured parts. By making greater use of computer technology it should be possible to reduce cost of producing parts made in small quantities.



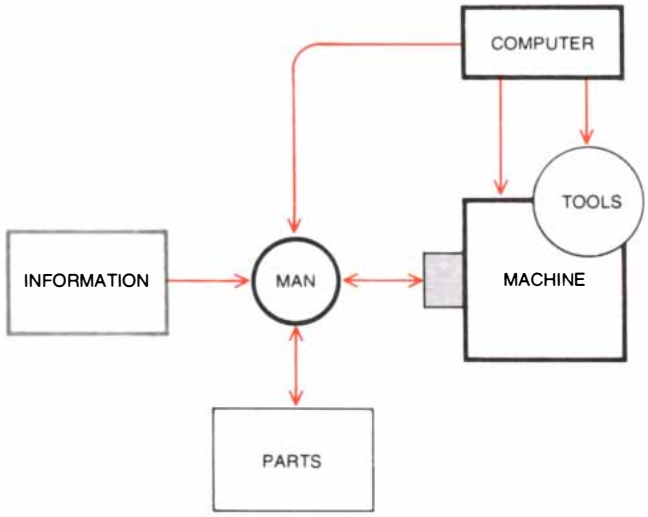
GENERAL-PURPOSE MACHINE, such as a lathe, a milling machine or a drill press, requires the services of a skilled operator. His job is to load the machine, select the proper tools and provide the machine settings needed to accomplish a given task. His instructions are usually contained in a blueprint of the desired part.



NUMERICALLY CONTROLLED MACHINE TOOL developed at the Massachusetts Institute of Technology relieved the operator of the job of interpreting a blueprint in selecting machine settings. Machine motion is controlled by instructions on a punched tape. Operator must still select tools and load and unload the machine.



AUTOMATIC TOOL CHANGING was the next step in raising the productivity of general-purpose machine tools. In this system the punched tape contains not only machine-guidance instructions but also information for selecting the right tool from a bank of from 20 to 100 tools. Tool-changing time can be as little as two seconds.



COMPUTER NUMERICAL CONTROL transferred information storage from punched tape to the more capacious and flexible memory of a computer. This change not only made editing and altering programs easier but also made a computer available for a variety of other tasks, such as logging the time each tool is in use.

We can now consider the characteristics and capabilities of a computer-managed parts-manufacturing system. In a CMPM system a number of machines are linked together, not only through a common control computer but also through a part-transfer, load/unload system [*see illustration on page 29*]. Apart from tool and machine maintenance the only direct operator activity is attaching the workpieces to special pallets (load) and removing the finished parts (unload). Once a workpiece has been loaded and the computer has been informed that it is ready to enter the system, the computer takes over, routes the workpiece to the necessary machines, selects the proper tools, executes the proper operations and when the part is finished returns it to the operator for unloading. At the machine itself all seven of the required functions have been automated, with the result that the operator services a system of machines rather than a single machine.

Such a system is clearly well suited to the segment of industry we are considering: the manufacture of parts in small batches. Parts are produced in batches because of the "setup" costs associated with changing a machine from making part *A* to making part *B*. Fixtures that allow part *A* to be clamped accurately in the machine must be replaced by fixtures suitable for part *B;* tools for part *B* must be drawn from the "toolroom" and inserted in place of the tools used for part *A;* finally, parts must be run and carefully checked to ensure that all is in order. Obviously if only a small batch of parts is made, the setup costs per part are high. If a large batch is made, the setup costs are low but the subsequent inventory costs are high. The result is an optimum batch size for any particular part. In a computer-managed parts-manufacturing system, however, the usual concept of an economic batch size does

not apply because the setup costs are exceedingly low. It can be efficient to run a "batch" consisting of only a single part. Since the batches are small, the cost of holding in inventory parts that are either finished or in process becomes almost negligible.

I hardly need to emphasize that computer-managed parts-manufacturing systems call for considerably less direct labor than more conventional manufacturing systems. The data, which are still limited, indicate that a CMPM system, compared with a standard job shop of similar capacity, needs only between 10 and 30 percent as much direct labor.

It is also true, of course, that CMPM machines are more complex and costly than conventional machines. Does the labor saved offset the increased cost? In undertaking to answer this question one must bear in mind that a metal-cutting machine is productive only when it is actively cutting metal. It is not being productive when it is shut down, when it is being set up to make a different part, when parts are being loaded and unloaded or when tools are being changed. Although estimates vary, the average machine tool in a conventional shop is cutting metal only between 3 and 10 percent of the time, whereas in a CMPM system the metal-cutting time can be 50 percent or more. It is this far greater utilization that offsets the higher individual-machine cost. It is anticipated that for a given productive capacity the capital investment required for a highly sophisticated CMPM system will be no more than that for a more conventional shop.

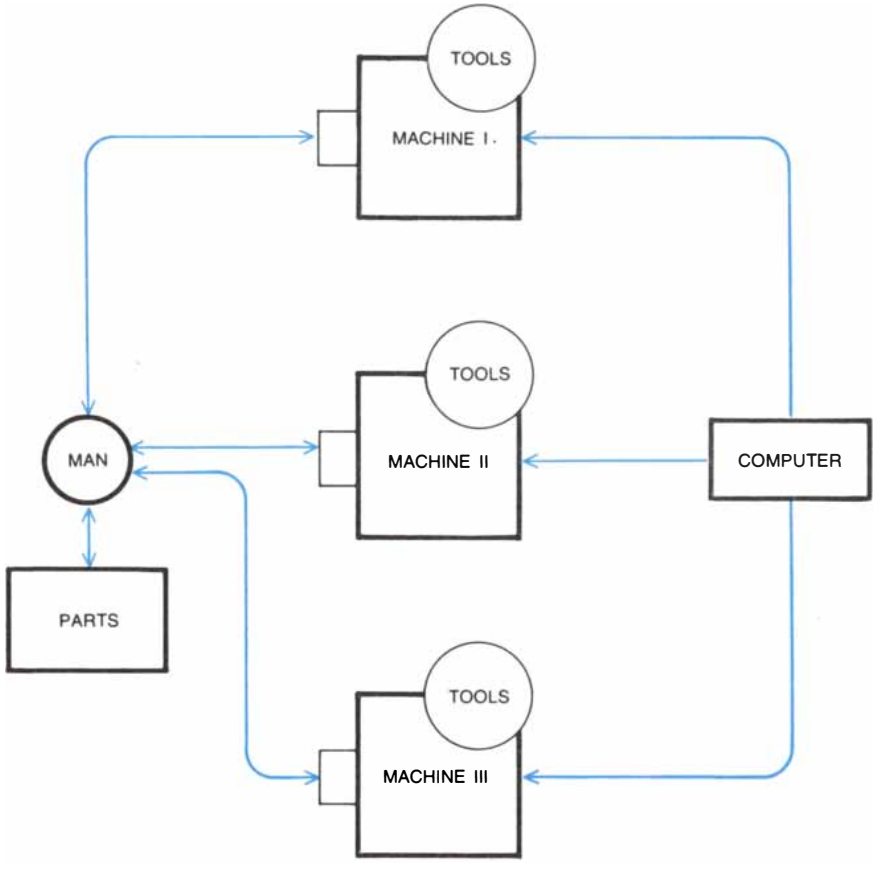
In addition to the direct savings in labor costs one can expect substantial savings from reduced inventories. Far fewer parts, whether finished or in process, will be "waiting," either for the next operation or for assembly. The anticipated result is an overall cost reduction by a factor of from five to 10; in other words, a cost reduction of from 80 to 90 percent.

Although CMPM systems are only now emerging, a small number of prototype systems have been set up for production or demonstration. Perhaps the most ambitious system (and the one for which information is hardest to obtain) is operating at Karl-Marx-Stadt in East Germany. The system, enclosed in a special air-conditioned building as large as two football fields, employs linear induction motors to propel the work pallets, which are floated on air-cushion supports. The system can handle workpieces as large as $1 \times 1 \times 1.6$ meters.

In the U.S. the CMPM system that probably has been in operation longest is the Heavy Machining Center at the Ingersoll-Rand plant in Roanoke, Va. Built by the Sundstrand Corporation, the installation has six machines arranged around a looped transfer system [*see illustrations on pages 22 through 24*]. The system is capable of manufacturing some 500 completely different parts. At any one time it can accommodate as many as 16 parts of different design, either being machined, waiting in queue to be machined or on the transfer system. In the present configuration of the system there are about 500 tools, 200 of which are in automatic tool-changing carousels at any one time. The entire system, which can be regarded as replacing a typical shop of 30 machines and 30 workers, is operated by a team of only three people and a supervisor.

Some of the most striking advances in automatic production on the batch scale are being incorporated in the variable-mission manufacturing system now being developed by Cincinnati Milacron Inc. [*see illustration on page 25*]. This system not only has the general character of computer-managed parts manufacture seen in other systems but also provides for the processing of low-volume parts at higher rates than the rates that can be achieved with more conventional numerically controlled machines. For example, an ingenious five-axis "manufacturing center" automatically changes clusters of tools mounted on a single head, so that a number of operations can be performed simultaneously. By means of a novel scheme of handling workpieces from above, the Cincinnati Milacron system provides efficient management of coolants and chips, together with easy access for inspection and servicing.

The country that has the greatest experience with computer-managed parts manufacture may be Japan, where a number of operating systems have been developed. Because of Japan's (prerecession) labor shortage and the country's well-known need to maintain economic advantages in production, the Japanese are supporting a large research and development program directed toward an "unmanned factory." (Between 1965 and 1973 Japanese industrial productivity climbed at the remarkable rate of more than 13 percent per year, more



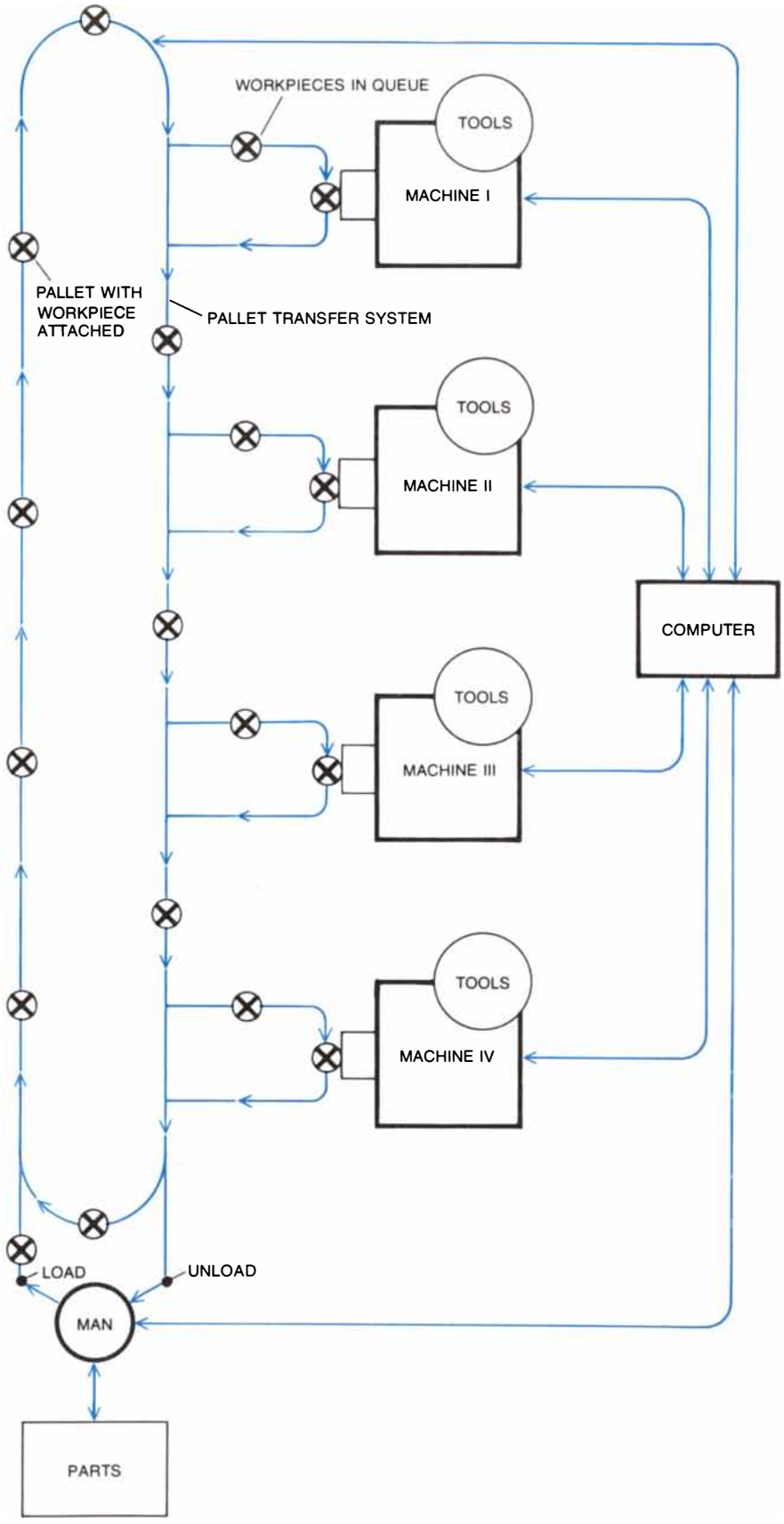
DIRECT NUMERICAL CONTROL places a battery of machines under the control of a single computer. Machines can be of different types and programmed to carry out different tasks. Such systems may or may not be provided with automatic tool-changing capabilities.

than four times the rate of increase in the U.S.)

So much for the description of computer-managed parts manufacture and its present state of development. What about the future? At this point I must shift from presenting facts to offering personal opinion. If cost reductions of perhaps 80 percent can be predicted for CMPM systems, why are so few systems in operation? There are two kinds of reasons: economic and technological.

The principal economic reason is that we are not yet sure just how much can be saved through computer-managed parts manufacture in various branches of industry. As we have seen, it has been estimated that from 50 to 75 percent of all parts manufacturing is done in batches of 50 or less and might therefore be suited to computer-managed parts manufacture. We do not know, however, precisely how large the economic lever is. The existing systems have been justified for the manufacture of specific sets of parts, and the justification has included a generous amount of faith. CMPM systems are expensive; they can cost millions of dollars. Unless we can help to show through research and demonstration that CMPM systems are indeed economic, reliable and amenable to future changes in product mix and quantity, few companies will be willing to take the financial risk involved in finding out for themselves.

The technological reasons for the slow adoption of CMPM systems are mainly associated with the development of computer systems and programs to run them. Although individual companies have created highly successful operating systems, none, to my knowledge, has the full operational flexibility that is needed to secure the maximum performance from a CMPM system. From the point of view of the U.S. as a whole it seems wasteful for each machine-tool manufacturer entering the CMPM market to completely evolve its own proprietary system. It would seem more efficient to develop a commonly accepted overall system framework that would ensure the compatibility of the hardware and the software developed by various manufacturers. It is my own belief that if the estimates of CMPM cost effectiveness are reasonably accurate and if the U.S. hopes to maintain levels of productivity higher than those of other countries, it will have to mount a national program for the development of computer-managed parts manufacture. Such a program should have the ultimate goal of achieving totally computer-integrated manufacturing systems.



COMPUTER-MANAGED PARTS MANUFACTURING represents the latest stage in the application of computer technology for reducing the cost of machining parts in small batches. Since the computer memory contains all the information needed to machine a part, a "batch" can efficiently consist of only one unit. The system depicted here is essentially the same as that employed in Ingersoll-Rand's Heavy Machining Center in Roanoke, Va.

The Most Primitive Objects in the Solar System

Minerals found in the meteorites known as carbonaceous chondrites represent samples of the solid grains that condensed directly out of the gaseous nebula that gave birth to the sun and the planets

by Lawrence Grossman

Of the various samples we now have of the solar system the most primitive are the meteorites known as chondrites. Many of these objects consist of materials that were formed 4.6 billion years ago, well before the most ancient rocks on the surface of the earth and the moon had crystallized. The chondrites appear to have been preserved in smaller bodies that escaped the violent processes that destroyed the earliest rocks on the earth and the moon and created new ones from the remnants. Furthermore, there is considerable evidence that one class of chondrites, the carbonaceous chondrites, are mixtures of minerals that have survived unaltered since the time they condensed out of the solar nebula: the cloud of gas and dust that gave rise to the sun and the planets.

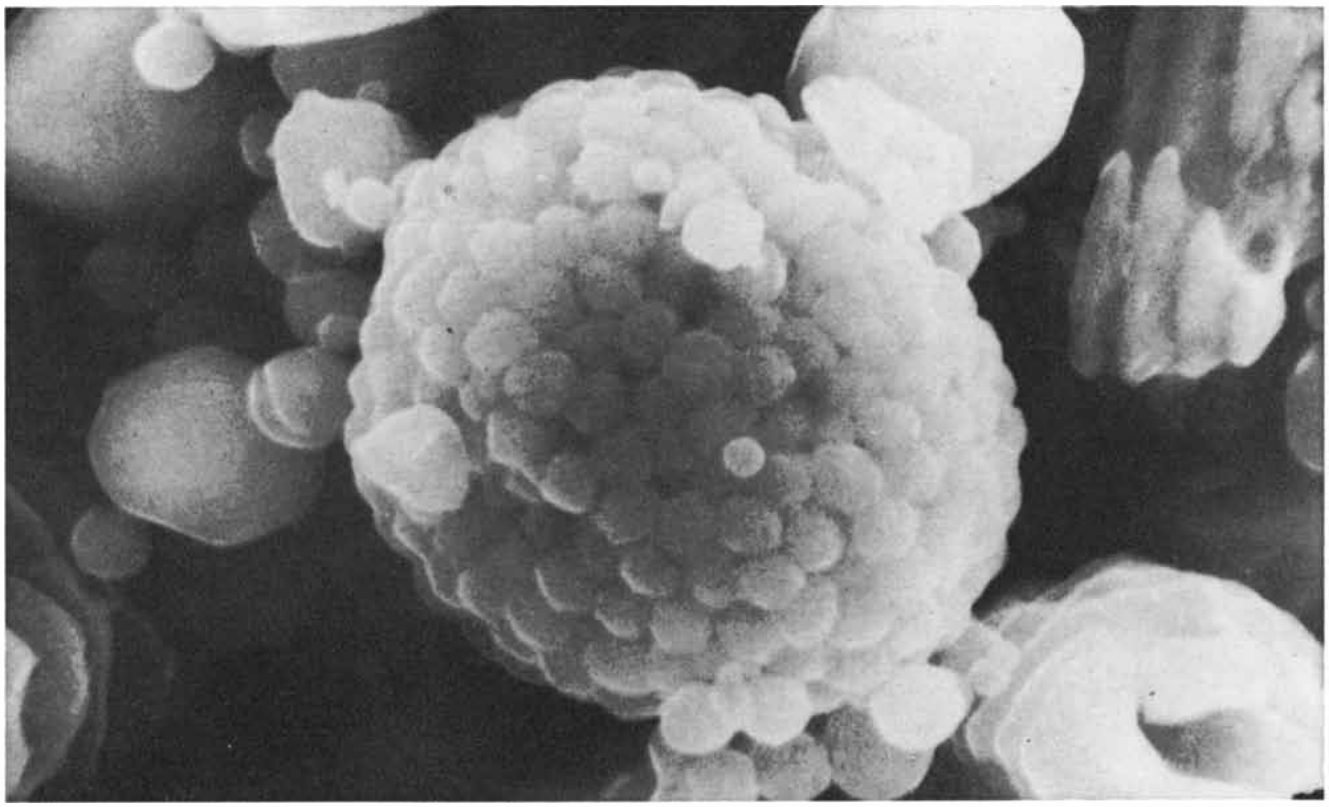
Chondrites are stony meteorites that usually contain chondrules, spheres of silicate minerals and glass ranging up to the size of a pea. Several kinds of chondrites are known, but here we shall focus on the carbonaceous chondrites, so named because they contain an appreciable amount of carbon. The carbonaceous chondrites of Type 1, also designated C1 chondrites, have a special significance. First, they contain no chondrules. Second, of all the various types of meteorites they have the highest concentration of water and volatile elements such as lead, bismuth and carbon. In fact, for all but the most volatile elements (hydrogen, oxygen, carbon and the noble gases) the relative abundances of the elements in the C1 chondrites are the same as the abundances in the sun. Since the sun accounts for more than 99 percent of the mass of the solar system, this means that for all practical purposes the relative proportions of the nonvolatile elements in the C1 chondrites are characteristic of the proportions in the solar system as a whole. Evidently the nonvolatile elements did not separate from one another as the material of the C1 chondrites formed. By the same token, these materials could not have undergone the evolutionary processes that fractionated the elements during the formation of the rocks of the earth and the moon. Such arguments suggest that the minerals observed in the most primitive chondrites may be the solid grains that condensed out of the solar nebula at the birth of the solar system.

The carbonaceous chondrites of Type 2 and Type 3, designated C2 and C3, are also very primitive meteorites. They contain chondrules and have a different chemical and mineralogical composition from C1 chondrites and from each other. These characteristics are clues that the different types of carbonaceous chondrites formed under slightly different conditions.

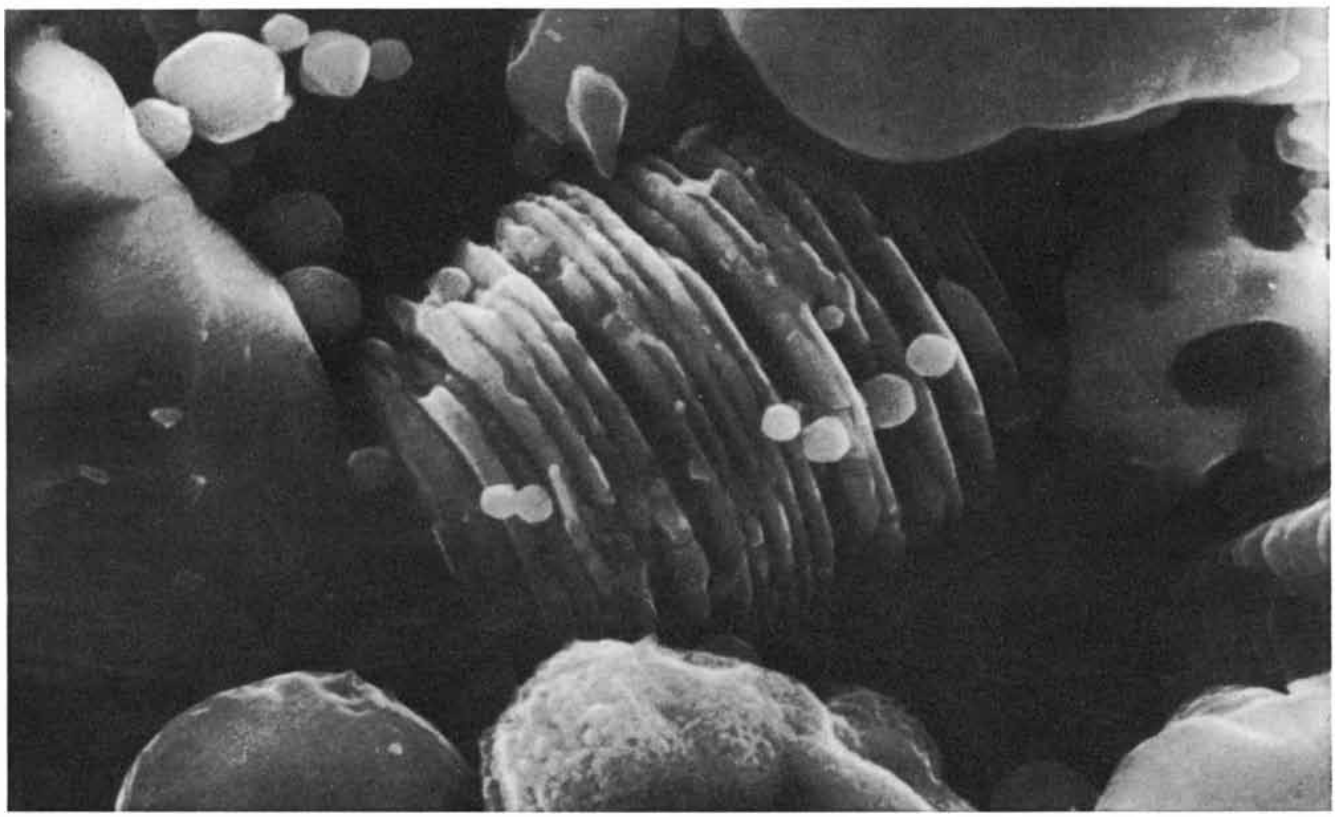
The solar system is believed to have formed out of a cold cloud of interstellar gas and dust. The cloud contracted under the influence of its own gravity and then spun down to a disk. A. G. W. Cameron of the Harvard College Observatory and his collaborators have utilized hydrodynamic models to determine the initial physical conditions within the rotating disk. If it is assumed that the disk had twice the mass of the sun, the models suggest that in the center the temperature was some 2,000 degrees Kelvin and the pressure was about 10^{-2} atmosphere (a hundredth of the atmospheric pressure at sea level on the earth today). At a distance of 10 astronomical units from the center in the central plane of the disk (10 times the distance from the earth to the sun), the temperature was only 100 degrees K. and the pressure was 10^{-5} atmosphere. Above and below the central plane of the disk the temperature and pressure also decreased.

If such conditions did obtain in the solar nebula, the temperature in the center would have been sufficient to have evaporated any solid grains that had existed in the original cold cloud. At distances beyond 10 astronomical units, however, the grains would have survived. Moreover, as most of the material in the inner region of the nebula continued to move toward the center of the disk, where it gave rise to the sun, what was left behind began to cool. In this remaining material, which eventually gave birth to the planets, the evaporated grains started to recondense, and a complex pattern of flows and eddies developed. Such movements may have been capable of transporting grains from one part of the nebula to another. The condensation was probably 95 percent complete within 50,000 years of the time the disk first formed.

J. W. Larimer of the University of Chicago was the first to attempt to calculate the sequence in which elements and compounds condensed out of the cooling solar nebula. His results were an approximation in the sense that he did not try to account for the changing composition of the vapor in the nebula as the major elements were removed by their condensation in solid phases. I have made similar calculations that take the removal into account by assuming that



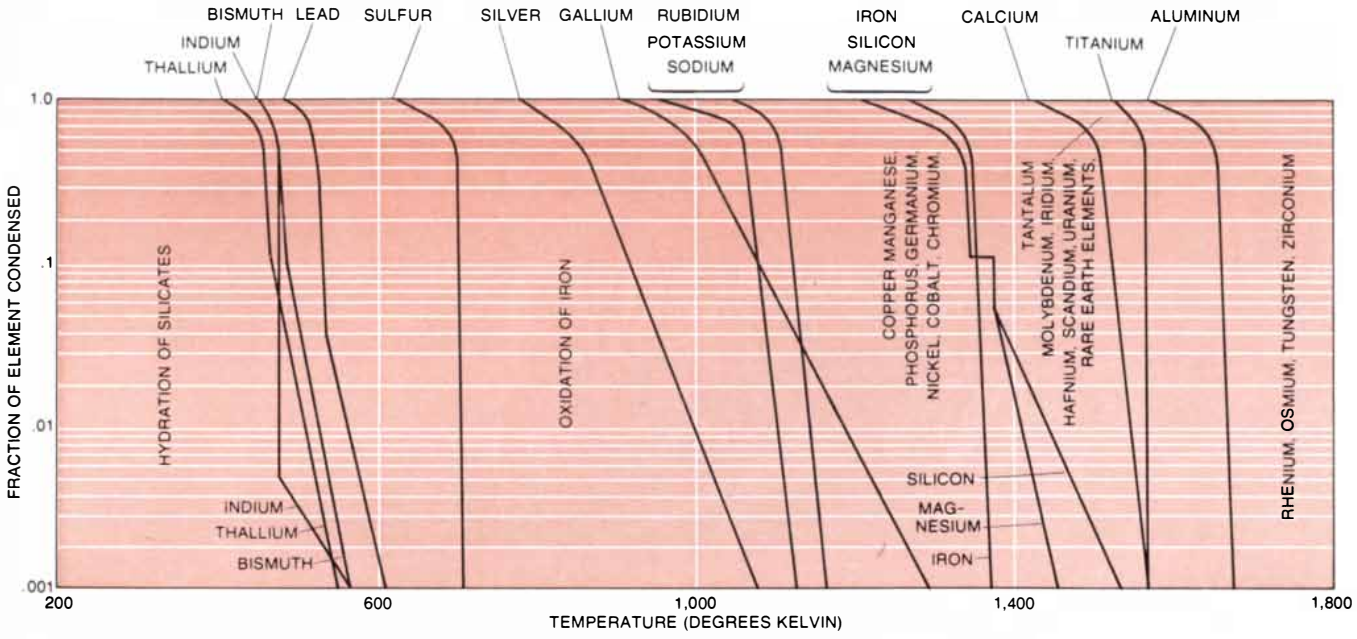
RASPBERRY-SHAPED AGGREGATE of crystals of magnetite (Fe_3O_4) may have condensed out of the nebula that gave rise to the sun and the planets at a temperature of some 400 degrees Kelvin (degrees Celsius above absolute zero). The aggregate was separated from the Orgueil meteorite, a carbonaceous chondrite of Type 1 (also designated a C1 chondrite). Shown here in a scanning electron micrograph made by R. S. Lewis of University of Chicago, aggregate has a diameter of about 10 micrometers (.01 millimeter).



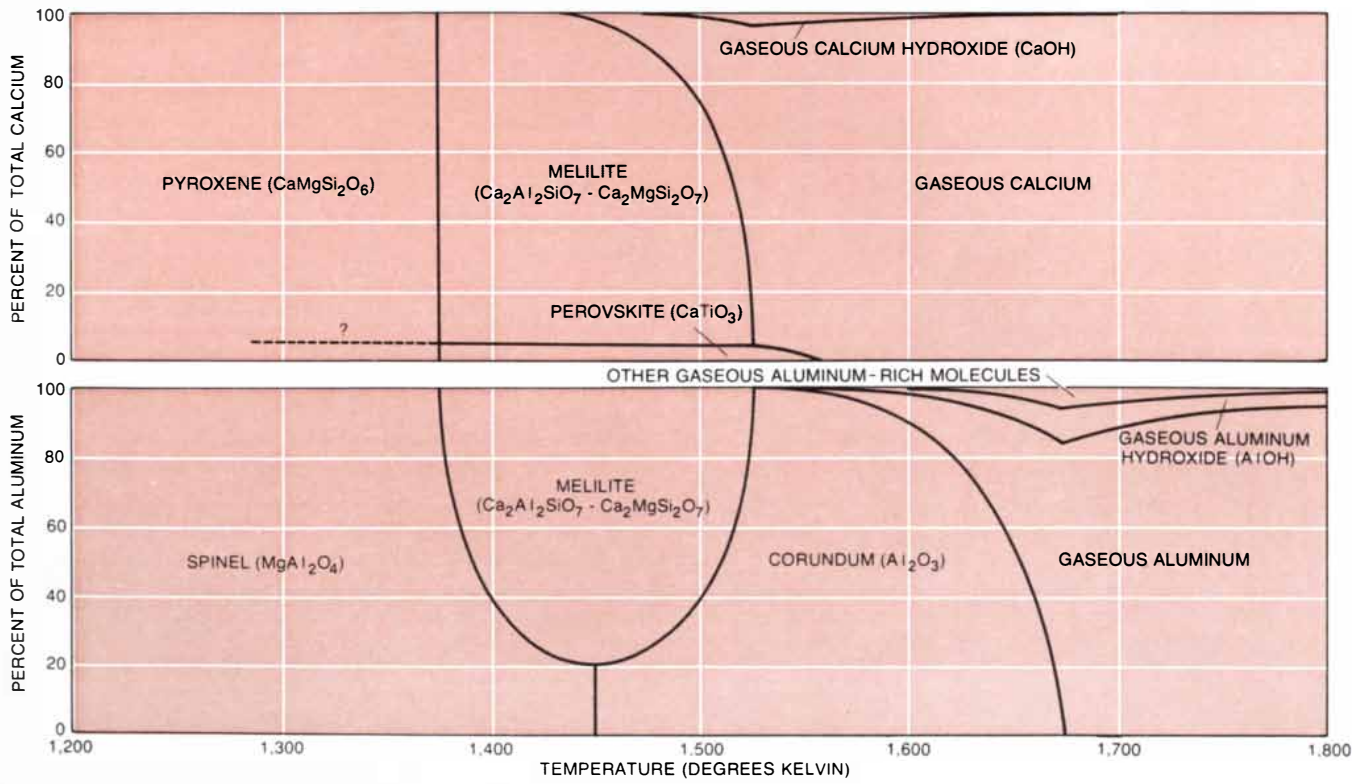
STACK OF CIRCULAR PLATES of magnetite was also separated from the Orgueil meteorite. The entire stack is about nine micrometers high. This scanning electron micrograph was also made by Lewis. Magnetite has not been found in this form on the earth.

the grains and the vapor were in chemical equilibrium.

If the composition of the system is known, the distribution of the elements among crystalline phases and gaseous species can be calculated from thermodynamic data for any temperature and pressure. The results I shall discuss here are derived from such a thermodynamic model, based on estimates of the composition of the solar system that are obtained from chemical analyses of C1 chondrites for the nonvolatile elements and from the spectrum of the sun for volatile elements. The initial temperature of 2,000 degrees K. and the pressure of 10^{-4} atmosphere I have chosen for the calculations are consistent with the estimates of those quantities for the inner portion of the nebula yielded by the hydrodynamic models. The general



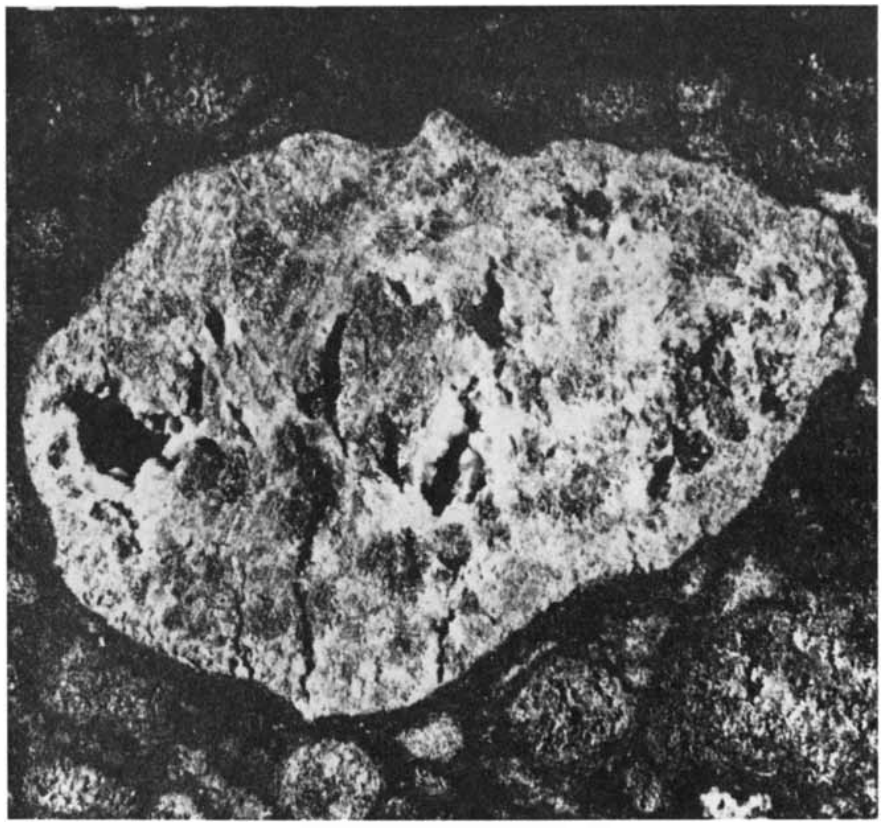
ELEMENTS CONDENSED from the solar nebula at different temperatures. The most volatile elements condensed first, followed by the less volatile ones. The general features of the sequence in which they condensed are shown, assuming that the average total pressure in one part of the nebula was 10^{-4} atmosphere. (One atmosphere is roughly the air pressure at sea level on the earth today.)



DISTRIBUTIONS OF CALCIUM AND ALUMINUM between their gaseous species and crystalline phases are compared as a function of temperature in the solar nebula. Aluminum had almost totally condensed by the time temperature reached 1,550 degrees K. Calcium had almost totally condensed at 1,475 degrees. Composition of melilite varies between $Ca_2Al_2SiO_7$ and $Ca_2MgSi_2O_7$.

features of the condensation sequence are relatively insensitive to pressure.

As the temperature dropped from 2,000 to 1,450 degrees, the relatively abundant elements aluminum, titanium and calcium condensed as silicates and oxides. At the same time a variety of relatively rare elements with high boiling points condensed as metals and oxides; they included some of the platinum metals, the rare earths and other elements such as zirconium, tantalum, uranium and scandium. The bulk of the condensable matter in the solar system condensed as the temperature fell from 1,450 to 1,000 degrees. A metallic iron-nickel-cobalt-chromium alloy appeared; magnesium, silicon, sodium and potassium condensed as silicates. Some trace elements such as germanium, copper, gallium and silver also condensed in this temperature region. Finally, as the solar nebula cooled to below 1,000 degrees, sulfur condensed as a result of reactions with metallic iron; some of the iron became oxidized, the silicates became hydrated (bound with water) and the most volatile trace elements (lead, bismuth, indium, thallium and mercury) condensed. I shall discuss in detail the theoretical predictions for the processes in each of the three temperature ranges, along with some of the evidence from carbonaceous chondrites showing that the processes played a major role in the origin of such meteorites.



WHITE INCLUSION on a cut surface of the Allende meteorite, photographed by E. A. King, Jr., of the University of Houston, may be a sample of material that condensed out of the solar nebula at the highest temperatures: between 1,557 and 1,375 degrees K. The Allende meteorite is a C3 chondrite. The spherical bodies surrounding the inclusion are chondrules, which are lacking in C1 chondrites. Inclusion is nine millimeters from left to right.

The amount of aluminum in various crystalline phases with respect to the amount in the vapor phase at various temperatures in the solar nebula can be calculated [*see lower illustration on opposite page*]. Aluminum began to condense in combination with oxygen as grains of the mineral corundum (Al_2O_3) at 1,671 degrees. By the time the temperature had dropped to 1,550 degrees the aluminum had been almost totally removed from the vapor. At 1,528 degrees the corundum started to react with gaseous calcium and silicon to form melilite. When the temperature reached 1,451 degrees, the remaining corundum reacted completely with the vapor in the solar nebula to form spinel ($MgAl_2O_4$).

Melilite is a mineral whose composition can vary continuously between $Ca_2Al_2SiO_7$, known as gehlenite, and $Ca_2MgSi_2O_7$, known as akermanite. The first melilite to form was pure gehlenite. Below 1,450 degrees, however, a large amount of akermanite would have dissolved in the crystal structure of the gehlenite as atoms of aluminum would have been replaced by atoms of magnesium and silicon that had struck the surface of the grains and diffused inward. The displaced aluminum atoms had to diffuse outward, where they formed particles of spinel on the surface of the melilite grains. As the temperature continued to drop, melilite became less important as a consumer of aluminum and simultaneously spinel became more important.

As aluminum was being consumed in this sequence of minerals, calcium was following a similar sequence. Calcium would have begun to condense with titanium and oxygen as grains of the mineral perovskite ($CaTiO_3$) at 1,557 degrees. The ratio of titanium to calcium in the solar system is so low, however, that only 3 percent of the calcium could have condensed as this mineral even when all the titanium had been consumed by it at 1,525 degrees. Most of the calcium condensed as melilite that later reacted with the vapor in the solar nebula to form pyroxene ($CaMgSi_2O_6$). At that temperature, 1,375 degrees, the melilite would have contained about 85 percent akermanite in solid solution.

How well does the observed composition of carbonaceous chondrites accord with the predictions of the thermodynamic model? There is an abundance of irregularly shaped white inclusions in the C3 chondrite known as the Allende meteorite and in other carbonaceous chondrites [*see illustration above*]. These objects can be extracted from the meteorite for analysis without being seriously contaminated by the matrix in which they are embedded. The composition of the inclusions compares well with the chemical composition calculated for the high-temperature condensates at two different temperatures [*see illustration on page 36*]. The similarity suggests that the white inclusions could actually be samples of the highest-temperature condensates from the original solar nebula, which were in equilibrium with the vapor of the nebula between 1,400 and 1,380 degrees. Only the amount of silicon dioxide (SiO_2) in the inclusions is higher than the amount that is predicted to have condensed. The excess of silicon dioxide could imply that kinetic factors might have prevented the grains from attaining complete chemical equilibrium with the vapor.

One can think of many other different assemblages of minerals whose bulk chemical composition would be the same as the chemical composition of the white

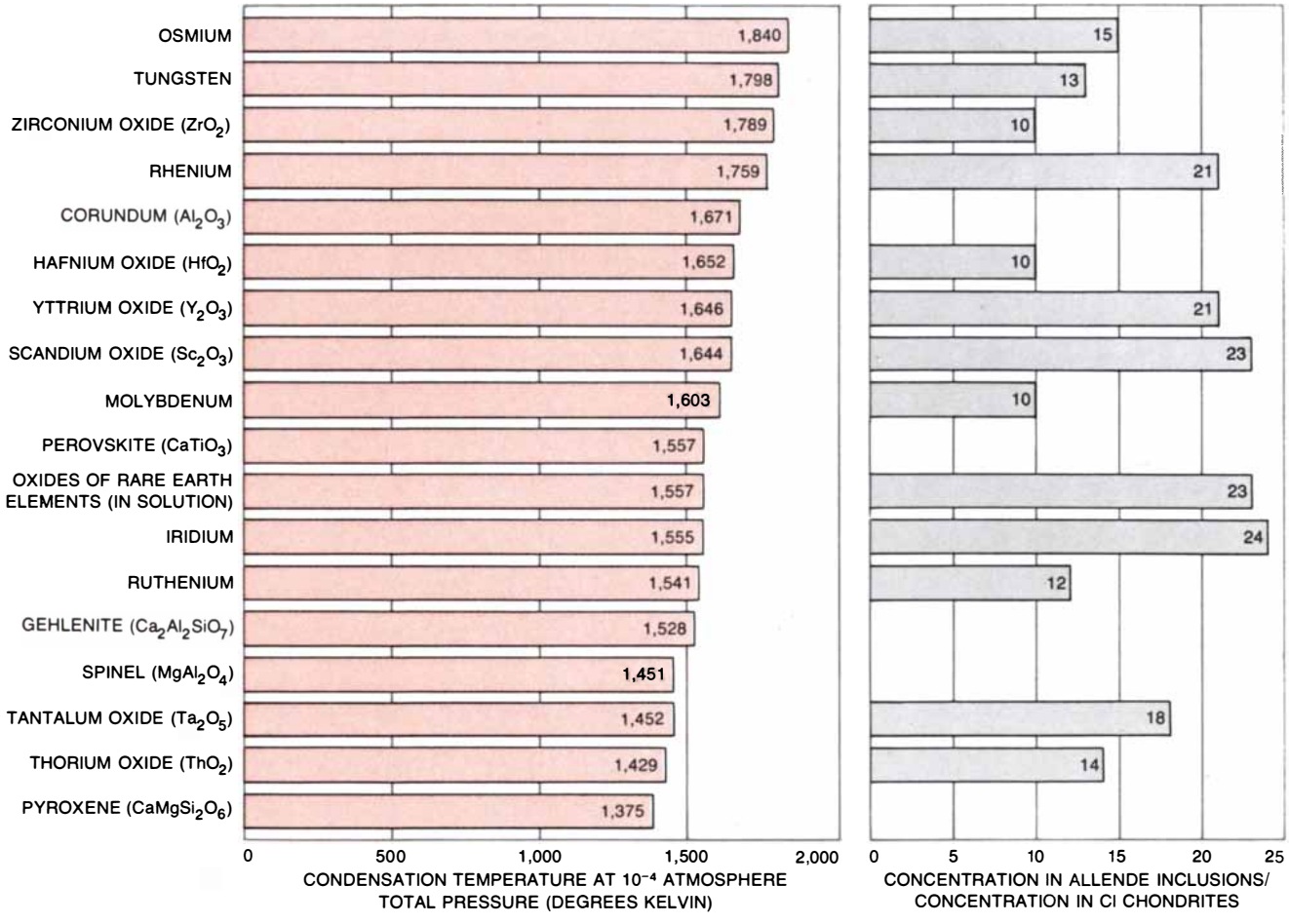
inclusions. Thus one could argue that the similarity between chemical compositions alone is by no means conclusive proof that the inclusions are matter that originated in the early solar nebula. Definitive evidence lies, however, in the mineralogy of the inclusions. They can be crushed to a powder and bombarded with X rays to get diagnostic diffraction patterns. They can be sliced and polished and studied with both the light microscope and the electron microprobe.

Such investigations have revealed that the inclusions consist predominantly of melilite, spinel, perovskite and pyroxene: precisely the same mineral phases predicted to be in equilibrium with the vapor of the solar nebula between 1,557 and 1,375 degrees. The interrelations and the textures of the minerals also suggest that they originated by condensation. The grains of perovskite are surrounded by spinel; the spinel and the melilite are grown together; rims of pyroxene surround the inclusions and cavities in their interior.

All these features are consistent with the form of the minerals predicted by the thermodynamic model. Usually the melilite in the inclusions contains less than 30 percent akermanite, implying that the melilite was unable to maintain chemical equilibrium with the vapor in the nebula below 1,425 or 1,400 degrees. The pyroxene rims are interpreted as signaling the beginning of the reaction of melilite with the vapor in the nebula at 1,375 degrees, but the fact that much of the melilite inside the inclusions was not converted to pyroxene suggests that the rates of diffusion were not fast enough to keep pace with the demands of chemical equilibrium, perhaps because the nebula was cooling very quickly. Moreover, scanning electron photomicrographs of the interior of cavities in some of the white inclusions reveal delicate free-standing crystals of melilite and other minerals [*see illustration at top left on page 37*]. Such well-formed crystals projecting into cavities are typical of minerals in terrestrial and lunar rocks that have been deposited out of a vapor.

Many trace elements have such a low vapor pressure (the pressure exerted by the vapor phase on the solid or liquid phase in equilibrium) or form such stable, nonvolatile oxides that the temperatures at which they condensed are as high as the temperatures at which the much more abundant aluminum and calcium minerals condensed [*see illustration below*]. The sensitive technique of neutron activation has been used to determine the concentrations of such elements in the white inclusions. When the inclusions are bombarded with neutrons, atoms in the material are made radioactive. The pattern of radioactive decay yields precise information about the composition of the material.

The abundances of all the nonvolatile elements in the calcium-rich inclusions of C3 chondrites are about 20 times



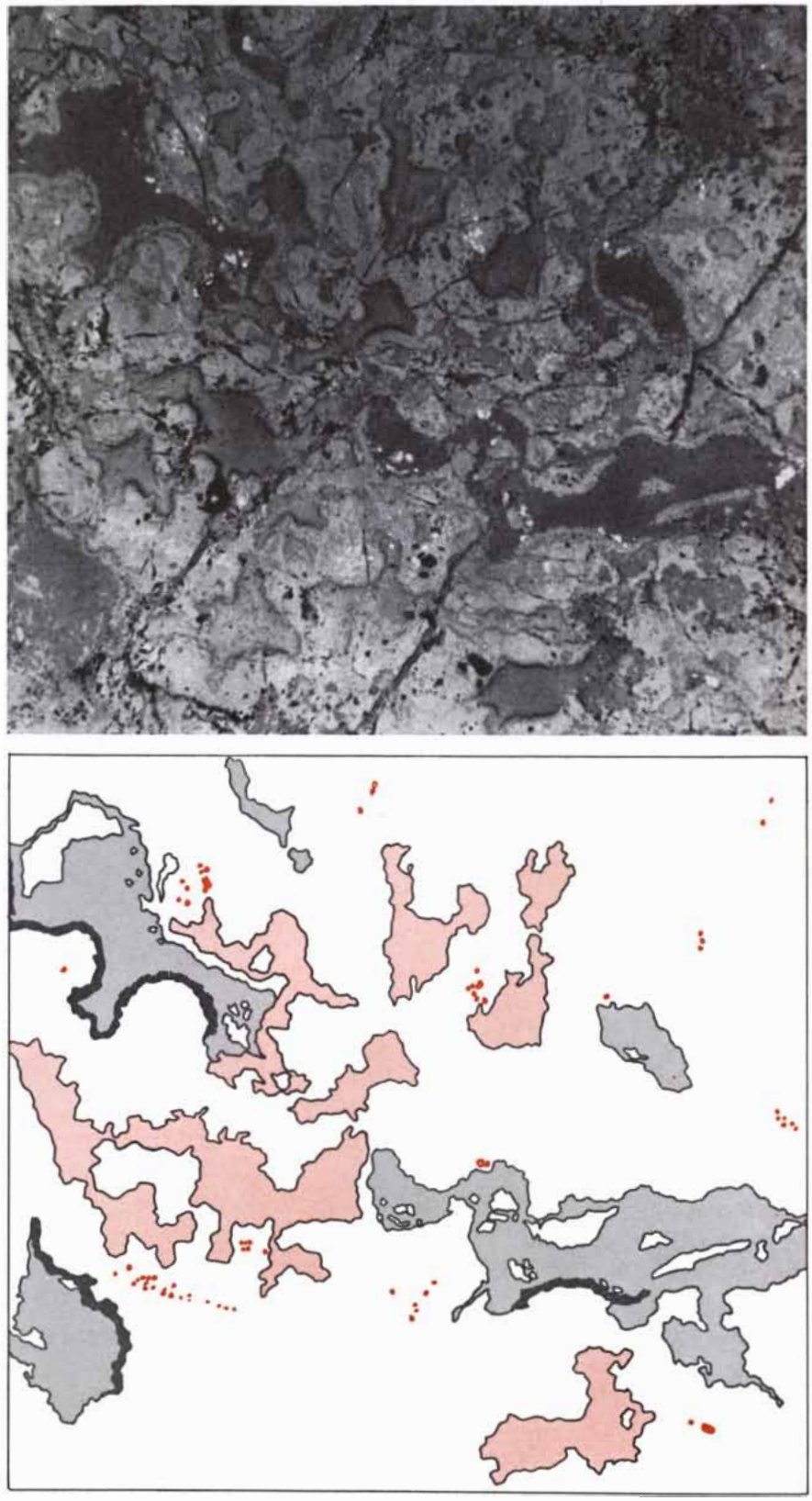
TRACE ELEMENTS IN THE INCLUSIONS of the Allende meteorite (osmium, tungsten, rhenium, molybdenum, iridium and ruthenium) all condense at very high temperatures. In the chart at the left the condensation temperatures of these trace elements are compared with the condensation temperatures of the major high-temperature minerals in the inclusions. The horizontal bar to the right of each element gives the temperature at which the element would condense at a pressure of 10^{-4} atmosphere. The bars that appear at the far right give the ratios of the concentration of these elements in the inclusions to their concentration in CI chondrites.

higher than the abundances in C1 chondrites. The nonvolatile elements in the inclusions span a wide range of chemical properties and geochemical behavior; the only property common to them all is that they have condensation temperatures in the same range as the condensation temperatures of the inclusions' major elements. The concentrations of several more volatile elements such as iron, copper and gallium have also been determined; they are all substantially less abundant in the inclusions than they are in C1 chondrites. Even palladium, a volatile metal of the platinum group, is strongly depleted in spite of an enrichment of the nonvolatile platinum metals ruthenium and iridium, which are its constant companions in normal geochemical processes.

The marked enrichment of the nonvolatile trace elements and the depletion of the volatile ones is thus strong evidence that the calcium-rich inclusions in the Allende meteorite and other C3 chondrites are samples of the highest-temperature condensates from the solar nebula. Some of the nonvolatile trace elements may have acted as nuclei for the condensation of the major minerals; conversely, the calcium minerals may have provided nuclei for the condensation of the trace elements. Some of the trace elements undoubtedly condensed in solid solution with the major minerals, as the rare-earth elements present in perovskite did.

We have seen that calcium, aluminum and titanium had almost totally condensed by the time the temperature in the nebula had reached 1,475 degrees. The thermodynamic model predicts, however, that until the temperature had dropped another 100 degrees these were the only abundant elements that had completely condensed in the solar system. At 1,370 degrees only 12 percent of the total amount of magnesium, 15 percent of the silicon and none of the iron had condensed out of the gas of the solar nebula. They are the most abundant nonvolatile elements in the solar system. If the nebular gases had somehow dissipated at this temperature before any more elements had condensed, the planets would now have compositions completely different from that of the chondrites.

The calculations indicate that at 1,365 degrees metallic iron would have begun to condense, carrying nickel, cobalt and chromium along with it. At 1,294 degrees gaseous molecular phosphorus reacted with the surface of the



POLISHED SURFACE OF A WHITE INCLUSION from the Allende meteorite was photographed by L. H. Fuchs of the Argonne National Laboratory (*top*). Total width of the field of the photomicrograph is one millimeter. Map (*bottom*) locates the natural cavities (*light gray*) and the minerals spinel (*white areas*), gehlenite (*light color*) and perovskite (*dark color*). The rims surrounding the cavities consist of pyroxene (*dark gray*). These are the same minerals that are calculated to have condensed out of the solar nebula at temperatures between 1,557 and 1,375 degrees at a pressure of 10^{-4} atmosphere. The spatial relations of the minerals suggest that inclusion was indeed formed by the calculated sequence of reactions.

metal grains, thus forming the mineral schreibersite (Fe_3P). The thermodynamic model further predicts that soon after iron first condensed, magnesium and silicon began to accumulate as the mineral forsterite (Mg_2SiO_4) at 1,361 degrees. The forsterite rapidly removed magnesium from the vapor and then began to react with the nebular gas, removing the remaining silicon as well, to form enstatite ($MgSiO_3$) at 1,294 degrees. By the time the temperature had dropped to 1,250 degrees more than 90 percent of the iron, magnesium and silicon had condensed.

An ion of ferrous iron (an iron atom lacking two electrons) can easily substitute for a magnesium ion in the crystal structure of forsterite and enstatite to form the compounds $(Mg,Fe)_2SiO_4$ or $(Mg,Fe)SiO_3$. At this stage in the evolution of the solar system, however, the ratio of molecular hydrogen (H_2) to water was very high, about 2,400 : 1. Molecular hydrogen easily gives up its electrons, so that metallic iron was prevented from oxidizing to ferrous iron. As a result, when the silicate minerals forsterite and enstatite first formed at high temperature, they contained vanishingly small concentrations of oxidized iron.

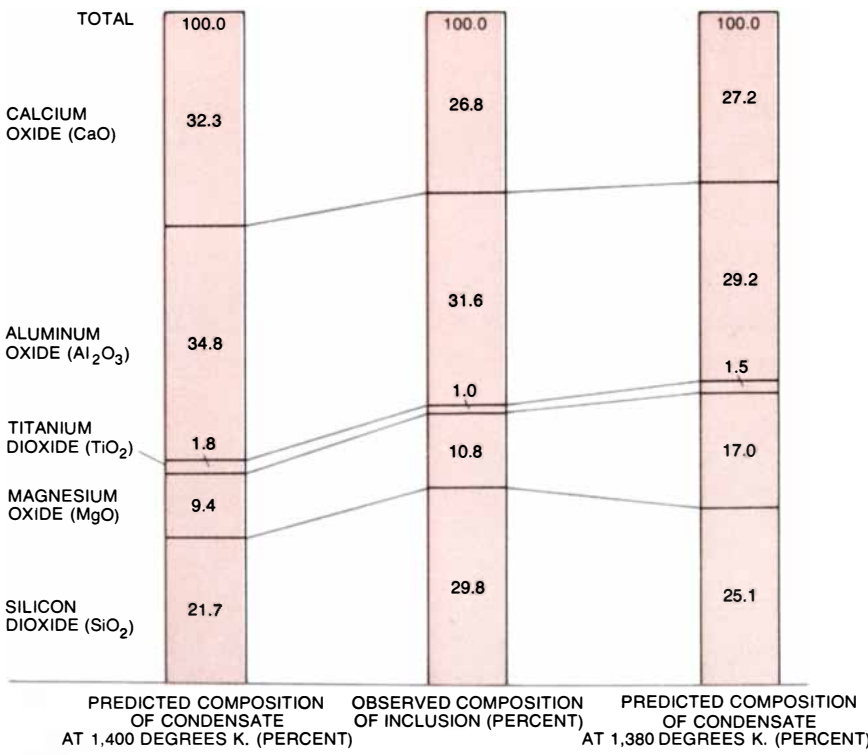
What kind of evidence is there for this sequence of events in carbonaceous chondrites? The C2 chondrites contain abundant well-formed crystals of forsterite and enstatite and also irregularly shaped aggregations of such crystals loosely held together. Most of these grains contain little oxidized iron, usually less than a few percent of either Fe_2SiO_4 or $FeSiO_3$. Examination of the aggregates with the petrographic microscope shows that there are many grains of metal about one micrometer in size trapped inside and between the silicate crystals. The metal grains consist mostly of iron, but they also contain small quantities of nickel, cobalt and chromium. Although schreibersite has not yet been detected, most of the grains are about 1 percent phosphorus. The observed compositions are almost exactly those predicted for alloys that formed out of the solar nebula over the same temperature range as forsterite.

In addition to the metal grains, another type of inclusion is found inside the crystals of forsterite and enstatite in C2 chondrites. These inclusions are minute ellipsoids of glass that are rich in calcium oxide (CaO) and aluminum oxide (Al_2O_3), often containing more than 20 percent of each by weight. They are chemically similar to the calcium-rich, high-temperature condensate inclusions from the Allende meteorite I have mentioned, and they may even be related to them. Their glassy nature implies that they were melted and rapidly chilled before they were incorporated into the crystals of forsterite and enstatite. The evidence thus suggests that the aggregates in C2 chondrites could have condensed from the solar nebula at intermediate temperatures.

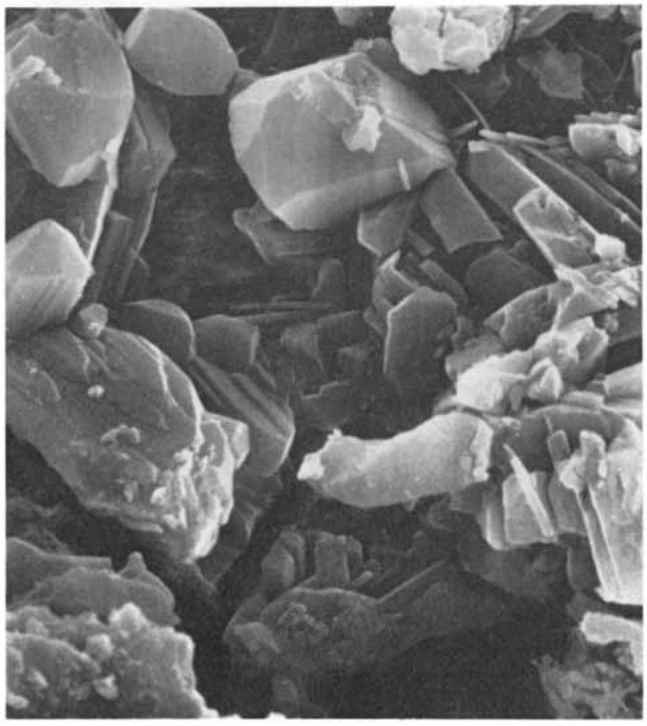
The sequence of events may have been as follows. Minerals rich in calcium and aluminum condensed above 1,375 degrees. They were melted during some kind of high-energy event in the nebula while they were suspended in the vapor. Then they cooled rapidly and froze into glass. Forsterite and nickel-iron alloy began to condense together at about 1,360 degrees, and the alloy grains were able to adjust their composition so that they were in equilibrium with the nebular gas until they were trapped inside growing crystals of forsterite along with the glass ellipsoids rich in calcium and aluminum. Once the metal grains were trapped they were unable to react with the gas any longer. The forsterite crystals continued to grow; some reacted to form enstatite and many began to stick to one another to form the aggregates. Thereafter the aggregates were mixed with lower-temperature condensates, which accumulated together to form the C2 chondrites, perhaps on bodies approximately the size of asteroids.

Now let me turn to reactions that are predicted to have ensued in the solar nebula below a temperature of 1,000 degrees. The most abundant gaseous molecule in the vapor was molecular hydrogen. At higher temperatures most of the carbon was present as carbon monoxide (CO). Below about 800 degrees the two molecules began to react with each other in the gas: one molecule of carbon monoxide combined with three molecules of hydrogen to form one molecule of methane (CH_4) and one of water. As a result the ratio of molecular hydrogen to water had dropped from 2,400 : 1 to about 640 : 1 by the time the temperature had reached 650 degrees. The lower ratio, together with the lower temperatures, allowed the metallic iron to oxidize and give rise to Fe_2SiO_4 and $FeSiO_3$, which could dissolve in the previously condensed forsterite and enstatite. Simultaneously gaseous hydrogen sulfide (H_2S) began to react with the surface of the metallic iron grains and the sulfur condensed as the mineral troilite (FeS) at 700 degrees.

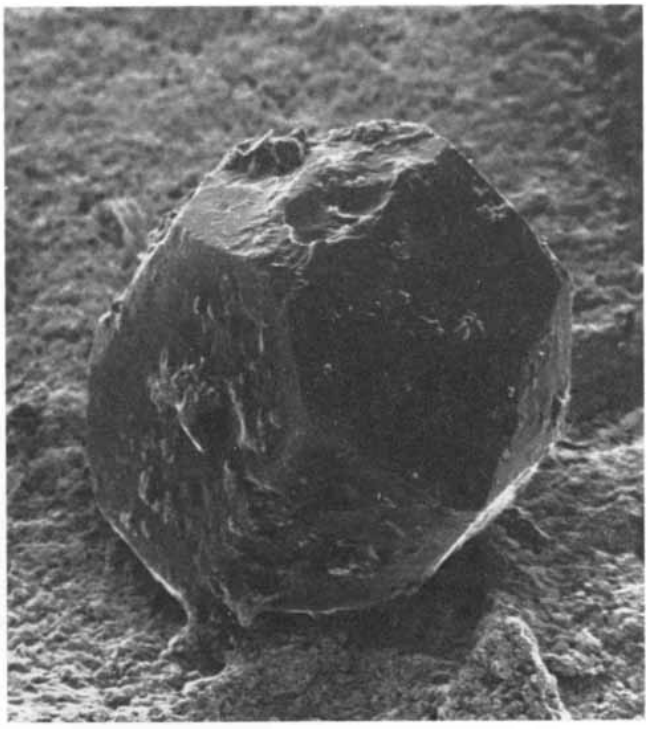
Edward Anders and his co-workers at the University of Chicago have concen-



CHEMICAL COMPOSITION OF A WHITE INCLUSION from the Allende meteorite is compared with calculated composition of condensates from solar nebula at two temperatures. In almost all cases predicted and observed compositions are very close to each other.

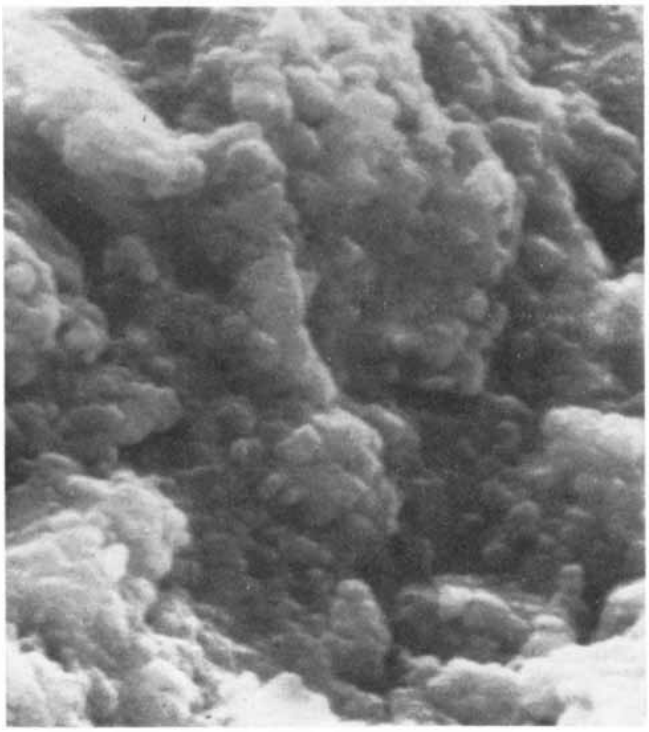


CALCIUM-RICH MINERALS in the inclusion in the Allende meteorite appear in this scanning electron micrograph. The fact that there are distinct well-formed crystals projecting into the cavity suggests that the minerals were formed by condensation from a vapor. The width of the field of view is about eight micrometers.

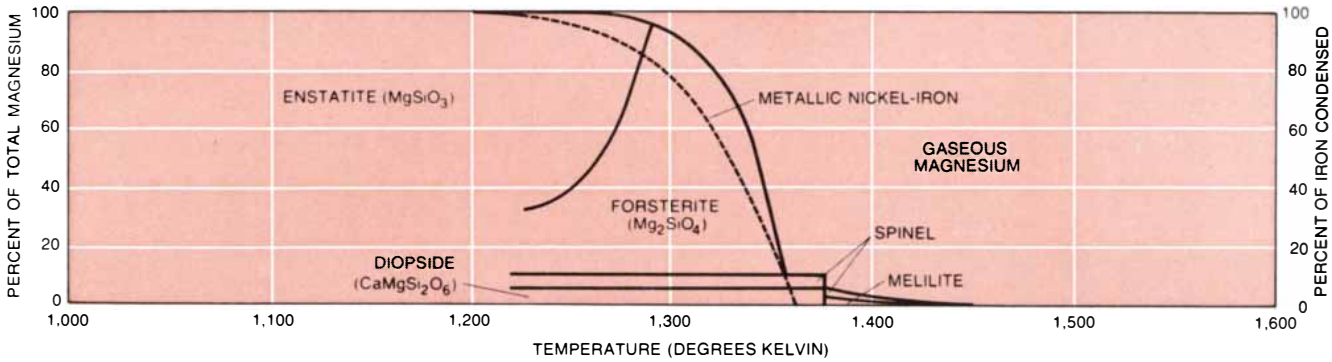


FORSTERITE CRYSTAL, also revealed in a scanning electron micrograph, was removed from the dark, fine-grained matrix substance of the Murchison meteorite, a C2 chondrite. Such crystals make up about 20 percent of the entire volume of the Murchison meteorite. The forsterite crystal itself is about .4 millimeter across.

MATRIX OF A C3 CHONDRITE, the Allende meteorite, is viewed in photomicrograph made by R. H. Beauchamp of Battelle Pacific Northwest Laboratories. Platelike material is half forsterite (Mg_2SiO_4) and half Fe_2SiO_4. Dark opaque grains are troilite (FeS). Composition is similar to composition predicted for materials that reached equilibrium with solar nebula between 700 and 400 degrees. The width of the field of view is 108 micrometers.



MATRIX OF A C2 CHONDRITE, the Murchison meteorite, is composed predominantly of hydrated layer-lattice silicates, that is, iron-magnesium silicate materials bound with water and having a layered crystal structure. It is believed the hydrated layer-lattice silicates condensed at temperatures of about 350 degrees. The field of view is 7.5 micrometers across. Even at this high magnification the exact forms of the tiny crystal grains are difficult to perceive.



DISTRIBUTION OF MAGNESIUM between gaseous molecules and condensed phases has been calculated for the solar nebula as it cooled. Condensation curve for nickel-iron is shown for reference (*broken line*). Magnesium had almost totally condensed by the time temperature had fallen to 1,275 degrees. Iron had almost totally condensed by the time temperature reached 1,250 degrees.

trated on the details of the condensation sequence below 700 degrees. Between 600 and 450 degrees the volatile elements lead, bismuth, indium and thallium condense as pure metals or as sulfides in solid solution. Organic compounds such as hydrocarbons are believed to have formed in the presence of catalysts in the nebula at about 400 degrees.

Again, what is the evidence from the meteorites? The high-temperature minerals I have mentioned as being present in the C3 chondrites are embedded in a fine-grained, dark-colored matrix. In order to study this material by light transmitted through it, special ultrathin slices of the meteorite must be prepared [*see illustration at bottom left on preceding page*]. The matrix is composed predominantly of tiny mineral plates about 10 micrometers long that have a chemical composition of about half Mg_2SiO_4 (forsterite) and half Fe_2SiO_4. Between the plates are opaque patches of troilite. The matrix is rich in lead, bismuth, indium and thallium and also contains about .8 percent carbon by weight. The composition is quite similar to the composition predicted for materials that reached equilibrium with the solar nebula between 700 and 400 degrees. There seems to be a slight tendency for the long axes of the silicate platelets to align themselves parallel to one another, suggesting that the particles were influenced by strong directional forces as they settled down onto the surface of the parent body of the Allende meteorite.

The calculations predict that in the final stages of the condensation of the solar nebula magnetite (Fe_3O_4) formed by the oxidation of preexisting nickel-iron alloy at 400 degrees. At 350 degrees the previously formed iron-magnesium silicate minerals began to react with water molecules to form hydrated silicates that have a layered crystal structure and are therefore called hydrated layer-lattice silicates.

The evidence for this sequence of events is found in the C2 chondrites. Like the C3 chondrites, the matrix of the C2 specimens is fine-grained and rich in volatile metals and organic material, but the mineralogy is different. The C2 chondrites lack the platelets of iron-magnesium silicate found in the matrix of the C3 chondrites but instead are composed predominantly of the hydrated layer-lattice silicates and magnetite [*see illustration at bottom right on preceding page*].

The C1 chondrites are composed of very few of the high-temperature minerals found in both the C3 and the C2 chondrites. Their mineralogy is quite similar to that of the matrix of the C2 chondrites. The morphology of the magnetite grains in the C1 specimens is peculiar [*see illustrations on page 31*]. Some of the magnetite is in the form of spherical aggregates of well-formed crystals, each crystal about a micrometer in diameter, and some is in remarkable stacks of circular plates about nine micrometers high. These stacks of plates are quite unlike any form of magnetite ever found on the earth. As a result it has been suggested that the meteoritic magnetite may have grown directly from the vapor of the solar nebula into these spectacular shapes.

We have seen, however, that magnetite first becomes stable in the solar nebula at 400 degrees, far below the temperature at which virtually all the iron had already condensed in the form of other minerals. Since only a negligible amount of iron remained in the vapor of the nebula, the implication is that the magnetite could not have condensed directly from the vapor but must have formed through the oxidation of preexisting iron minerals. If that had been the case, however, the shapes of the magnetite grains would have mimicked the shapes of the minerals that had been oxidized. The fact is, no other meteoritic minerals are known that have such peculiar shapes.

What can one conclude from these arguments? There does appear to be one way in which the magnetite could have condensed directly from the vapor. The pressures in the gas clouds of interstellar space are so low that grains are probably unable to condense out of them until the clouds contract into comparatively dense protostars such as the solar nebula. We have seen that the hydrodynamic models suggest that the peak temperatures reached inside the solar nebula were only a few hundred degrees at distances greater than several astronomical units from the center. In this region of the disk temperatures were never high enough to stabilize the various high-temperature minerals, and the first grains that condensed out of the nebula would have been the minerals in equilibrium with a low-temperature gas, phases such as magnetite and the layer-lattice silicates. Therefore these minerals could have condensed directly from the vapor in the absence of any high-temperature condensate as a precursor.

As I have indicated, there is much evidence that the carbonaceous chondrites are mixtures of unaltered minerals that condensed over a wide range of temperatures in the solar nebula before the planets formed. Locked inside the grains of material embedded in the carbonaceous chondrites are clues to the origin of the planets and the solar system as a whole. Research over the next decade will undoubtedly focus on analyses of the minerals, trace elements and isotopes in fragments separated from different carbonaceous chondrites in order to improve our understanding of the processes by which the solar system came into being.

We want to be useful ...and even interesting

Advice on TLC: snip, save, char

The mighty force of fashion powers more than the fashion industries. It governs the furnishing of the laboratory no less than the furnishing of the home. Don't knock it.

If talk of thin-layer chromatography (TLC) is slightly less fashionable today than when Eastman Chromagram sheet was introduced, it's because there are that many more instruments and methods to talk about now than then. TLC still works fine at separating molecular species, but its contribution toward building an impressive capital budget is a bit weak.

Our precoated sheet failed to obsolete the glass TLC plate. It did find its niche, though. A Chromagram sheet costs about the same as a precoated plate, but if only a piece will do, one needn't waste a whole plate. There are still those to whom such considerations matter. They use our product even though our poly(vinyl alcohol) binder has precluded charring in the visualization procedure.

Luckily our source of poly(vinyl alcohol) dried up. A composition based on poly(acrylic acid) has replaced it. With this substitution, nay improvement, for users of currently manufactured silica gel- or alumina-coated forms of Chromagram sheet (with or without incorporated fluorescent indicator), charring is no longer precluded!

How do you like that for candor?

You can order Eastman Chromagram sheet from Curtin Matheson Scientific, Fisher Scientific, North-Strong, Preiser Scientific, Sargent-Welch, SciChemCo, Scientific Products, or VWR Scientific (East).

Project for a slow week

Let us say he is good with tools and ought to be kept busy in a meaningful way. Let us say the professional staff makes use of the photographic process at a level that averages somewhat more than a single 135 cartridge of film processed per week but less than 100 gallons of processing solutions and wash water per day. Let us say there is no capacious municipal sewer system in which entities like acetic acid and thiosulfate ion that have served their photographic purpose can join all the other oxygen-demanding effluvia that characterize centers of civilization. Let us merely state for the record that Kodak pamphlet J-43, "A Simple Waste-Treatment System for Small Volumes of Photographic-Processing Wastes," is available for the asking from Kodak, Dept. 412-L, Rochester, N.Y. 14650. If there is also concern for recycling of a valuable resource, we can be asked to throw in J-9, "Silver Recovery with the KODAK Chemical Recovery Cartridge."

A perceptive type of customer

Weren't we once promoting a TLC method of mass-screening neonates for metabolic anomalies?

Yes.

Is it now in wide use?

No.

Why not?

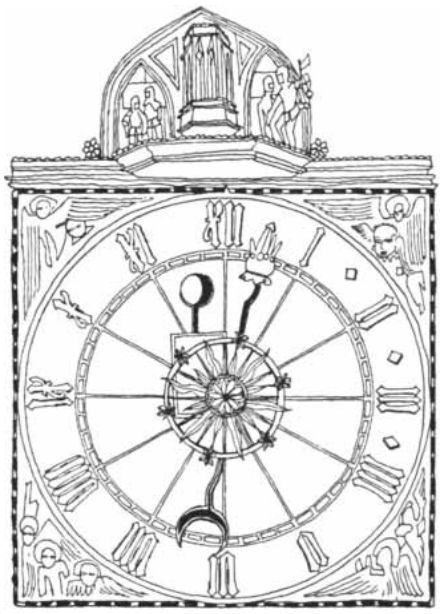
Because of a scarcity of people not only wise in blood chemistry but willing to spend their careers looking at thousands of different TLC patterns before finding one that justifies the cry "Aha!" From our long experience in medical x-ray film, we should have known how long it takes to develop such keen perceptions.

It is up to our industry to maintain a flow of what it believes to be improvements and up to the professional customers to decide whether they really are. At the moment the spotlight shines on a blend of terbium-activated gadolinium oxysulfide and lanthanum oxysulfide. When instead of a spotlight, x-ray photons in the diagnostically useful 50 to 150 kV range fall on this blend, formulated as fluorescing screens in contact with the emulsion on either side of a sheet of x-ray film, light is emitted at around 540 nm in the green, with a few small spikes in the blue and violet. All but 2 or 3 percent of the film exposing is done by light from these screens. The trick is very old but has generally been done with phosphors that largely emit a broad band of blue light, such as calcium tungstate. A couple of years ago we put on the market screens that emit principally in the ultraviolet, more of which is absorbed in the emulsion instead of spreading out to degrade the image on the opposite side of the film base.

Now, the lanthanide phosphors. In practical medical use, a pair of the new Kodak Lanex regular screens absorb about 60 percent of the x-radiation reaching them, as compared with 40 percent for the same thickness in $CaWO_4$. Absorption is the name of the game. Information-bearing x-ray photons that have passed through the patient are precious. With emulsions sensitized to green light, the absorption is convertible into several desirable currencies: reduction in patient dosage (popular, but not always the best choice for the patient's benefit); ability to freeze motion by shortened exposure time; better geometrical sharpness by reduction of focal-spot size when less tube current is required; use of the small focal spot to produce magnified images for improved diagnostic capability in some examinations; use of slower films for less "quantum mottle" from statistical fluctuation than when the delineation job is done with fewer of those mighty x-ray quanta.

Maybe flexibility is the name of the game.

Kodak



SCIENCE AND THE CITIZEN

Teravolt Territory

The new Congress will probably be asked to support three new accelerator facilities that will carry the study of subnuclear particles to new energy levels beginning in the 1980's. It is now eight years since Congress authorized construction of the world's most powerful research instrument, the 440-GeV (billion-electron-volt) accelerator at the Fermi National Accelerator Laboratory ("Fermilab"). The new machines have been recommended by a committee on new high-energy facilities headed by Victor F. Weisskopf of the Massachusetts Institute of Technology. The recommendations reflect the desire of the physical community to maintain three strong, geographically separated centers of high-energy research: the Stanford Linear Accelerator Center (SLAC) on the West Coast, Fermilab in the Middle West and the Brookhaven National Laboratory on the East Coast.

For SLAC, Congress will be asked to authorize "PEP," a positron-electron-proton storage ring capable of storing particles of up to 15 GeV circulating in opposite directions. PEP will be modeled on SPEAR, the successful electron-positron colliding-beam facility at SLAC that stores particles produced by the two-mile, 30-GeV linear accelerator. The SPEAR ring, an oval track with a circumference of 230 meters, is limited to storing particles of 4.5 GeV. The roughly circular PEP ring will have a circumference of about two kilometers. Designed jointly by physicists from Stanford University and the University of California at Berkeley, it will cost an estimated $53 million (at 1974 prices).

For Fermilab, Congress will be asked to support a development program looking toward the creation of a proton beam of about 1,000 GeV, or one teravolt (TeV), in the present accelerator tunnel by adding a new doughnut-shaped vacuum tube fitted with superconducting magnets. The proposed installation is referred to as the energy doubler, because it would work with protons accelerated to 400 or 500 GeV in the original ring. In testimony before the Joint Committee on Atomic Energy in the last Congress, R. R. Wilson, the director of Fermilab, pointed out that the superconducting ring would be justified if only to keep the laboratory's electric-power bill from reaching $10 million per year. (It was $3.5 million in 1973.) Wilson predicted that with the proposed ring in operation the power bill would be less than $4 million per year. He estimated that the energy doubler would cost $30 million. Also under study for Fermilab are rings capable of storing either protons of up to 1,000 GeV or electrons of 20 GeV. Collisions between protons in two 1,000-GeV beams would release an amount of energy equivalent to that released if the protons in a single beam of 2.2 million GeV, or 2,200 TeV, were to strike a target of protons at rest.

The third recommendation, involving only $3.5 million in the coming fiscal year, would be to establish manufacturing methods for building superconducting magnets on an industrial scale. The magnets would eventually be installed in a proton-proton colliding-beam facility called Isabelle, to be constructed at Brookhaven. The protons in each beam would have a minimum energy of 200 GeV. Isabelle, now being designed at Brookhaven, would cost $127 million at 1974 prices. The Weisskopf committee believes the three new facilities can be built over the next 10 years (and the present high-energy facilities kept running) at an annual cost of no more than $200 million.

Chilling Effect

Medical and biological research that involves the human fetus, or tissues from the fetus, has been all but abandoned in the U.S. as a result of the indictment in Boston of four physicians. The four are Agneta Philipson, Leon Sabath, David Charles and Leonard Berman; the charges against them concern an experiment they performed during 1971 and 1972, when all four were working at the Boston City Hospital.

The experiment was intended to investigate the efficacy of common antibiotics in pregnant women. It was proposed by Philipson, a Swedish physician who has since returned to Sweden; Sabath, who is now at the University of Minnesota School of Medicine, joined the project as an expert on antibiotics. Although the drugs employed in the study, erythromycin and clindamycin, are considered safe, Philipson and Sabath decided that in order to avoid all possible risk they would study only women who intended to terminate their pregnancy by induced abortion. They therefore enlisted the cooperation of David Charles, who was then performing abortions at the Boston City Hospital and is now at St. John's University in Newfoundland. Berman, a pathologist at the hospital, secured the aborted fetuses for study.

The experiment was approved by a hospital committee, and each woman who participated gave her written consent. The results were published in June, 1973, in *The New England Journal of Medicine*. It was found that pregnancy did alter the women's ability to metabolize the drugs, and that some antibiotics reach the fetus more effectively than others; the observations are considered to be of clinical significance.

The investigators are accused of violating an 1814 Massachusetts statute that prohibits grave-robbing. Their crime, according to the state, was performing unauthorized autopsies on the fetuses; in the prosecution's interpretation of the case, permission to examine the fetuses should have been obtained from the "mothers," whom the state considers to be "next of kin" to the fetuses. It was not then hospital practice to obtain such permission, essentially because an aborted fetus was considered to be not a corpse but a surgical tissue specimen. If the fetuses had not been employed in the experiment, they would have been incinerated (not formally cremated).

The issue raised by the indictments is

a familiar one in the debate over the morality and legality of abortion: whether or not the fetus is to be defined as a person. That issue was thought to have been settled in 1973 by the United States Supreme Court's decision in the case of Roe *v.* Wade, which declared unconstitutional most state laws regulating abortion. Opponents of abortion nonetheless continue to challenge that decision.

The same question is addressed more directly in the indictment for manslaughter of another Boston City Hospital physician, Kenneth Edelin. He is accused of killing a fetus, by "assault and beating," in the course of a routine induced abortion in October, 1973. The prosecution contends that the fetus was viable at the time of the abortion, although it could not have weighed much more than one and a half pounds and may have weighed quite a bit less. The Edelin case went to trial in January; as of this writing a verdict has not yet been reached.

Regardless of the outcome of either case, research on the fetus and on fetal tissue has virtually stopped. Sabath, whose work now does not involve fetal tissue, believes the study of the fetus could be "one of the richest areas of biological research." It could be particularly valuable, he says, in the investigation of cancer, congenital anomalies and certain diseases of childhood.

Frederick C. Robbins, dean of the Case Western Reserve University School of Medicine, points out additional areas of research in which the fetus or fetal tissue is essential. With Thomas H. Weller and John F. Enders, Robbins won a Nobel prize in 1954 for growing the poliomyelitis virus in a culture of fetal tissue; their work was essential to the development of a vaccine for poliomyelitis. Other kinds of fetal research, he says, involve the investigation of metabolism and development, both by examining the living fetus directly and by altering the physiology or environment of the fetus in anticipation of abortion. The experiment conducted by Philipson and her colleagues was of the latter kind. According to Robbins and others who are acquainted with the field, studies of this kind are not being made today. "Nobody is likely to work with fetal tissues right now," Robbins says.

Philipson, Sabath, Charles and Berman are now awaiting a hearing, at which several motions to dismiss the the charges against them will be made. Even if none of those motions is granted and the case goes to trial, it seems likely that the four investigators will eventually be acquitted. It is harder to predict how long the chilling effect of such proceedings on fetal research will last.

The Epidemiology of Heroin

As epidemiologists often point out, epidemiology is not merely the study of epidemics of infectious disease; it is the broad examination of the rates and patterns of disease in the community. By almost any standard drug abuse can be regarded as a disease; accordingly it can be profitably investigated by the methods of epidemiology. Heroin addiction in particular is examined in a group of epidemiological studies presented in December as a monograph supplement to *American Journal of Public Health*. The findings challenge some conventional assumptions about heroin use and the heroin user.

Mark H. Greene of the Center for Disease Control in Atlanta presents data indicating that Washington, D.C., experienced a heroin "epidemic" that peaked between 1968 and 1970. The data, developed by the Narcotics Treatment Administration of the District of Columbia, show that the incidence of first heroin use in the general population began to rise from a long-term base rate of .2 per 1,000 in 1961 to a high of 4.2 per 1,000 in 1969 and fell back to .2 per 1,000 in 1973. The population at risk were young people between the ages of 14 and 25. Of the individuals born between 1945 and 1954 still living in the District of Columbia approximately 10.5 percent of the males and 1.9 percent of the females had become heroin users by 1973 and had received treatment from the Narcotics Treatment Administration. In terms of the population most at risk (black males aged 14 to 25 years) the rate of first use of heroin rose from 1.9 per 1,000 in 1961 to a peak of 40.4 per 1,000 in 1969.

"In the epidemiologic model," writes Greene, "the infectious agent is heroin, the host and reservoir are both man, and *the vector is the drug-using peer.*" By tracing the spread of heroin use from person to person, the conventional notion of the "pusher" as the vector was discredited. In the large majority of instances a person was introduced to the use of heroin by a well-meaning friend, and in most instances the initiator was himself relatively new to the use of heroin. "Thus," Greene notes, "it is the *new user* of heroin who probably contributes most significantly to heroin use incidence."

In another paper Leon G. Hunt notes that data from heroin treatment programs throughout the U.S. show that new heroin use also peaked around 1968 in other major cities, which suggests that there was a simultaneous "national epidemic" of heroin use. Further analysis of the data, however, reveals that there was not a single peak but rather a series of local peaks ranging from 1967 to the present and that the sequence of local peak use of heroin is related to city size. Hunt finds that the number of new users of heroin peaked first in the larger cities and then shifted to medium-sized cities and is still peaking in places with populations smaller than 500,000. "This relationship implies," Hunt says, "that new heroin use may continue to appear in smaller cities in the future. If this relationship is correct, it has important implications for drug abuse policy and planning. First, it leads to an estimate for future levels of heroin use in 'new' cities. In the next five years, one might estimate that no more than 200,000 new users are likely to appear in places where heroin use has not yet become widespread. These estimated results suggest the need for continuous reallocation of drug treatment funds to smaller and smaller cities as peak use shifts and treatment demands decline in areas of older heroin use."

Reversal

A major French archaeological discovery that was declared fraudulent by many prehistorians in the 1920's has now regained credibility as a result of dating studies conducted at three independent laboratories. Beginning in 1924 a local farm boy and a local physician at Glozel, a rural community near Vichy, unearthed human and animal bones and a variety of artifacts: fishhooks, needles and harpoon points, all made of bone, pebbles engraved with images of deer and reindeer, fragments of polished stone axes, a kiln in which glass had been made and numerous objects made of fired clay, including more than 60 inscribed tablets. The physician, A. Morlet, soon became convinced that the material was of early Neolithic origin and therefore that at some time around 8000 B.C. France had been the cradle of both world literacy (the tablets) and glass technology (the kiln).

Morlet's hypothesis was understandably popular in France. At least one French prehistorian, Salomon Reinach, immediately endorsed the excavator's views, although Reinach put Glozel somewhat later in time: between 4000 and 3000 B.C. Other French scholars were less convinced of the site's antiquity. L. Franchet assigned it to the Iron

Age, around 100 B.C.; C. Julian, to the Gallo-Roman period, no earlier than A.D. 300.

None, however, suspected that the Glozel finds might be a hoax until the British prehistorian O. G. S. Crawford declared them to be so after a visit to the site. Crawford was supported in his view by at least one French scholar, Vayson de Pradenne. After initially favoring the Glozelians, the Abbé Breuil, a leading French prehistorian, reversed his opinion when an international commission formed in 1927 concluded that some of the artifacts from the site were of modern manufacture. The following year a second commission, principally French in membership, restudied Glozel and came to the opposite conclusion: the material was pronounced not only authentic but also Neolithic.

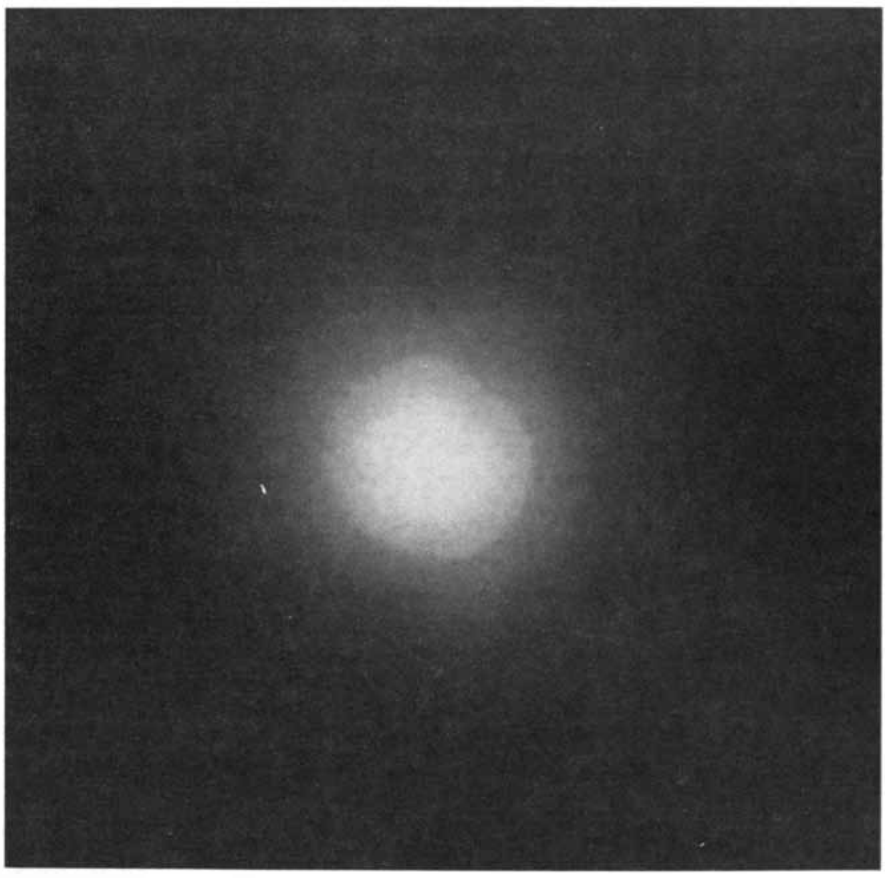
There the matter rested, with French honor saved at home but with the site so discredited abroad that prehistorians in general turned their backs on the Glozel finds. Last year, however, Hugh McKerrell of the National Museum of Antiquities of Scotland, Vagn Mejdahl of the Danish Atomic Energy Commission and Henri François and Guy Portal of the French Center for Nuclear Studies reopened the question when they completed a joint assessment of the clay artifacts by the technique of thermoluminescence dating.

Thermoluminescence dating is based on the fact that small crystal inclusions in pottery, such as grains of sand, accumulate energy as a result of being bombarded by ionizing radiation from the environment and from radioactive elements in the clay. When a sample of this crystalline material is heated, the energy is released in the form of light, the intensity of which can be measured with a sensitive detector. In general the greater the intensity of the light is, the longer the crystal has been bombarded and the older the pottery is. Moreover, the luminescence of the crystal is abolished by heating, so that the light intensity is a measure of the time that has elapsed since the pottery was fired.

In a recent report in *Antiquity,* McKerrell, Mejdahl, François and Portal relate that they took samples from more than 20 of the Glozel clay artifacts and tested identical subsamples in their separate laboratories. Their findings, consistent in all instances, indicate that the artifacts were fired about 300 B.C. The probable error in the estimated age (plus or minus 400 years) is large because the investigators' allowance for background radiation could only be estimated and is subject to correction. They believe that when actual background measurements become available, the probable error will be narrowed to plus or minus 200 years.

The age of the Glozel material was also assessed by other methods. Carbon-14 measurements of collagen from samples of bone proved to be consistent with the estimated age of the pottery. The fluorine-dating method, which was responsible for the exposure of the Piltdown Man hoax some 20 years ago, was also applied to samples of bone. This method measures the uptake of fluorine from ground water; if young bone, such as the fraudulent Piltdown jaw, is found in association with genuinely old bone, such as the Piltdown skull, the deception becomes evident because the bone that has been buried the longest will have absorbed the most fluorine from percolating ground water. All the Glozel bone specimens that were tested contained similar amounts of fluorine.

The investigators conclude that Franchet and Julian seem to have been right in dating the Glozel material around the beginning of the Christian Era. To dispose of the Glozel question once and for all, they suggest, the site should now be revisited, and whatever artifacts the new excavations may yield should be similarly analyzed by modern dating techniques.



BETELGEUSE, a giant star some 500 light-years from the earth, is seen as an extended image in this enhanced photograph made recently by astronomers at the Kitt Peak National Observatory. The slightly mottled surface and tenuous outer edge of the image are thought to be partly genuine features of the star and partly artifacts of the photographic technique.

Zooming in on Betelgeuse

It is virtually an axiom of astronomy that with the sole exception of the sun stars always appear in even the largest telescopes as essentially point sources of light. This statement is no longer true. In a recent announcement astronomers at the Kitt Peak National Observatory report that they have succeeded for the first time in resolving an extended photographic image of a distant star [*see illustration at left*]. Moreover, they have found that close scrutiny of the recorded data reveals surface features reminiscent of the convection patterns observed on the surface of the sun.

The subject of the new photographs was the giant reddish star Alpha Orionis, commonly known as Betelgeuse, one of the brightest stars visible to the unaided

eye. The photographs were made by Roger Lynds, Jack Harvey and Peter Worden, who used a novel interferometric technique in conjunction with the 158-inch Mayall telescope on Kitt Peak. The Mayall reflector, currently the second-largest optical telescope in the world, has been in operation for less than two years.

According to Lynds, the method employed by the Kitt Peak group to reconstruct an image of Betelgeuse differs from the earlier technique of "speckle interferometry," which indicated the presence of stellar surface features but did not produce an actual picture of the star. Both techniques rely on short exposure times to minimize distortions caused by turbulence in the earth's atmosphere. In the Kitt Peak approach the faint image of Betelgeuse was amplified by means of a special image intensifier, and the resulting pictures were then processed by a large computer to remove whatever extraneous fuzziness remained.

The slightly mottled appearance of the star's surface in the reconstructed photographs is thought by the astronomers to result from temperature inhomogeneities similar in structure to the vast regions of hotter and cooler gases apparent on the sun's surface. Additional data are being processed to learn more about the surface features of Betelgeuse, a star with a diameter some 800 times greater than that of the sun.

Hot Spot

Yellowstone National Park, which has some 10,000 geysers, hot springs and other thermal phenomena, is one of the world's most famous hot spots. Now it appears that the park may provide the first direct evidence for the existence of the thermal plumes that are widely believed to provide the driving mechanism that moves large lithospheric plates (and the continents they carry) about on the surface of the earth. The evidence was described by H. M. Iyer of the U.S. Geological Survey at a meeting of the American Geophysical Union.

A thermal plume is envisioned as a column of magma (molten rock, hot liquid and gas) a few hundred kilometers in diameter that rises from deep in the hot mantle of the earth. According to the hypothesis that some 20 thermal plumes provide the driving force for the movement of crustal plates, a plume reaching the lithosphere spreads radially under the crust, giving rise to the crustal movements. At the surface a plume creates a hot spot with volcanic activity.

Yellowstone has been the scene of intense volcanic activity over the past two million years. The most recent volcanic eruption, which took place about 600,000 years ago, created the caldera, or giant crater, that occupies most of the central part of the park. The volcanic activity and the other thermal phenomena in the park are known to have been caused by a body of magma under the park, but until recently there were no means for ascertaining the width and depth of the magma body.

Since Yellowstone is also a region of much earthquake activity, the U.S. Geological Survey began in 1963 to operate a network of seismograph stations to monitor the earthquakes. The seismographs record the waves originating from both local and distant earthquakes. It is the distant ones, according to Iyer, that have provided the evidence that Yellowstone may sit at the top of a thermal plume. Iyer and his colleagues John R. Evans and John M. Coakley found a large delay in the waves from distant earthquakes that had traveled through the magma under the park. By plotting the paths of the waves thus delayed, Iyer and his associates were able to estimate for the first time the shape and dimensions of the body of magma under the park. It appears to be rather pear-shaped, some 50 kilometers wide at the middle and at least 150 kilometers deep. The full depth has not yet been determined. Preliminary evidence from mobile seismic stations, Iyer said, indicates that the magma body may extend downward several hundred kilometers into the mantle. Readings from additional seismograph stations now being installed, he said, may provide enough information to resolve the question.

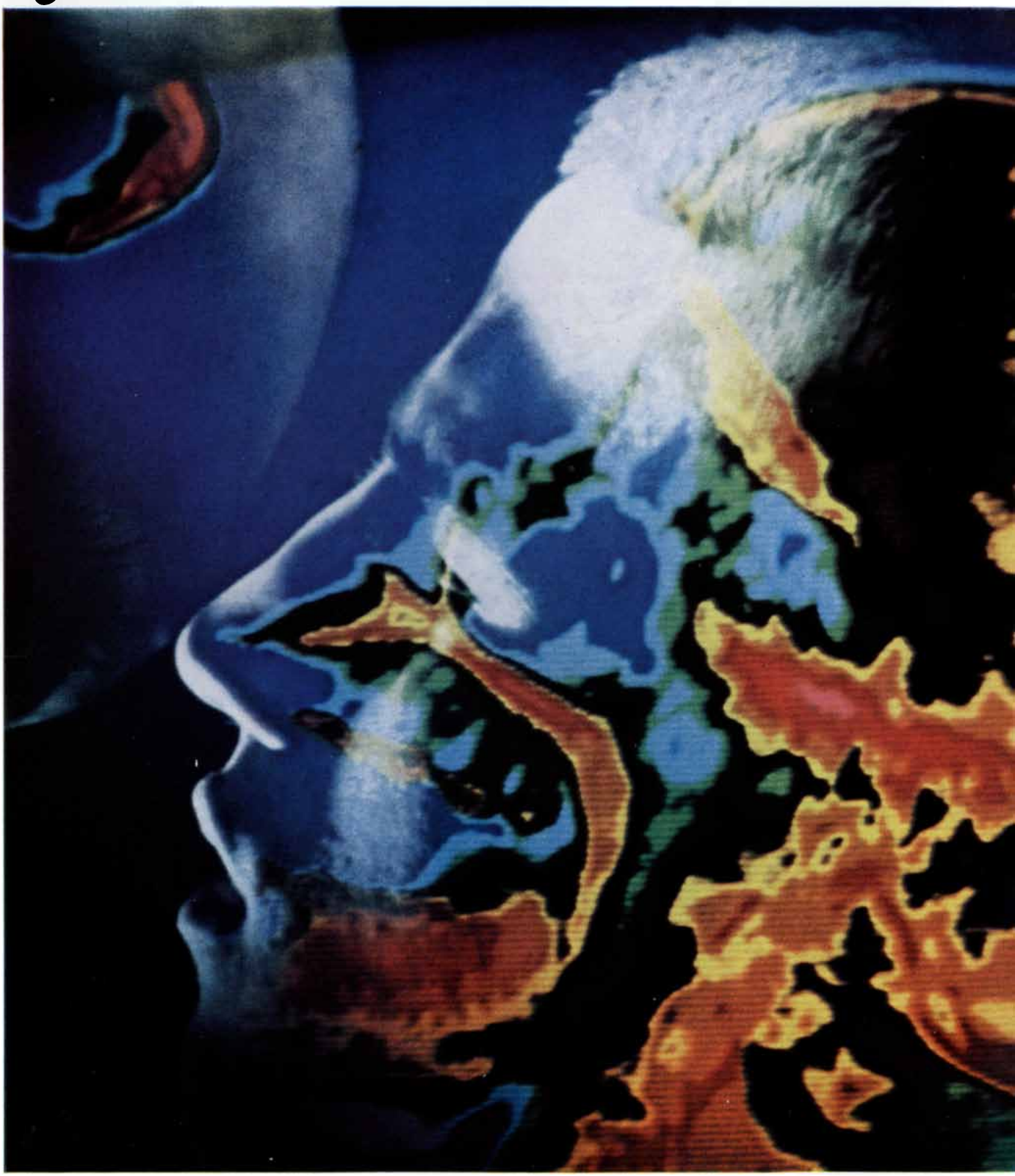
Naked Lunch

Some people like dark meat and some like light, and some even eat the gizzard, but no one eats the feathers. It was therefore a matter of more than academic genetic interest when, in 1957, Ursula K. Abbott of the University of California at Davis discovered a mutant strain of chicken without feathers. A food animal is a converter of feed into food; the more efficiently it converts, the better. To eliminate an entire tissue that does not constitute food should be a great help. The interaction between biology and agricultural economics is seldom simple, however, and featherless birds have yet to reach the market. At the University of Maryland, Max Rubin and Daniel E. Bigbee have made some studies of the nutritional requirements and processing characteristics of experimental featherless chickens.

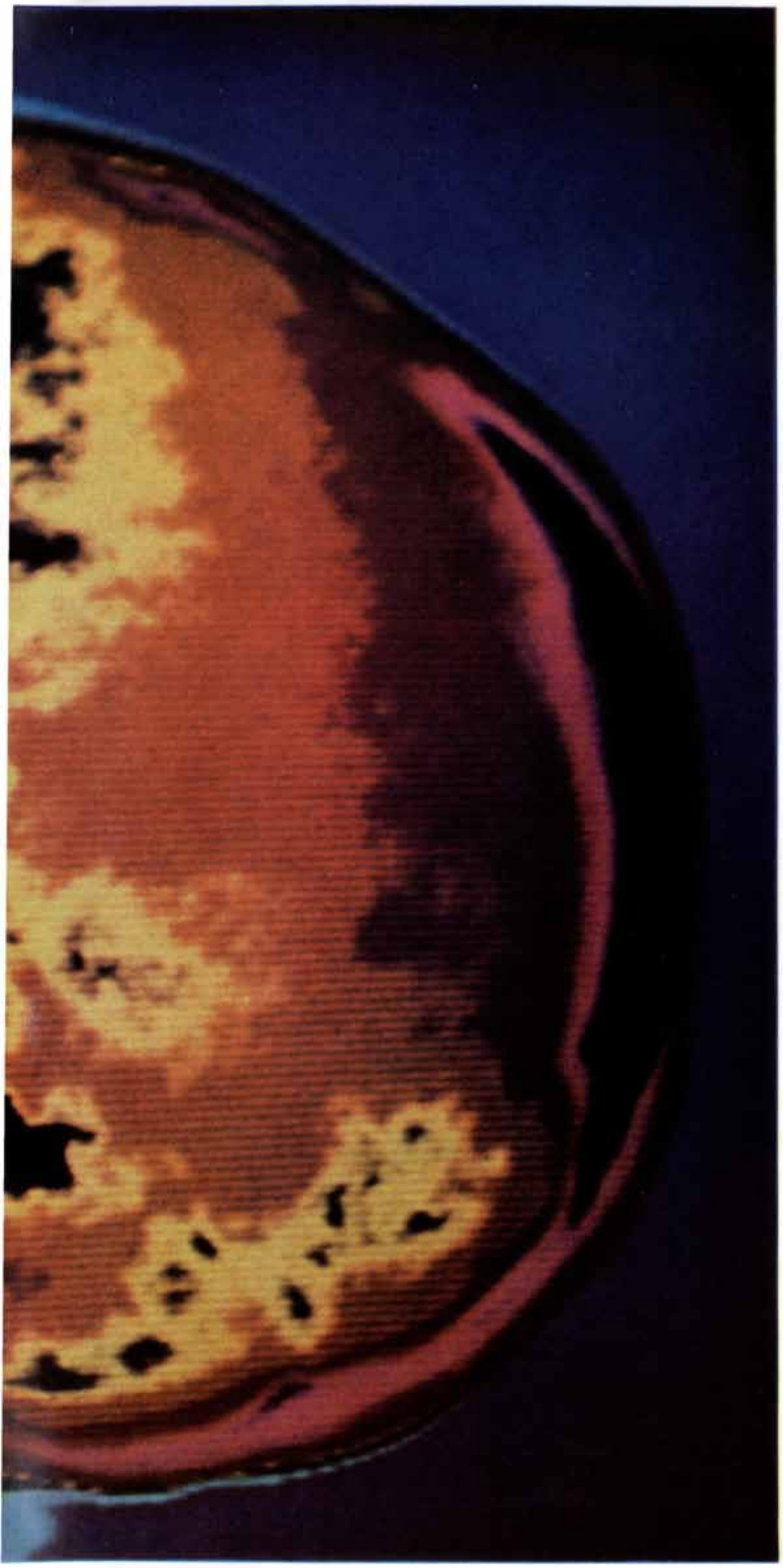
Abbott found that featherlessness (actually a lack of scales; feathers are modified versions of scales such as those that cover a chicken's feet) is governed by a single recessive gene. If a feathered bird is mated with a featherless one, the progeny are feathered; if the progeny are mated in turn, about 25 percent of the second-generation birds are featherless. The trouble is that a naked chicken is a cold chicken. The mutation is a temperature-sensitive lethal one: without special handling featherless chickens die of excessive heat loss. (On the other hand, they survive in heat waves that may kill some normal chickens.) In the absence of a heat wave the featherless animals can be raised in a heated environment, but heating a chicken house into the 80-degree range costs money. Moreover, even in a warm space the featherless chicks consume more food energy maintaining their body heat than normal chicks, and that energy loss would presumably have to be more than balanced by the energy saved through the nonproduction of feathers if the chickens are to be economically viable. Rubin has calculated that about 25 percent of the protein in the live weight of a normal chicken is in the protein-rich feathers. He finds that the feathers have a high priority for the essential amino acids supplied by the feed; feather growth has precedence over the total growth rate. In other words, a featherless bird makes more efficient use of the essential amino acids to produce muscle than its feathered siblings do during the early growth period, and it gains weight more rapidly.

Taste tests have indicated that the meat of featherless broilers is similar to that of their feathered siblings. The featherless birds lose less weight in cooking, according to Bigbee, because their muscles contain less fat. The nominally "featherless" birds retain about 1 percent of the normal plumage in the form of primitive feathers. These would have to be removed if a commercial meat chicken were developed, and it turns out that the rubber-fingered rollers that whip the plumage off chickens in broiler factories have more trouble with the primitive feathers than with full plumage. Bigbee has established, however, that the rollers can manage. That leaves a final question. Featherless chickens have smooth skin. If a featherless broiler ever reaches the market, how will consumers, accustomed to the familiar pocked texture of a properly defeathered feathered bird, react?

Why jump into a you've studied all the



specialized field until possibilities?



If you're a new physician, you might not have decided on a field of specialization yet. That's easy to understand. But while you're deciding, you can be getting valuable experience and expert guidance towards that decision as a physician in the U.S. Air Force. Air Force physicians are exposed to many fields of medicine. They practice with highly professional people in modern facilities. And all Air Force physicians have seemingly unlimited opportunities for furthering their education and professional growth. Air Force graduate medical education programs are available. And because Air Force physicians participate in Group Practice (a collection of specialists in different fields complementing each other by expanding total capability) they all expand their individual ability and horizons with each new patient.

As an Air Force physician (and officer) you can apply for worldwide work assignments. The salary is excellent. And you get a 30-day paid vacation each and every year.

If you're still deciding on a medical specialty...or a future...look into the Air Force before you make a final decision. We think you'll like what you see. For the location of the Air Force Medical Recruiting organization serving your area, fill in the coupon.

Health Care Opportunities
P.O. Box AF
Peoria, IL 61614

C-SA-25

Yes, I am interested in Air Force Health Care opportunities. I understand there is no obligation.

Name _____ Sex ☐ (M) ☐ (F)
(Please Print)

Address _____ City _____

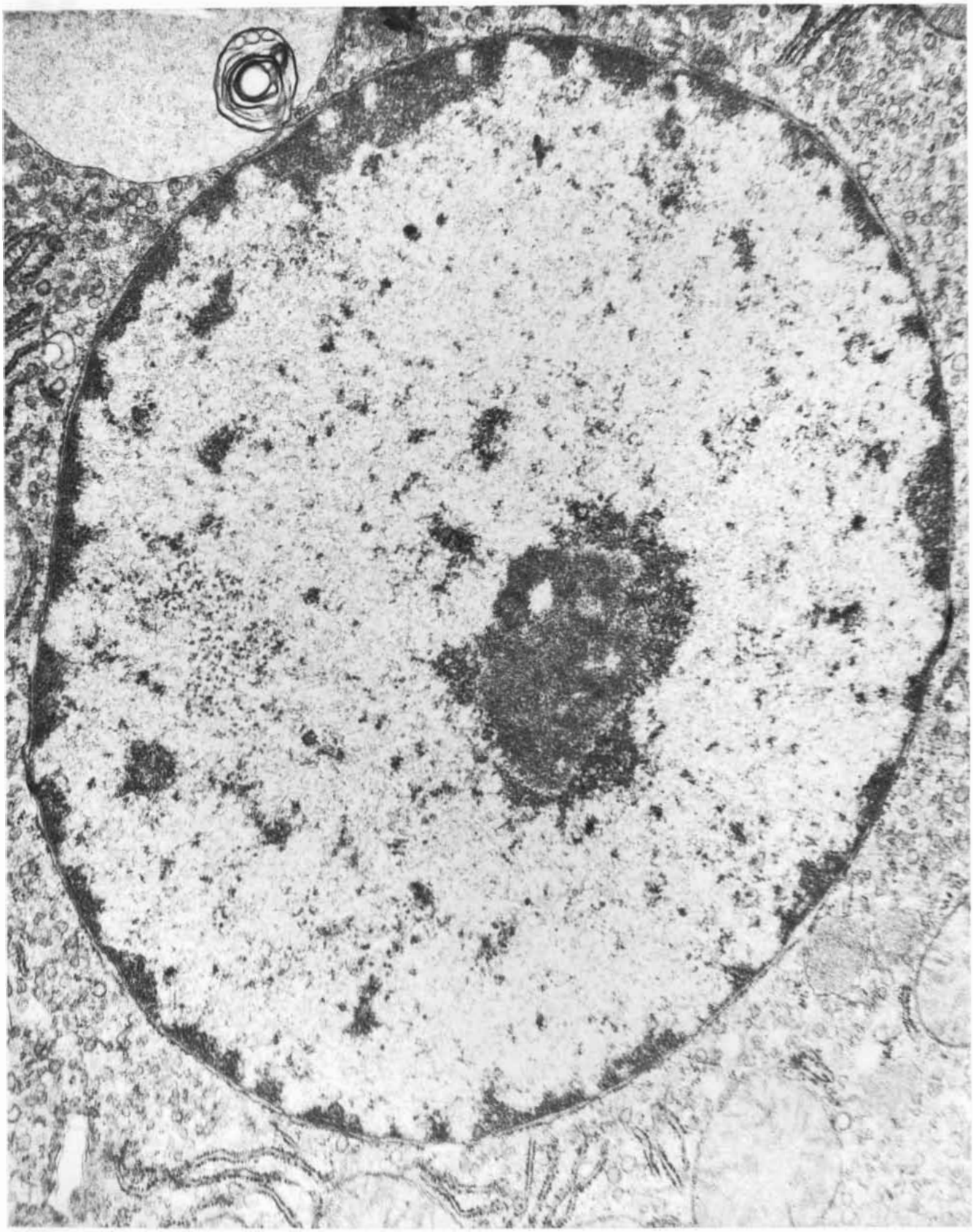
State _____ Zip _____ Phone _____

Soc. Sec. # _____ Date of Birth _____

Specialty _____

School _____ Year of Degree _____

Health Care at its best. Air Force.



NUCLEUS OF MOUSE LIVER CELL is enlarged 22,000 diameters in this electron micrograph made by David E. Comings and Tadashi Okada of the City of Hope Medical Center. The nucleus, bounded by its thin double membrane, is filled with chromatin, the material that accomplishes the genetic and control functions of the cell. In chromatin the genetic material, DNA, is complexed with proteins whose role is apparently to regulate the activity of the genes. The dark material near the membrane and elsewhere in the cell is "condensed" chromatin, which is not actively synthesizing RNA; the lighter material that fills most of the nucleus is "extended" chromatin, which is synthesizing RNA. The fine granular material that can be seen at left center in the nucleus consists of granules of RNA complexed with protein. The large dark structure visible near the center is the nucleolus, a region of specialized chromatin.

CHROMOSOMAL PROTEINS AND GENE REGULATION

The role of the proteins associated with DNA in the nuclei of higher organisms is beginning to be understood. Apparently the histones keep genes turned off and the nonhistone proteins selectively turn them on

by Gary S. Stein, Janet Swinehart Stein and Lewis J. Kleinsmith

As is now well known, the genes, which transmit the hereditary information from one generation to the next and direct the function of every living cell, are made of DNA. In the cells of higher organisms, however, the genes are arrayed on chromosomes, and chromosomes are nucleoproteins: complexes of nucleic acids and proteins. If the DNA is the genetic material, what is the function of the chromosomal proteins? Apparently they play a major role in maintaining the structure of the genetic material and in regulating the activity of the genes, that is, in determining which genes in each cell are turned on, and when. The control of genes is central to such fundamental processes as differentiation, embryonic development and hormone action, and to such abnormal processes as cancer, metabolic diseases and those birth defects that are related to the expression of genetic information. Here we shall be concerned primarily with the regulatory activity of the chromosomal proteins, an area of investigation that has begun to yield significant results in the past few years.

Gene Regulation

A cell's genetic information is encoded in the sequences of nucleotides that constitute its DNA. To utilize this information the cell transcribes nucleotide sequences into the complementary strands of RNA that are translated into chains of amino acids to form proteins. The control of gene activity resides in the selectivity of transcription in both time and place. In embryonic development, for example, a single fertilized egg containing all the organism's genes proliferates to give rise to a vast variety of differentiated cells that are specialized for different functions. In the course of development hundreds of thousands of genes are turned on and off. Moreover, in each of the ultimate cell types only a certain combination of genes is transcribed, or expressed, in accordance with the cell's specific function. In red blood cells it may be the gene coding for hemoglobin that is turned on; in muscle cells, the gene for myoglobin; in fibroblasts, the gene for the connective-tissue protein collagen.

And yet, as far as anyone can tell, the DNA in each of these specialized cell types is the same. The constancy of DNA in all the cells of a given organism was first established in 1948 and has been confirmed in many ways, most elegantly in the experiments of J. B. Gurdon of the University of Oxford. He transplanted nuclei from differentiated intestinal cells of the frog into frog eggs whose nuclei had been removed. The eggs developed normally. Such results demonstrated that the nuclei of intestinal cells and all the other differentiated cells of the frog contain all the frog's genetic information, even though only a limited portion of that information is expressed.

Even within a given differentiated cell certain information is expressed at one time and remains unexpressed at other times. The expression can be modified when resting cells are activated to divide and when cells are stimulated by specific hormones; in both instances complex changes in the cell's metabolism are effected by the switching of the cell's genetic program. In general less than 10 percent of the total genetic information is expressed at any one time. Specific regulatory mechanisms activate and inactivate appropriate regions of the genome, or full set of genes, for transcription as the needs of the cell change.

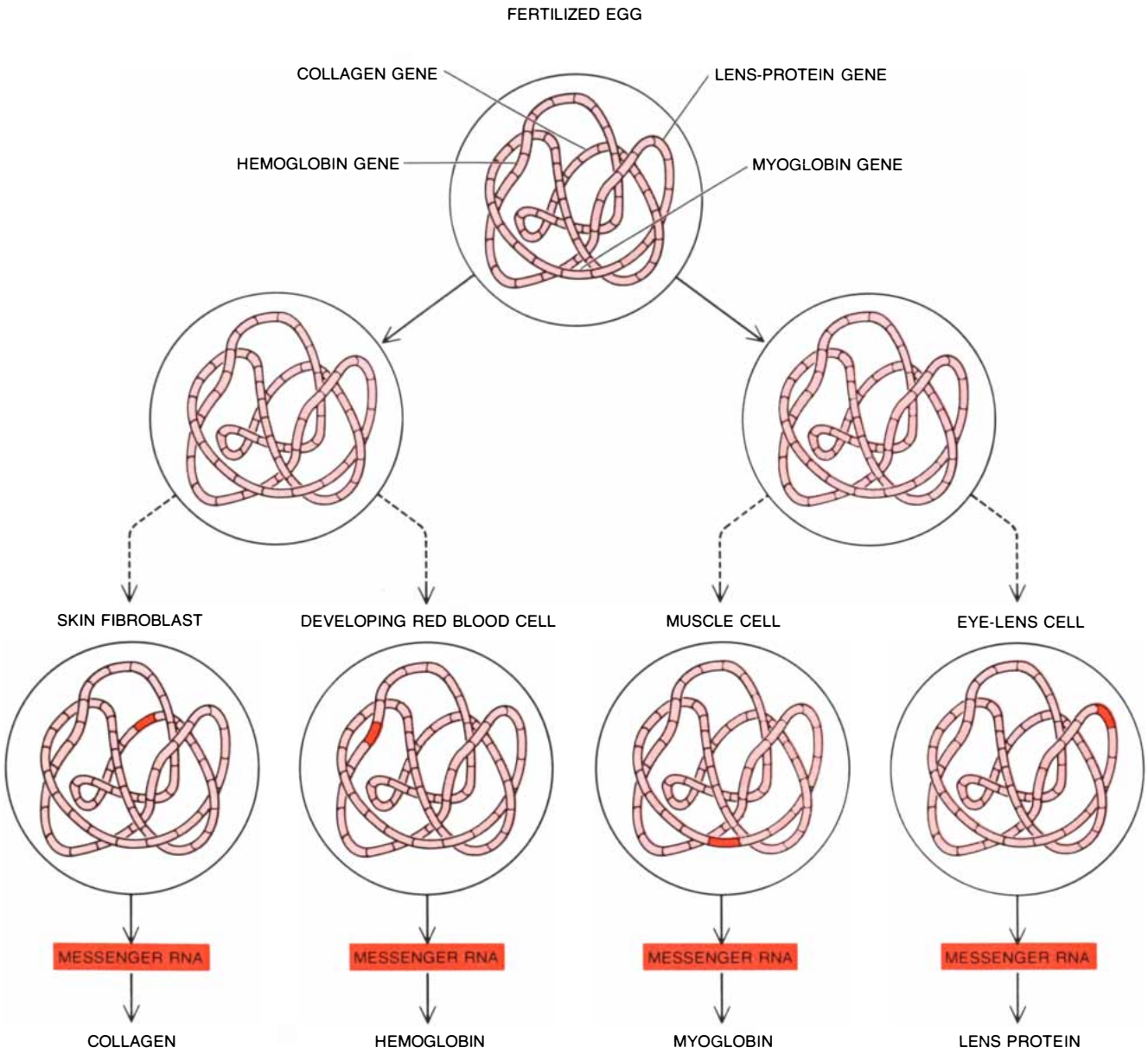
In bacteria, which lack a nucleus and where the genome is considerably less complicated than it is in higher organisms, substantial progress has been made toward elucidating the regulatory mechanism. In the colon bacillus *Escherichia coli* the transcription of individual genetic sequences into RNA has been shown to be regulated by the binding and release of specific regulators at appropriate sites on the DNA. In the cells of higher plants and animals the genetic material is more complex both structurally and functionally. The DNA is contained within a membrane-bound nucleus. And it is packaged there with proteins and with a small amount of RNA to form the complex substance called chromatin or, in its most organized form, the chromosomes. No specific regulatory proteins have yet been identified in such nucleated, or eukaryotic, cells, but the evidence does implicate the chromosomal proteins as being the regulatory elements.

The Histones

Historically these proteins have been divided into two groups, the histones and the nonhistone proteins. The histones are best defined by their chemical characteristics: they are proteins with a positive charge, enriched in the basic amino acids arginine and lysine and completely lacking in the amino acid tryptophan. Because of their basic nature they can readily be extracted with dilute hydrochloric or sulfuric acid. They

can be fractionated into five principal classes, most of which are present in all eukaryotic cells. Histones were discovered in the late 19th century, but it was not until 1943 that Edgar and Ellen Stedman, who were then working at the University of Edinburgh, drew attention to them as possible regulatory molecules. The Stedmans' data suggested that actively growing tissues contained less histone than nongrowing tissues, and so they concluded that histones function as repressors, or inhibitors of biological activity. They were correct in postulating an inhibitory function for the histones, but their conclusion was based on measurements of histone content that have since been shown to be inaccurate; today we know that the histone content of active and inactive cells is the same.

The first definitive biochemical studies of the effects of histones on DNA function were undertaken in the early 1960's after the development of cell-free systems for the synthesis of RNA. In such systems DNA or chromatin is provided to serve as a template for the transcription of RNA in the test tube. Other components of the reaction mixture are the precursors of the RNA subunits and RNA polymerase, the enzyme that catalyzes the polymerization of the subunits into the strands of RNA. In 1962 Ru-chih Huang and James Bonner of the California Institute of Technology demonstrated that adding histones to such a cell-free system inhibited RNA synthesis. (Maximal inhibition of RNA synthesis was observed at a histone-to-DNA ratio of 1 : 1, which is about the ratio in which histone and DNA are normally found in the nucleus.) The inhibition was reversed by the addition of more DNA, demonstrating that the histones blocked transcription by binding to DNA rather than by inhibiting the RNA polymerase. At about the same time Vincent G. Allfrey and Alfred E. Mirsky of Rockefeller University showed that the selective removal of histones from isolated nuclei stepped up the rate



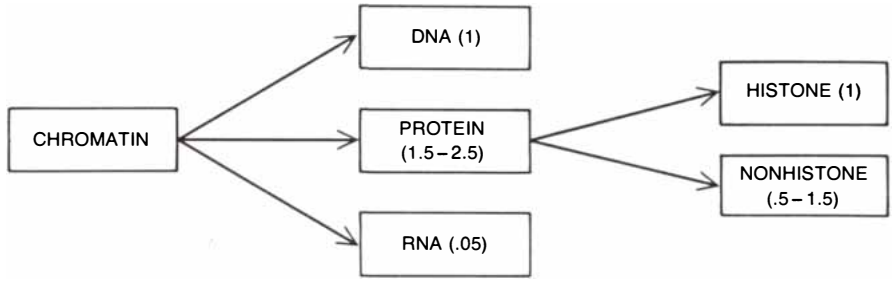
GENE REGULATION is exemplified in the course of development. The fertilized egg (*top*) contains all the organism's genetic information encoded in DNA; there are genes for the four special proteins indicated here and for thousands of others. The egg divides and cells proliferate and differentiate (*broken arrows*) into specialized cell types in various tissues. Each differentiated cell type still contains all the organism's genes, but the genes are regulated so that only those genes in each cell that are appropriate for the cell's specific functions are "turned on": transcribed into messenger RNA and subsequently translated into protein (*bottom*).

of RNA synthesis. Both types of experiment pointed to the conclusion that histones normally function to inhibit the capacity of genes to be transcribed into RNA.

It soon became apparent that histones are also involved in determining the structural properties of the chromatin. In particular the addition of histones increases the stability of the DNA double helix. That fact provides a clue to a biological function of the histones, because a tightening up of the helix could affect the availability, or exposure, of nucleotide sequences for transcription into RNA.

Another feature of the metabolism of histones that suggests their biological role is the relation between histone synthesis and DNA replication. Before a cell divides into two daughter cells all its DNA must be duplicated so that each daughter cell can receive a complete copy of the genetic information. DNA synthesis is confined to a defined period of the cell's life cycle, the S phase. It is tightly coupled with histone synthesis. Thaddeus Borun, Elliot Robbins and Matthew D. Scharff of the Albert Einstein College of Medicine and Gerald C. Mueller of the University of Wisconsin demonstrated that histone and DNA are synthesized simultaneously and that inhibition of DNA replication is accompanied by an immediate shutdown of histone synthesis. The tight coupling of DNA synthesis and histone synthesis is further supported by experiments indicating that the messenger RNA's that code for histones are only associated with the cell's protein-synthesizing organelles and translated into proteins at the time of DNA synthesis. (It makes sense that if only limited regions of the genome are to be transcribed in a given cell at a given time, histones must be available immediately to repress newly replicated segments of the DNA that are to remain unexpressed.)

Apart from the synthesis of new histone molecules, the interaction of histones with DNA can be altered by the modification of previously existing histones. There are enzymatic reactions that modify histone structure through the addition of acetate groups, methyl groups and phosphate groups. The exact significance of these reactions is not yet known, but histone acetylation has been shown by Allfrey and others to be correlated with the general level of RNA synthesis; the phosphorylation of certain histones, G. Roger Chalkley of the University of Iowa has observed, is correlated with the rate of cell division. Such reactions may be important in controlling the overall activity and structure of the genome.

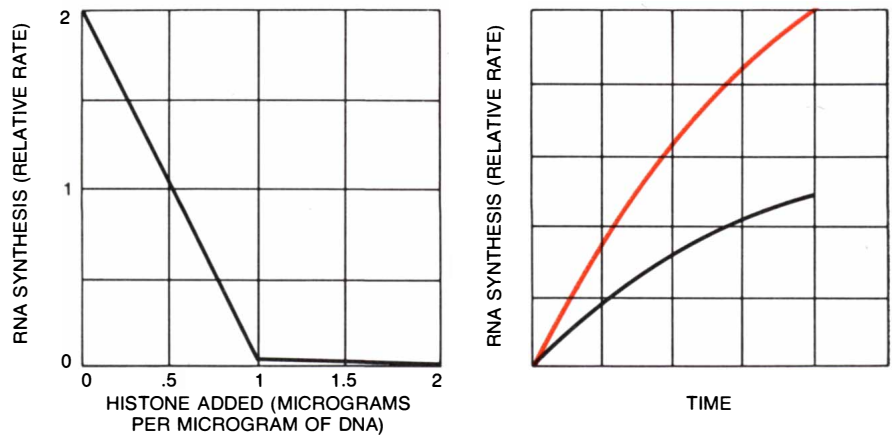


CHROMATIN, the complex material in the nucleus of cells of higher organisms, is made up of DNA, protein and a small amount of RNA. There are two categories of protein: histones and nonhistone proteins. The numbers indicate the relative amounts of each component of chromatin. There is about as much histone as DNA but nonhistone ratio varies.

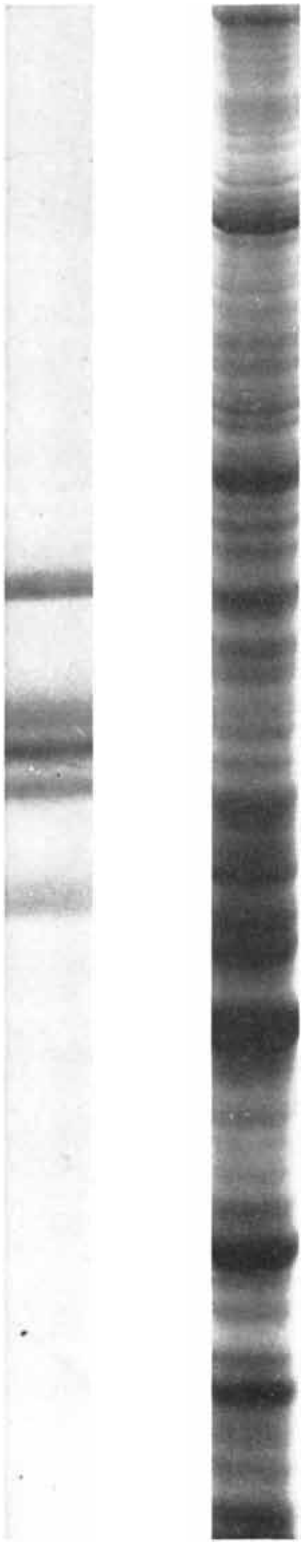
All in all, it is clear that histones are involved in the maintenance of chromatin structure as well as in the repression of DNA-dependent RNA synthesis. Nevertheless, several lines of evidence suggest that histones exhibit a uniformity and a consequent lack of specificity that preclude their ability to recognize and influence particular genes. Similar amounts of histone are present in active and in inactive tissues, and the amount in a cell remains constant as the cell's metabolism varies. Similar amounts of histone are also present in active and in inactive regions of the genome in a given cell, that is, in "extended" chromatin that is actively synthesizing RNA and in "condensed" chromatin, which is not synthesizing RNA. In the nuclei of certain cells of insect larvae the regions of the genome that are in the process of synthesizing RNA are extended in comparatively huge "puffs." Hewson Swift of the University of Chicago measured the ratios of histone to DNA in the inactive and the puffed regions and found they were identical. Perhaps the most striking evidence for the lack of histone specificity is that the amino acid sequence of one histone fraction has been compared in two very different kinds of cells, those of pea seedlings and of calf thymus, and they turn out to differ by only two amino acids. Histones, it seems clear, are regulatory molecules involved in the control of gene transcription, but they must be regulatory in a nonspecific sense.

Nonhistone Proteins

In the search for molecules with the capacity for the regulation of specific genes attention has recently been focused on the nonhistone proteins of the chromosomes, which are defined simply as the proteins other than histones that are associated with chromatin. Like the histones, they are synthesized outside the nucleus in the cytoplasm. The histones are transported into the nucleus and are associated with DNA immediately after synthesis. Studies in our laboratories at the University of Florida



INHIBITING EFFECT of histones on gene activity was first demonstrated in the 1960's. Ru-chih Huang and James Bonner of the California Institute of Technology found (*left*) that histones inhibited RNA synthesis in cell-free systems. Vincent G. Allfrey and Alfred E. Mirsky of Rockefeller University showed (*right*) that the rate of RNA synthesis was greater in nuclei from which histones had been removed (*color*) than in control nuclei (*black*).



HISTONE AND NONHISTONE proteins isolated from chromatin were subjected to electrophoresis in polyacrylamide-gel columns, in which individual proteins are separated because they migrate at different speeds in an electric field. The histones (*left*) fall into just five clearly defined groups. The nonhistone-protein fraction (*right*) is far more heterogeneous than that. Amino acid analysis indicates that the histones are richer in the most basic amino acids, whereas nonhistone proteins have a high content of the most acidic amino acids.

College of Medicine and at the University of Michigan have shown that for the nonhistone proteins the picture is somewhat more complex. Some of the nonhistone proteins become associated with DNA right after synthesis, but various periods of time elapse before others appear as components of the chromatin. As a group the nonhistone proteins exhibit a faster rate of turnover than the histones, with the lifetime of individual nonhistone fractions varying considerably; some have a lifetime of only several minutes and others are at least as stable as cellular DNA and histones. It is not clear whether, on leaving their association with DNA, the nonhistone proteins are degraded or enter a pool of proteins in the nucleus or the cytoplasm. What is clear is that whereas histones and DNA appear to constitute permanent components of the genome, the nonhistone chromosomal proteins are at least partly in a state of dynamic flux.

In contrast to the histones, the nonhistone proteins exhibit a tremendous amount of heterogeneity, structurally as well as functionally [*see illustration at left*]. They range in molecular weight from under 10,000 daltons to over 150,000. It is difficult to determine accurately how many species of nonhistone proteins are associated with the genome, but several lines of evidence suggest that the number is extremely large. One indication of their functional diversity is the variety of complex enzyme systems found among them. There are enzymes such as the polymerases involved in RNA and DNA synthesis and repair, enzymes involved in the processing and degradation of proteins and enzymes that modify nucleic acids and proteins by adding or removing acetate, methyl and phosphate groups. In addition to these enzymatic components the nonhistone fraction is also thought to contain proteins that have structural as well as regulatory roles in chromatin. In the remainder of this article, however, we shall focus on the major lines of evidence supporting the idea that at least some of the nonhistone proteins are involved in regulating the activity of specific genes.

There are significant differences between the types of nonhistone proteins found in the cells of different species. Since the genetic information encoded in the DNA is unique for each species, the observed species specificity of nonhistone proteins is consistent with a regulatory function. Even within the same organism there are different nonhistone proteins in different tissues. Although all cells of the organism contain the same DNA, variation in the complement of nonhistone proteins in the various cell types is consistent with the selective utilization of genetic information by those cell types.

If nonhistone proteins are engaged in specific gene regulation, then the spectrum of these proteins in a cell should change when gene activity changes. This prediction has now been tested in a large number of experimental systems where variations in gene expression are known to occur, and in almost all cases the predicted changes in nonhistone proteins have been observed.

Variations in Synthesis

Variations in gene expression occur throughout the cell cycle, for example. The most pronounced differences in transcription are observed between the S phase (the period of DNA synthesis) and mitosis (when the cell divides). In our laboratories and others the nonhistone chromosomal proteins associated with the genome during each phase of the cell cycle have been examined with respect to composition and metabolism, and similarly pronounced differences have been observed. To take another example, the steroid hormones, such as estrogen and hydrocortisone, are known to exert effects on target cells by binding in the nucleus and ultimately altering the pattern of gene transcription. It has now been shown that several steroid hormones cause changes in the composition and metabolism of nonhistone chromosomal proteins.

Mouse neuroblastoma cells (nerve-cancer cells) in tissue culture provide an interesting model system for studying differentiation. They are readily induced to differentiate into cells with the structural, biochemical and electrical properties characteristic of nerve cells. One of the system's unique advantages is that the induced differentiation is completely reversible. Michelle Zornetzer at our laboratory in Florida has observed that in the differentiated state there is a selective decrease in the synthesis of nonhistone proteins of high molecular weight and that this phenomenon is reversed when the state of differentiation is reversed [*see illustration on page 52*]. She also showed that the differentiation of such cells is associated with differences in RNA synthesis. That is, there is a strong correlation between the types of nonhistone proteins associated with the genome and the alterations in gene expression that accompany differentiation.

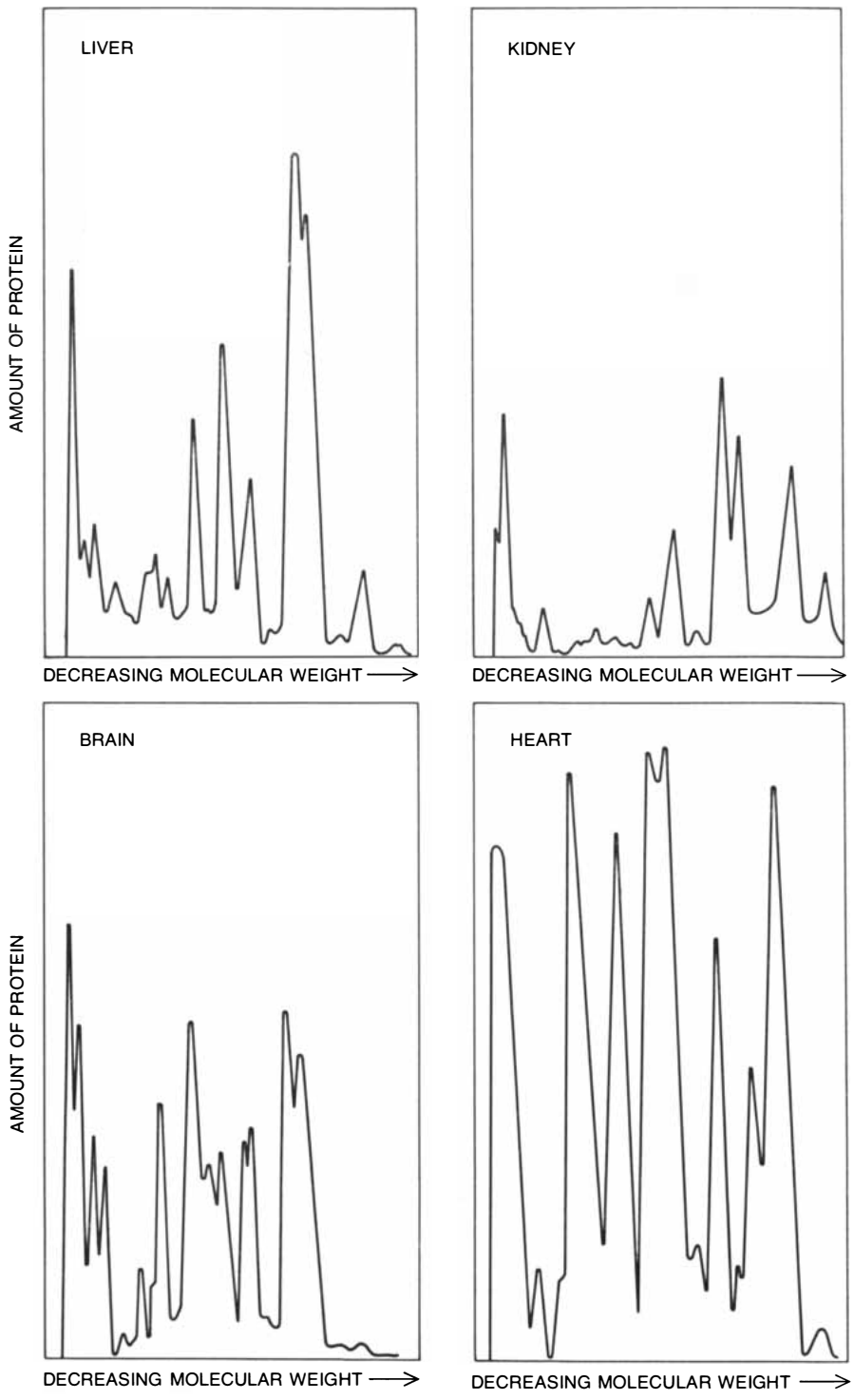
It is possible to separate chromatin into two fractions that differ from each other in both structure and genetic ac-

tivity: the condensed chromatin fraction, which is structurally compact and relatively inactive in RNA synthesis, and the extended chromatin fraction, which actively synthesizes RNA. As we have mentioned, the ratio of histone to DNA is the same in these two types of chromatin. The relative amount of nonhistone protein is considerably higher, however, in the extended, active form of the chromatin. Here again the presence of nonhistone proteins is correlated with a structural state of chromatin characterized by active RNA synthesis.

Changes in the nonhistones also accompany the modifications in the structural and biochemical properties of a cell infected and transformed by viruses that cause cancer. Transformed cells exhibit the lack of control of growth that is characteristic of malignancy. Recent results from our laboratories indicate that almost immediately following the infection of cells with tumor viruses there are changes in the composition and metabolism of nonhistone chromosomal proteins. These modifications become permanent in cells that are transformed and rendered malignant. So far there is no definitive evidence for a cause-and-effect relation, but it seems reasonable to conclude that nonhistone proteins could have a primary function in mediating the virus-induced modifications in gene expression that are associated with malignancy.

Still another distinct property of nonhistone chromosomal proteins supporting the concept that they are involved in specific gene regulation is their ability to bind DNA. In bacterial cells in which specific regulatory proteins have been identified and isolated it has turned out that they function by binding to specific sequences in appropriate regions of the DNA. In a similar fashion specific regulatory molecules in eukaryotic cells might be expected to bind to particular types of DNA. In 1970 experiments performed in one of our laboratories (Kleinsmith's) first demonstrated that some of the nonhistone proteins do specifically bind to DNA.

These experiments employed the technique of DNA-cellulose chromatography developed by Bruce M. Alberts of Princeton University. In this procedure DNA is adsorbed onto a cellulose matrix that is packed into a glass column, and the proteins under study are passed through the column. Proteins that bind to DNA become lodged in the column and can be removed for analysis. When nonhistone proteins prepared from rat liver are passed through such a column made from rat DNA, a small percentage of the proteins stick. When these DNA-binding proteins are removed from the column and passed through a new column made from a bacterial DNA, most of them pass right through [*see illustration on page 53*]. In other words, some of the nonhistone proteins must be able to recognize specific nucleotide sequences in DNA, so that they bind to their host DNA but not to a foreign DNA. Such a specific interaction between these proteins and DNA is exactly what would be expected if some of them are involved in regulating the activity of specific genes. This specific binding of DNA and nonhistone proteins has been



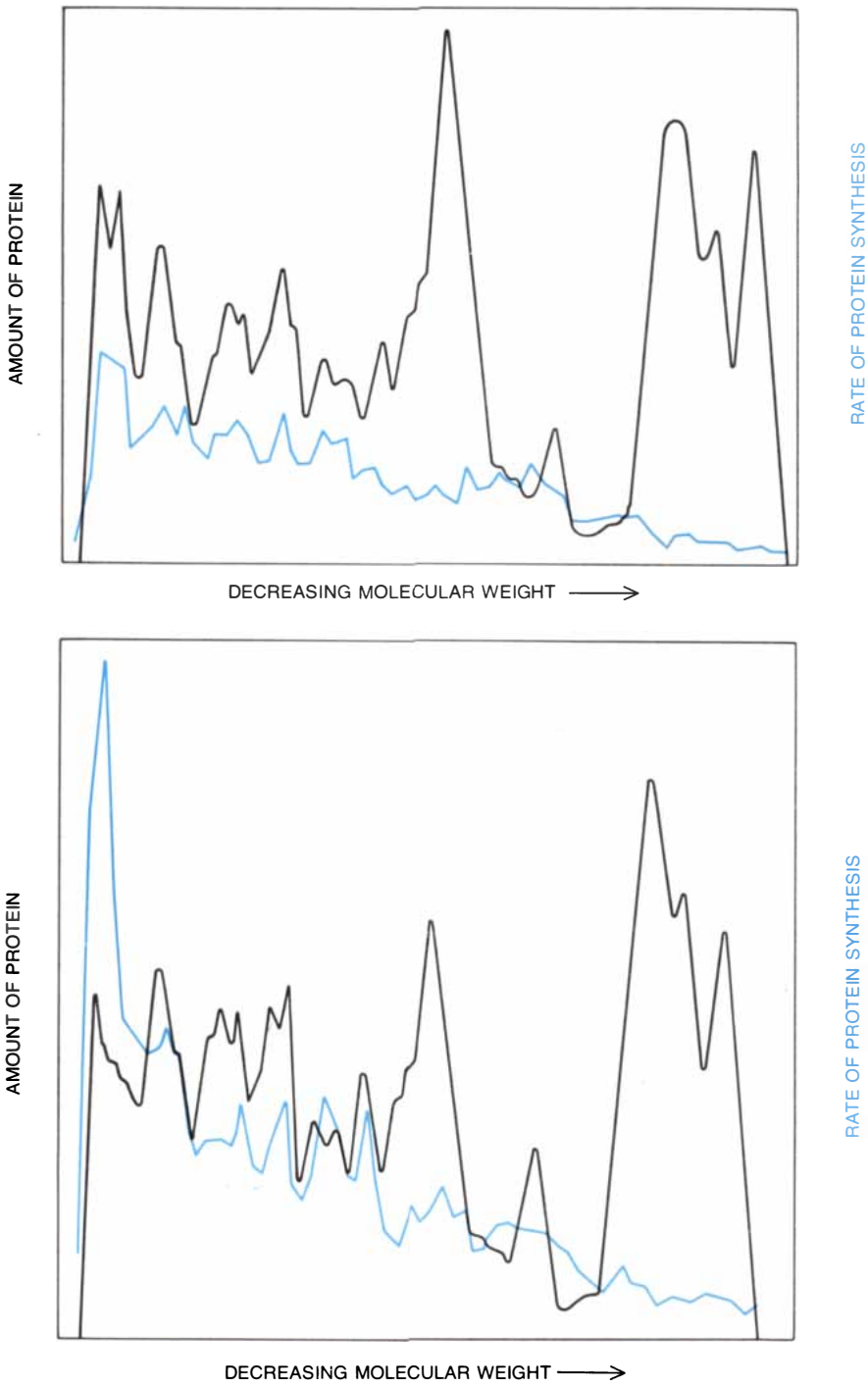
DIFFERENT SPECTRUM of nonhistone proteins is present in different tissues. Each curve records fluctuations in the optical density, and therefore in the amount of protein, along the length of a gel such as the one at the right in the illustration on the opposite page. The pattern of peaks, and thus of individual nonhistone proteins, is different for gels containing nonhistone proteins extracted from four different specialized tissues of the rat.

substantiated by observations in several other laboratories, including those of Allfrey, Bonner and of Lubomir S. Hnilica of the University of Texas at Houston.

Experimental Proof

In order to determine directly whether the presence of nonhistone proteins can alter gene transcription it is necessary to utilize systems in which the effects of nonhistone proteins on RNA synthesis can be measured precisely. In 1968 a report from the laboratory of Tung-Yue Wang at the State University of New York at Buffalo indicated that adding nonhistone proteins to a cell-free system containing purified DNA and RNA polymerase could reverse the inhibition of RNA synthesis that is normally produced by the presence of histones. These results were consistent with an earlier report by Thomas A. Langan, who was then working at Rockefeller University, that when histone is complexed with nonhistone protein, the histone exerts only a partial inhibitory effect on RNA synthesis. Similar approaches were employed by Ching-Sung Teng and Terrell H. Hamilton of the University of Texas at Austin and by Thomas C. Spelsberg and Hnilica in Houston. Both groups found, in different systems, that nonhistone chromosomal proteins prevent the complete inhibition of RNA synthesis by histones. Wang's group went on to demonstrate that the augmented transcription produced by the addition of nonhistone proteins represented the activation of regions of the genome that had previously been suppressed. Taken together, these studies clearly demonstrated that nonhistone chromosomal proteins can influence transcription. Still, definitive interpretation of the results was difficult because in the cell-free systems the components of chromatin were mixed together under conditions that may not allow them to become associated as they are in the nucleus of an intact cell.

The most direct evidence that nonhistone chromosomal proteins play a role in the regulation of tissue-specific transcription comes from chromatin reconstitution experiments, in which the various components of chromatin are taken apart and then put back together in various combinations. In pioneering experiments along these lines R. Stewart Gilmour and John Paul of the Beatson Institute for Cancer Research in Glasgow isolated chromatins from rabbit thymus and rabbit bone marrow and dissociated them into their DNA, histone and nonhistone components. Then they reconstituted the chromatins by a method developed by Huang and Bonner. In the control experiment the DNA, histones and nonhistone proteins from rabbit thymus were simply put back together again. The reconstituted chromatin was then allowed to serve as a template for RNA synthesis in the presence of RNA polymerase. The RNA that was transcribed behaved, in tests of its ability to bind to DNA from rabbit thymus or bone marrow, just like normally synthesized thymus RNA. Then Gilmour and Paul pooled the DNA and histones from thymus and bone marrow and reconstituted the chromatin by adding thymus nonhistone proteins to the pool. The RNA made from this hybrid chromatin

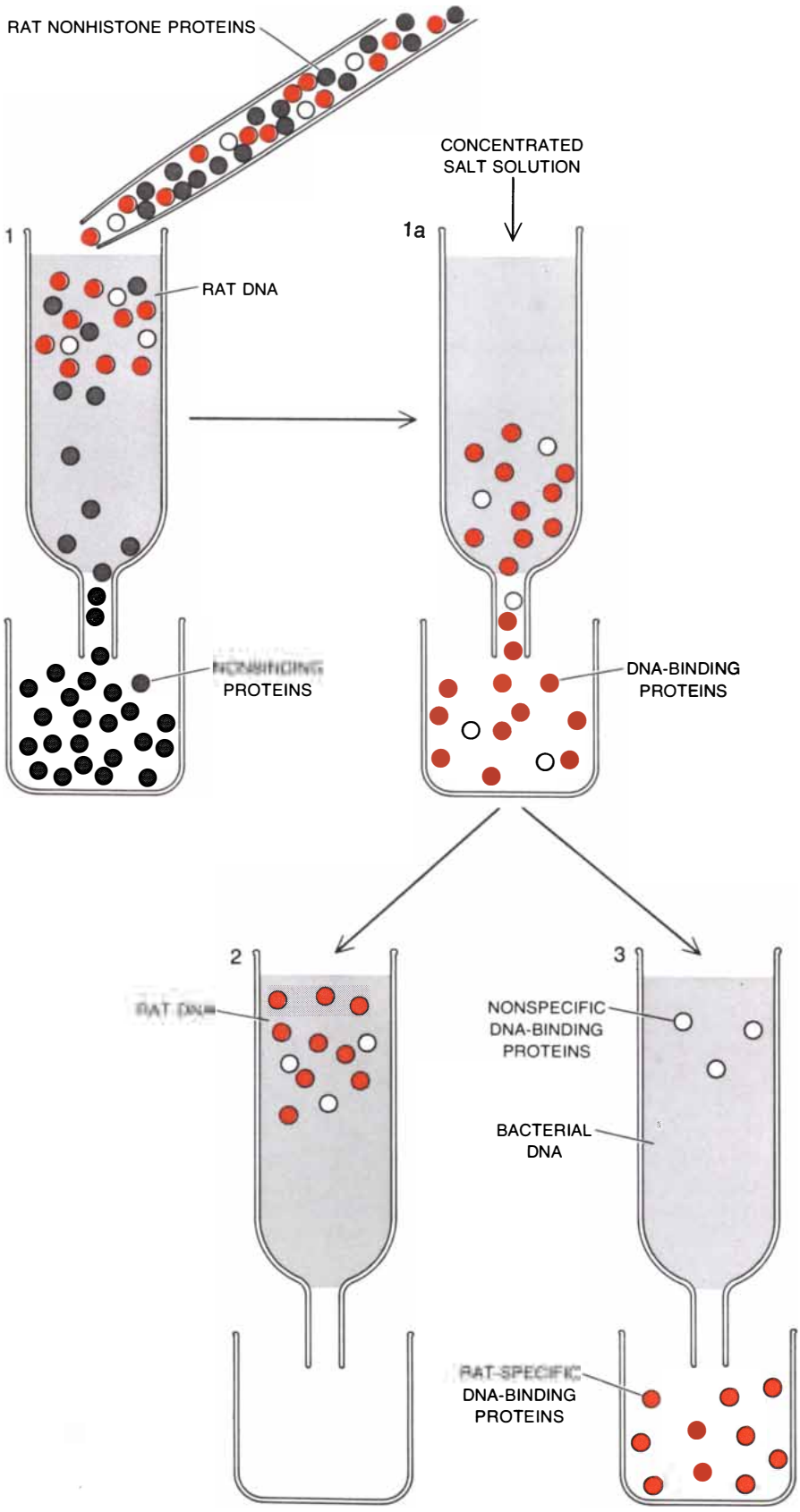


RATE OF SYNTHESIS of certain nonhistone proteins, and consequently the amount of each one of them, varies in the same kind of cell at different stages of differentiation. The colored curves in these two graphs show the rates of synthesis, and the black curves show the amounts, of nonhistone proteins that are found in differentiated mouse nerve-cancer cells (*graph at top*) and in nondifferentiated ones (*graph at bottom*). There is a selective decrease in rate of synthesis of high-molecular-weight proteins in the differentiated state.

again behaved like thymus RNA—even though the chromatin included DNA and histones from bone marrow. On the other hand, when the pooled DNA and histones were reconstituted with nonhistone proteins from bone marrow, the RNA made from the hybrid chromatin behaved like bone-marrow RNA [*see illustration on next page*]. The experiments indicated strongly that it is the presence of tissue-specific nonhistone proteins that determines just which genes will be transcribed in various tissues.

How about the same tissues at different times? Are nonhistone proteins also responsible for the transient changes in transcriptional activity that are associated with changes in the metabolic state of a single cell type? The cell cycle provides an ideal model system for directly investigating the influence of nonhistones on the transcriptional activity of the genome, since chromatin isolated from S-phase cells (the period of active DNA replication) has greater activity in RNA synthesis than chromatin isolated from cells in mitosis. Chromatin-reconstitution experiments conducted by one of us (Gary Stein) and John Farber at Temple University demonstrated that it is the nonhistone proteins that are responsible for this modulation of gene transcription during the cell cycle. Chromatin reconstituted from pooled DNA and histones with nonhistone proteins from mitotic cells exhibited a lower capacity for RNA synthesis than chromatin reconstituted in the presence of S-phase nonhistone proteins [*see illustration on page 55*]. To eliminate the possibility that histones are involved in these alterations in gene transcription, reconstitution experiments were also performed in the presence of either S-phase or mitotic histones. In this case the transcriptional activities of the two types of reconstituted chromatins were identical, showing that histones do not dictate differences in the availability of DNA sequences for RNA synthesis during the cell cycle.

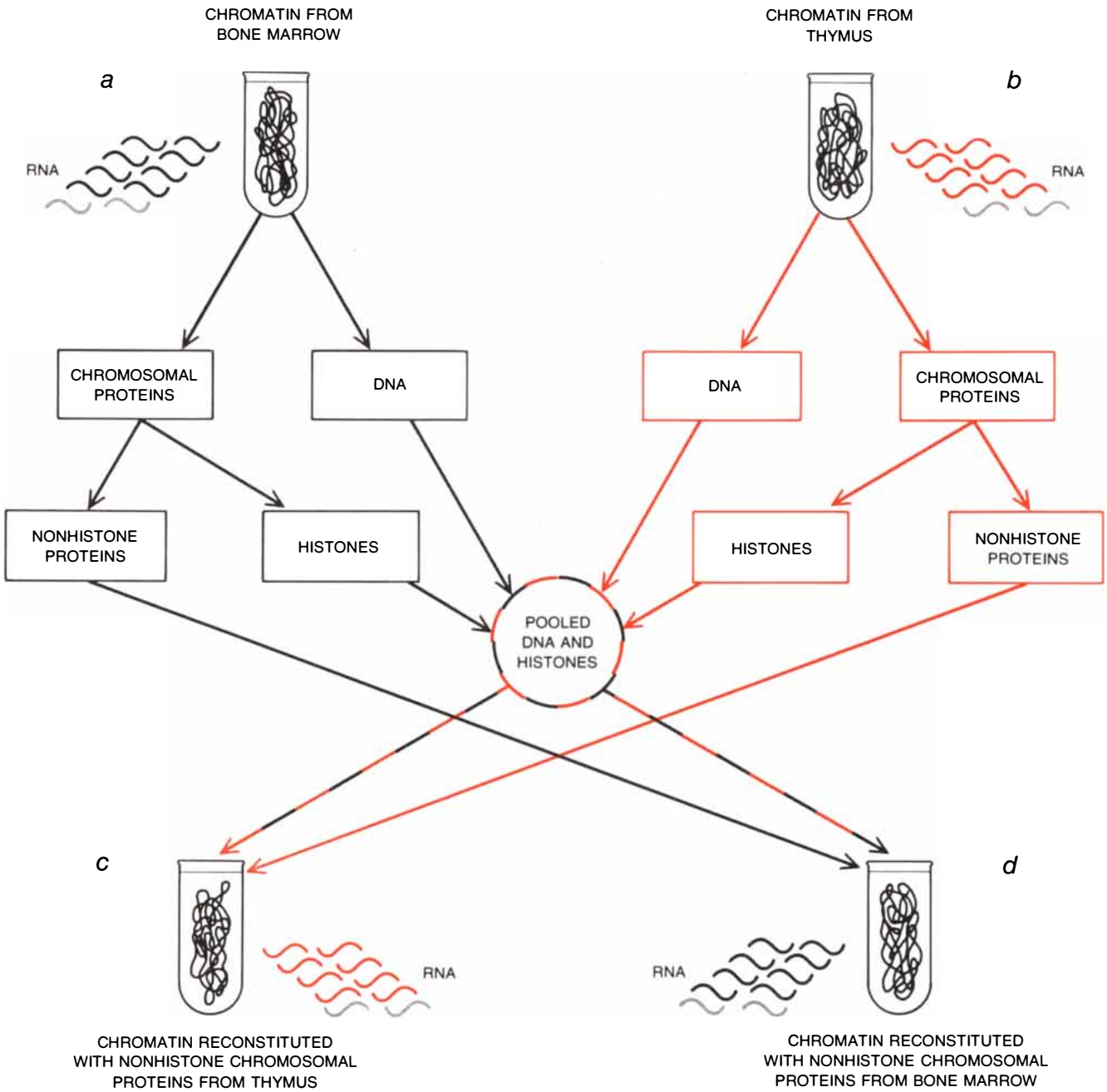
To demonstrate clearly that nonhistone proteins regulate the activity of individual genes one needs to test for the transcription of a specific gene. In the past few years it has become possible to synthesize in the laboratory radioactive copies of individual gene sequences by working "backward" from a known RNA to its complementary DNA. These radioactive gene copies will specifically bind to the complementary RNA's for which they code, and so they can be used as sensitive probes for monitoring the presence of RNA's that have been tran-



SPECIFIC BINDING of nonhistone proteins to a DNA was demonstrated by passing nonhistone proteins from rat liver through a column of cellulose to which rat DNA had been adsorbed. Some of the proteins stayed in the column, bound to DNA; the nonbinding proteins (*gray*) emerged (*1*). The bound proteins (*color and white*) were dislodged with a concentrated salt solution (*1a*). These DNA-binding proteins were then passed through columns containing rat DNA or DNA from the bacterium *Escherichia coli*. All became bound to rat DNA (*2*) but only a small proportion were bound to *E. coli* DNA (*3*). The ones that bound only to rat DNA but not to *E. coli* DNA were proteins that are specific for rat DNA.

scribed from particular genes [*see top illustration on page 56*].

Gilmour and Paul recently employed such sophisticated probes in an exciting series of chromatin-reconstitution experiments. Chromatin was isolated from fetal mouse liver, which normally synthesizes globin (the protein of hemoglobin), and from mouse brain, which does not make globin. Each chromatin preparation was transcribed in a cell-free system, and the RNA's synthesized were characterized by testing their ability to bind to radioactive copies of globin genes. In keeping with the pattern of transcription observed in intact cells, fetal-liver chromatin synthesized RNA coding for globin but brain chromatin did not. Gilmour and Paul then proceeded to dissociate the chromatins and reconstitute them in the presence of various types of nonhistone proteins. They found that when brain chromatin was reconstituted in the presence of nonhistone proteins from brain tissue, it would not synthesize globin RNA, but that when the reconstitution was performed in the presence of nonhistone proteins from fetal liver, the globin gene was transcribed. Such results strongly support the concept that it is the presence of specific nonhistone proteins that determines that the globin gene is transcribed



CHROMATIN-RECONSTITUTION EXPERIMENT conducted by R. Stewart Gilmour and John Paul of the Beatson Institute for Cancer Research showed that nonhistone proteins are responsible for tissue-specific transcription. Chromatin was isolated from rabbit thymus and bone marrow. Each chromatin made a characteristic RNA (*a*, *b*). When each chromatin was dissociated into DNA, histone and nonhistone fractions and reconstituted, it made the same RNA. Then the chromatins were dissociated and the DNA and histones from the two chromatins were pooled. When the pooled DNA and histones were combined with nonhistone proteins from thymus, the reconstituted chromatin made RNA that behaved like thymus RNA (*c*). When pooled DNA and histones were combined with bone-marrow nonhistone proteins, the reconstituted chromatin made RNA that behaved like bone-marrow RNA (*d*).

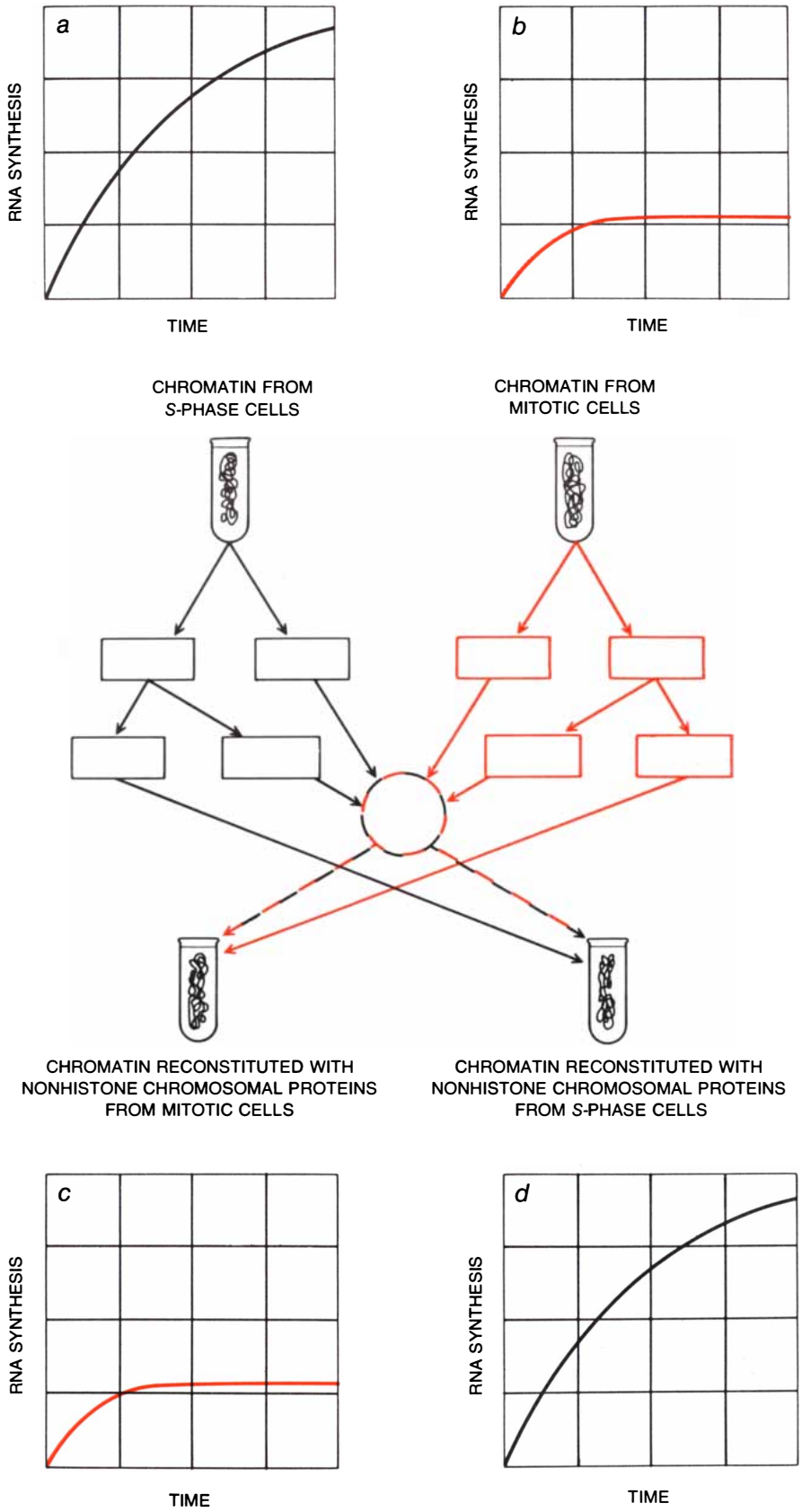
in fetal-liver tissue but not in brain tissue.

With a similar approach our group in Florida has recently demonstrated that nonhistone chromosomal proteins are responsible for restricting the transcription of histone genes to the S phase of the cell cycle [*see bottom illustration on next page*]. Using a radioactive histone-DNA probe to examine RNA made in the test tube, we first demonstrated that chromatin from S-phase cells transcribed histone RNA's whereas chromatin from cells in the G_1 phase of the cell cycle, the period preceding DNA replication, did not. This finding was consistent with the earlier evidence for the coupling of histone synthesis and DNA replication. We then showed that chromatin reconstituted with S-phase nonhistone proteins transcribed the RNA's for histones to the same extent as native S-phase chromatin. In contrast, chromatin reconstituted with nonhistone proteins isolated from G_1 chromatin showed no detectable transcription of histone RNA's. What this means is that nonhistone chromosomal proteins dictate the elaboration of information from specific genes during defined periods of the cell cycle.

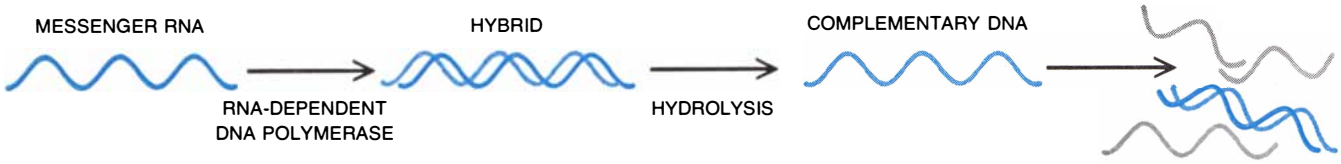
A Possible Mechanism

Assuming that nonhistone chromosomal proteins do play a key role in regulating the transcription of individual genes, both in different specialized cells and at different times in individual cell types, how might such control be exerted? What is the molecular mechanism? One striking property of the nonhistone proteins that may offer a clue is their extensive phosphorylation. In a series of experiments begun in 1966 one of us (Kleinsmith), Allfrey and Mirsky observed that nonhistone proteins are continually being modified in the nucleus by the addition and removal of phosphate groups. We speculated that because phosphate groups are negatively charged such modifications might alter both the structure and the functional interactions of the nonhistone proteins in chromatin. Since these early observations were made workers in many laboratories, studying different biological systems, have noted that phosphorylation of nonhistone proteins is associated with cellular differentiation and the activation of genes. Moreover, the ability of nonhistone proteins to stimulate RNA synthesis in cell-free systems has been shown to be dependent on their state of phosphorylation.

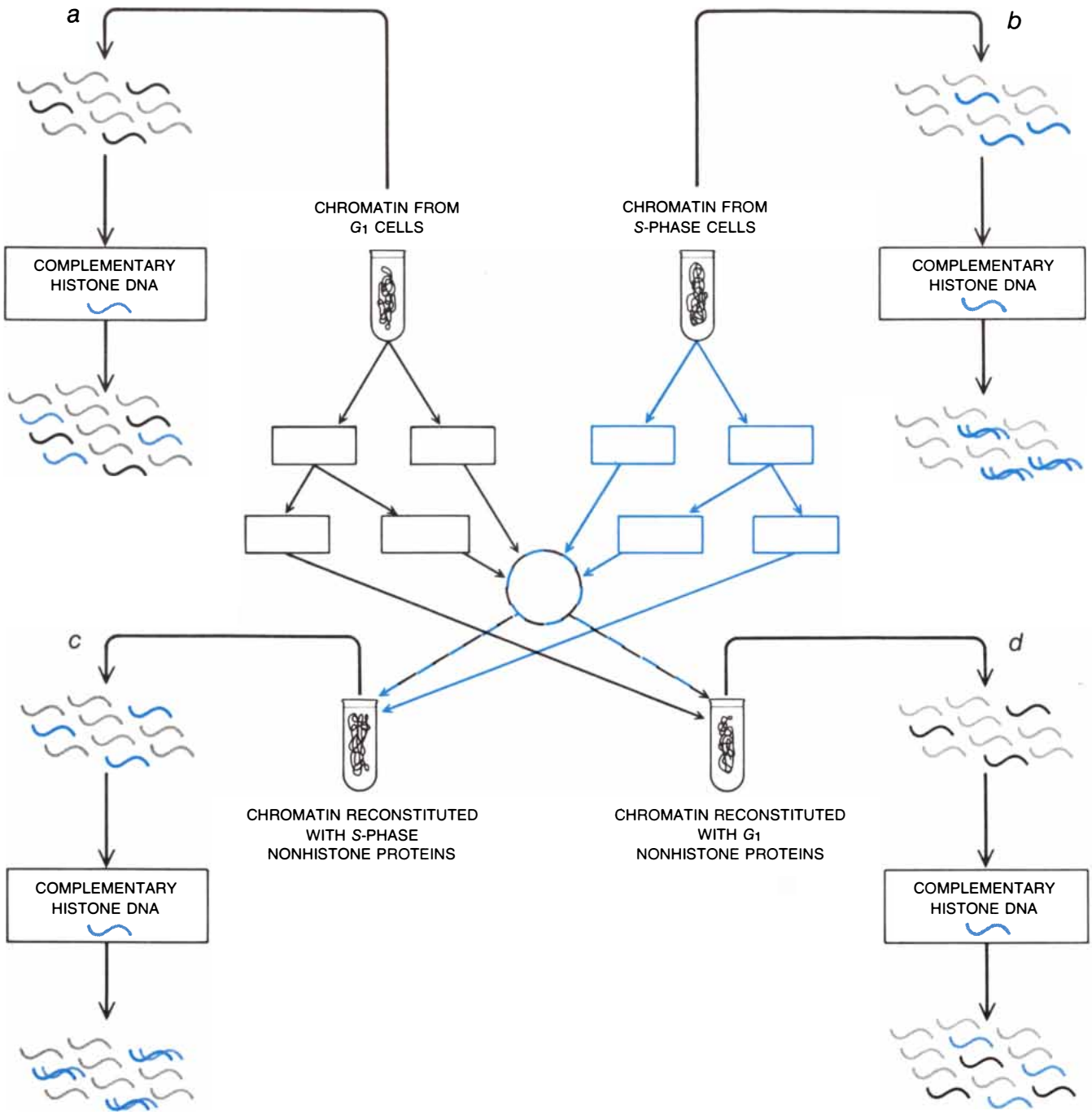
Such observations suggest that the



ANOTHER RECONSTITUTION EXPERIMENT showed that nonhistone proteins dictate variations in transcription at different stages of the cell cycle. Chromatin isolated from cells in the *S* phase synthesizes RNA faster (*a*) than chromatin isolated during mitosis (*b*). The chromatins were dissociated and the DNA and histones were pooled as in the experiment illustrated on the opposite page. Chromatin reconstituted with nonhistone proteins from mitotic cells synthesized RNA at the rate characteristic of mitotic-cell chromatin (*c*). Chromatin reconstituted with *S*-phase nonhistones made RNA at the faster *S*-phase rate (*d*).



HYBRIDIZATION EXPERIMENTS identify specific gene products. A known RNA (*dark color*) is isolated. In the presence of the enzyme RNA-dependent DNA polymerase ("reverse transcriptase") the RNA serves as a template for the assembly of a complementary strand of DNA (*light color*) that can be labeled through the incorporation of a radioactive subunit. The hybrid RNA-DNA is dissociated by hydrolysis, leaving a radioactive complementary DNA, in effect a synthesized gene. It can serve as a probe for identifying the RNA from which it was made: when it is added to a solution of RNA's, it will hybridize only with that RNA (*right*).



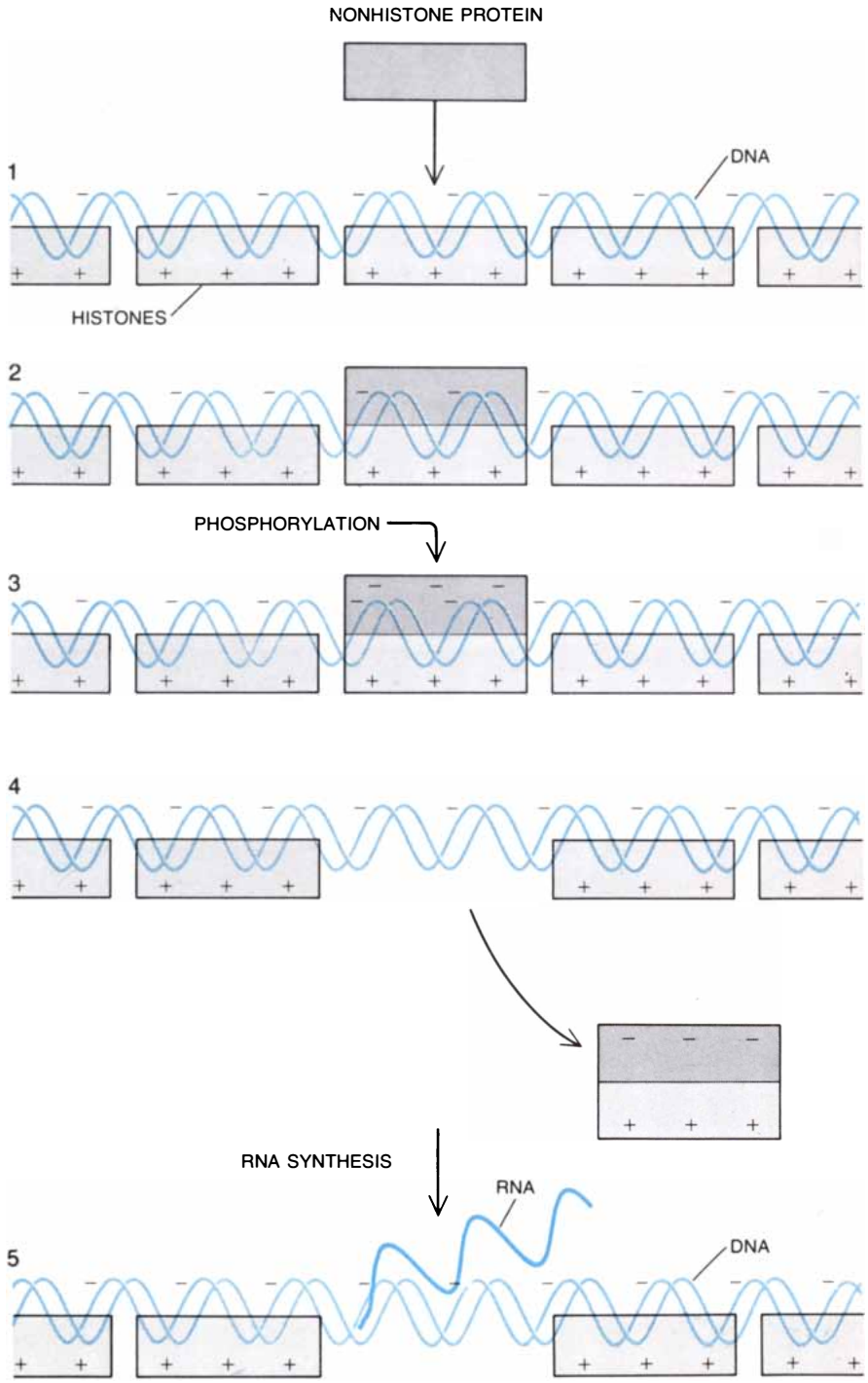
RECONSTITUTION AND HYBRIDIZATION were combined in this experiment. A radioactive DNA complementary to histone RNA was prepared and then used as a probe (*light color*) with which to demonstrate that whereas chromatin from G_1 cells does not make histone RNA (*a*), chromatin from S-phase cells does (*dark color, b*). Then the chromatins were dissociated, the histones and DNA were pooled and then reconstituted with two kinds of nonhistone proteins as in the earlier experiments. RNA's were prepared from the two reconstituted chromatins and were tested with the DNA probe. Chromatin reconstituted with S-phase nonhistone proteins synthesized histone RNA just as natural S-phase chromatin did (*c*). Chromatin reconstituted with G_1 nonhistone proteins did not (*d*).

phosphorylation of nonhistone proteins may somehow be involved in the mechanism by which these molecules regulate gene transcription. One proposed model invokes both the binding of specific DNA and phosphorylation [*see illustration on this page*]. We suggest that a specific nonhistone protein binds to a specific site on histone-repressed DNA. At that point the nonhistone protein becomes phosphorylated, and because phosphate groups are negatively charged the nonhistone protein begins to repel the negatively charged DNA and also becomes strongly associated with the positively charged histones. The effect of these two processes is to cause the displacement of the histone–nonhistone-protein complex from the DNA. The resulting naked area of DNA, now free of its inhibitory histone, can be transcribed into RNA.

Although this histone-displacement model is highly speculative, a number of observations are consistent with its general features. As we have mentioned, genes are more actively transcribed during the S phase than during mitosis. We have recently observed that nonhistone proteins are also more actively phosphorylated during the S phase than during mitosis, and also that histones bind less tightly to DNA during the S phase than they do during mitosis. Some new experiments on chromatin reconstitution carried out in our laboratory in Florida have shown that the nonhistone proteins associated with the genome during the S phase are specifically responsible for the less tenacious binding of histones to DNA during that period of the cell cycle. These results at the very least point to a close relation among gene activity, nonhistone-protein phosphorylation and the binding of histones to DNA.

The histone-displacement model is only one of many models proposed to account for the role of the chromosomal proteins in gene regulation. Although the detailed mechanism has not yet been clearly demonstrated, understanding of chromosomal proteins and their interactions with DNA is advancing rapidly. Certainly the histone and nonhistone proteins play important roles in determining the structural and functional properties of the genome. Where histones are involved in the maintenance of chromatin structure and in the nonspecific repression of genetic sequences, nonhistone chromosomal proteins appear to recognize defined gene loci and thereby regulate the transcription of specific genetic information.

It may not be long before proteins that regulate the expression of specific genes are isolated, introducing the possibility of a certain kind of genetic engineering: the proteins might be inserted into cells in order to modify abnormalities in gene transcription associated with development, differentiation and a broad spectrum of diseases, including cancer. Such a capability might revolutionize man's ability to deal with some profoundly destructive disorders. As in the case of any attempt at genetic engineering, however, the ethical questions raised by such an approach would be deep and complex, and would require careful consideration.



GENE IS SWITCHED ON, according to a proposed model, when a nonhistone protein recognizes (*1*) a specific site on a stretch of DNA that is generally repressed by histones. The nonhistone protein binds to the site (*2*); then it undergoes phosphorylation and thus becomes negatively charged (*3*). It therefore repels the negatively charged DNA and moves off with the positively charged histone, leaving the DNA site bare (*4*). The site, which is no longer repressed by histone, is available for transcription into RNA (*dark color, 5*).

This 300-year-old portrait looks well restored.

Portrait of Colonel Nels Assersen, painted in 1623 by Georg Günther Kräil de Bemeberg. Both photographs are enlarged from Polaroid Type 55 negatives.

These two prints from Polaroid instant negatives reveal a problem which must be solved in our lifetime. The price of failure will be the loss of many of our noblest paintings from the past.

That is the opinion of Dr. Björn Hallström, director of the Institute of Technology of Artistic Materials and head of the Art Conservation School of the Royal Swedish Academy of Fine Arts in Stockholm.

Dr. Hallström is specifically concerned about the long-term effects of certain early relining techniques on paintings. In these commonly employed restoration procedures, a painting threatened by flaking pigment or decomposing canvas has a new fabric pasted to its back. The painting may then appear quite revitalized. But beneath the "restored" surface, deterioration may actually accelerate.

Our Polaroid prints, for example, show a detail from a portrait painted in 1623. The painting was relined around 1930 and photographed some 40 years later. Under normal light (left) it looks fine. But the ultraviolet reflectogram (right) shows dark areas of decomposition in the relining paste. In time, such deterioration irreversibly alters a painting's color and characteristic craquelure.

Dr. Hallström suggests that the Polaroid instant photographs which reveal this problem can also help to solve it. Essentially, he proposes a standardized system of photoanalysis for all public collections. Each painting would be photographed using different kinds of radiation.

White light, flatly illuminating the canvas, gives a normal picture. *Raked white light* reveals surface damage, undulations in the canvas, etc. *Ultraviolet light with UV absorbing filters* on the camera shows fluorescence in the painting and indicates previous retouching. *Ultraviolet light plus a lens filter which*

Until you look beneath the surface.

Polaroid®

absorbs visible fluorescent light gives a UV reflectogram which can reveal decomposition invisible to the naked eye.

Polaroid Type 105 (3¼ x 4¼ pack format) and Type 55 (4 x 5 sheet format) positive/negative films are basic to Dr. Hallström's system for several reasons:

First, of course, they are *instant* films. This is particularly helpful in terms of UV reflectograms because some filters separate fluorescent areas better than others. If the Polaroid print shows poor separation, the filter can be changed, and the picture retaken on the spot.

Second, Type 105 and Type 55 provide an instant high resolution negative with the positive print. So enlargements can be made for closer scrutiny.

Finally, the superb tonal qualities of these films are especially useful in recording the low-brightness range of UV reflectograms.

Because of these advantages, and because Hallström's Polaroid technique is quick, economical and effective, it could be used to analyze all of our great paintings. And deterioration could often be detected and stopped before it went too far. Furthermore, periodic examination of newly relined paintings would indicate the restoration techniques most likely to truly conserve the art which we hold in trust for future generations.

The conservation of art treasures: A new use for Polaroid Instant Positive/Negative Films.

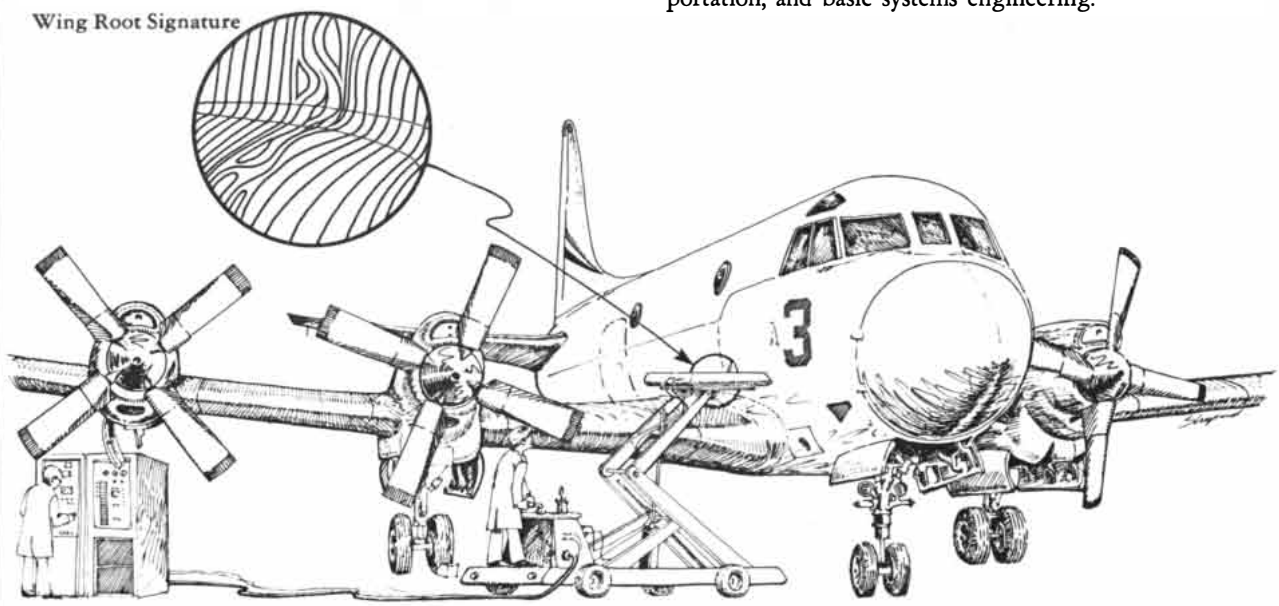
Conversation Pieces

Fault-Finding Without Tears

With today's huge high-speed aircraft, meticulously careful maintenance is essential to safety as well as to efficient operation. A vital element in every maintenance program is the kind of probing inspection that detects even invisible signs of corrosion, fatigue, and other early symptoms of deterioration in highly stressed structures.

This need has given rise to a whole new breed of test engineers. They use magnetism, high-frequency sound, penetrating dyes, and now the coherent light of laser beams to find the subtlest internal flaws before they become dangerous.

Under the innovative leadership of Dr. Pravin Bhuta, a TRW team has developed a system that uses holographic interferometry to reveal potential weaknesses in landing gear, wing panels, turbine blades, and other critical parts of aircraft. With the sponsorship of the U.S. Navy's Analytical Rework Program Office, the system has been successfully used in an ordinary maintenance environment.



The first tests were conducted in a TRW lab, however, where wing panels from a P-3 patrol pane were inspected. The prototype holographic systems not only found every flaw that had been previously located by conventional methods but also found several that had not been detected at all.

The next step was to do the same kind of job under workaday maintenance conditions *without* disassembling parts or removing paints or sealants. The completed system was taken to a Navy facility and the holographic equipment was mounted on a fork lift. It produced clear fringe patterns without external optics, whether it was pointed up, down, or sideways.

With this degree of mobility and flexibility, *in situ* inspection of critical parts becomes a practical reality not just for aircraft but for countless different kinds of structures. Compared with conventional methods, the saving in time alone is estimated to be as high as fifty percent.

When the technique has been fully developed, it will provide a cradle-to-grave record. Technicians will be able to compare the optical signature of the factory-new structure with later signatures, made during routine maintenance. Any significant differences will indicate the need for preventive repairs.

Dozens of promising ideas are under investigation at TRW, where we put the most advanced technology to work on the practical problems of defense, energy, transportation, and basic systems engineering.

For further information on the holographic interferometry system, write on your company letterhead to:

TRW®
SYSTEMS GROUP

Attention: Marketing Communications, E2/9043
One Space Park, Redondo Beach, California 90278

Dual-Resonance Models of Elementary Particles

In this new theoretical approach the strongly interacting particles classified as hadrons are viewed mathematically as massless strings whose ends move with the speed of light in multidimensional space

by John H. Schwarz

Over the past 20 years hundreds of different "elementary" particles have been produced and identified in collisions caused by high-energy particle accelerators. During this period many empirical and theoretical schemes have been devised in an attempt to bring order to a mass of data that at first seems utterly chaotic. Considerable progress has been made, and many components of the problem are now well understood, but physics still appears to be far from constructing a complete theory of elementary particles. This article concerns a promising new approach, known as dual-resonance theory, that may prove helpful in understanding the behavior of one of the two broad classes of elementary particles. That class is the hadrons, particles that are subject to one of the four known forces of physics: the strong force.

All theoretical reasoning in elementary-particle physics is based on two well-established principles: the special theory of relativity and the quantum theory. In addition to stating the famous relation between energy and mass ($E = mc^2$), special relativity describes how observations made by observers moving with different velocities are related. Theories are usually formulated covariantly, which means that such relations are automatically incorporated in the mathematical formalism that expresses the theory. According to the quantum theory, a particular experimental interaction—say the collision of two particles—has various possible outcomes, each with a definite probability. These probabilities are represented in terms of complex numbers called scattering amplitudes. The task of any particle theory is to enable one to calculate to a fair degree of accuracy the scattering amplitude of a given particle interaction. By making scattering amplitudes (expressed as functions of a suitable set of variables) the object of the quest, quantum theory is also incorporated into the new theoretical framework.

Elementary particles are subject to several different kinds of forces, which can be grouped into two broad categories: strong and nonstrong. The nonstrong category is usually further subdivided into the electromagnetic force, the gravitational force and the weak force. For the purpose of this discussion we can ignore these latter distinctions, since the strong interactions are the main concern of dual-resonance theory. The strong forces operate between elementary particles only at short distances (typically on the order of 10^{-13} centimeter) and are some 100 times stronger than the strongest of the nonstrong forces. The strong forces are the ones that hold the atomic nucleus together; they are also largely responsible for the properties of the protons and neutrons that make up the nucleus.

As a philosophical matter, I am inclined to believe it should be possible to find a complete, consistent and even elegant mathematical description of all the particles and their interactions. To make progress toward formulating such a theory it seems appropriate first to attack more modest problems with the aid of various approximations. Two hypothetical systems that might be simpler to consider are (1) hadrons subject only to strong forces and (2) nonhadrons subject only to nonstrong forces. Each of the two systems is at best only an approximation, because actually hadrons are also subject to nonstrong forces. It is conceivable that within each system one could find a consistent theory that would accurately describe large classes of experiments. It is just as conceivable, however, that any such approximation would involve throwing away logical elements that are essential for achieving a mathematically consistent theory of all the elementary particles. For the moment one can only hope that the first possibility applies. Accordingly the rest of this discussion will be focused on the hadrons and their strong interactions, with the implicit assumption that a consistent overall theory can be found.

The theoretical understanding of the hadrons achieved so far is based mostly on approaches that do not attempt to give exact formulas for all possible scattering amplitudes. Rather, they describe important general features of hadrons that should be contained or explained in a complete theory. Examples of such approaches include various statistical models and particle-classification schemes.

An interesting possibility arising from a study of statistical models by Rolf Hagedorn of the European Organization for Nuclear Research (CERN) is that there may be an infinite number of hadrons whose distribution is an exponentially increasing function of mass. Such a distribution would imply the existence of a universal maximum temperature (approximately 2×10^{12} degrees Kelvin). There is some experimental evidence in favor of the Hagedorn type of distribution, but it is not yet conclusive. This type of particle spectrum does, however, arise naturally (and inevitably) in dual-resonance models. Moreover, it is a feature that places them in sharp contrast to conventional quantum-field-theory

models, which are the leading competitors of dual-resonance models for producing a theory of hadrons. The field-theory models seem hard put to account for such a mass distribution, although they do not exclude the possibility.

The significance of a universal maximum temperature can be appreciated by considering the following hypothetical situation. Suppose a pot of hadrons is near that temperature and one attempts to heat it to a higher temperature by adding energy. For example, the hadrons might be bombarded with high-energy particles under conditions designed to give rise to a thermal equilibrium. The prediction of the Hagedorn hypothesis is that at the maximum temperature any additional input of energy results entirely in the creation of new particles rather than an increase in temperature. It has been suggested that such a "fourth law of thermodynamics" may have influenced the evolution of the universe at a very early stage of the "big bang."

There are two notably successful particle-classification schemes that account for different aspects of the observed spectrum of elementary particles. The first (based on work by Murray Gell-Mann and George Zweig of the California Institute of Technology) is called the quark model. Reduced to its essentials, the quark model requires that hadrons fall into certain groupings, or families, in which the different members are alike except for certain quantum numbers, or fundamental properties, such as electric charge. (Examples of more esoteric quantum numbers are the properties designated isotopic spin and hypercharge.) The quark model further states that the choice of groupings can be understood by supposing the hadrons are made according to certain rules out of three fundamental constituents (quarks) and their antiparticles (antiquarks). This type of reduction is analogous to the unraveling of molecules into atoms, atoms into electrons and nuclei, and nuclei into neutrons and protons; now one is stating that neutrons and protons are in turn composed of quarks.

At this point analogies apparently break down, because the production of quarks as free particles has not been observed in high-energy collisions. For this reason and others many theorists now believe the quarks are permanently bound inside the hadrons and cannot be produced as free particles under any circumstances. This statement is easy enough to make; inventing theories that account for such behavior is more difficult. A number of attempts to understand the permanent binding of quarks have been made both in the context of field-theory models and of dual-resonance models. So far none of these proposals is very persuasive. Thus as things stand now the "quark confinement" problem represents a serious and perhaps pivotal challenge to all efforts at constructing a complete theory of elementary particles.

The second important particle-classification scheme does not have a simple name, even though it is also very successful. It states that the hadrons can be grouped into families whose members have distinct values of angular momentum and mass but are alike in all other respects. (Angular momentum can have only discrete values—multiples of Planck's constant—in accordance with the quantum theory.) The members of each of these families are said to lie on a hypothetical curve called a Regge trajectory (after the Italian physicist Tullio Regge). The idea that all hadrons lie on such trajectories, sometimes called the Regge hypothesis (even though Regge's original work did not involve hadrons), is now widely accepted.

The empirical facts are much more specific, however, than even the Regge hypothesis suggests. The Regge trajectories turn out on observation to be nearly linear, meaning that the angular momentum of the particles on a particular trajectory is given to a good approximation by a linear function of the mass of the particle squared [*see illustration on opposite page*]. Furthermore, all the Regge trajectories for ordinary hadrons appear to be approximately parallel. Additional evidence supporting the Regge hypothesis and the linearity of ordinary Regge trajectories is deduced from the observed high-energy behavior of the interaction-probability functions known to particle physicists as scattering cross sections. Dual-resonance models have the property that in the first approximation all the particles lie on parallel linear Regge trajectories. This fact ensures that the particle spectrum and the scattering amplitudes have many realistic features. In field-theory models, on the other hand, it is difficult to account for the linearity of the Regge trajectories. To my mind this property of dual-resonance models is one of the most compelling reasons for taking the approach seriously.

In practice dual-resonance theory is the study of a class of mathematical models. The main challenge, of course, is to find the particular example that describes our universe. The models that have been constructed so far definitely do not. Still, they have many features that encourage the belief we are on the right track, for example the Hagedorn distribution and the parallel linear Regge trajectories. Another important feature of these models is implied by their name. What is meant by dual resonance?

The term "duality" has been used in high-energy physics with several different meanings. Confusion can be avoided by separately defining "theoretical duality" and "phenomenological duality." Historically, phenomenological duality was discovered first (in 1967), and it played an important role in motivating the formulation of theoretical duality a year later. Phenomenological duality is the empirical fact that when the smooth asymptotic behavior of a scattering cross section at high energy is extrapolated back to lower energies, the resulting curve tends to represent an average of the bumps and dips of the true cross section [*see top illustration on page 64*]. The bumps are manifestations of the short-lived, unstable particles called res-

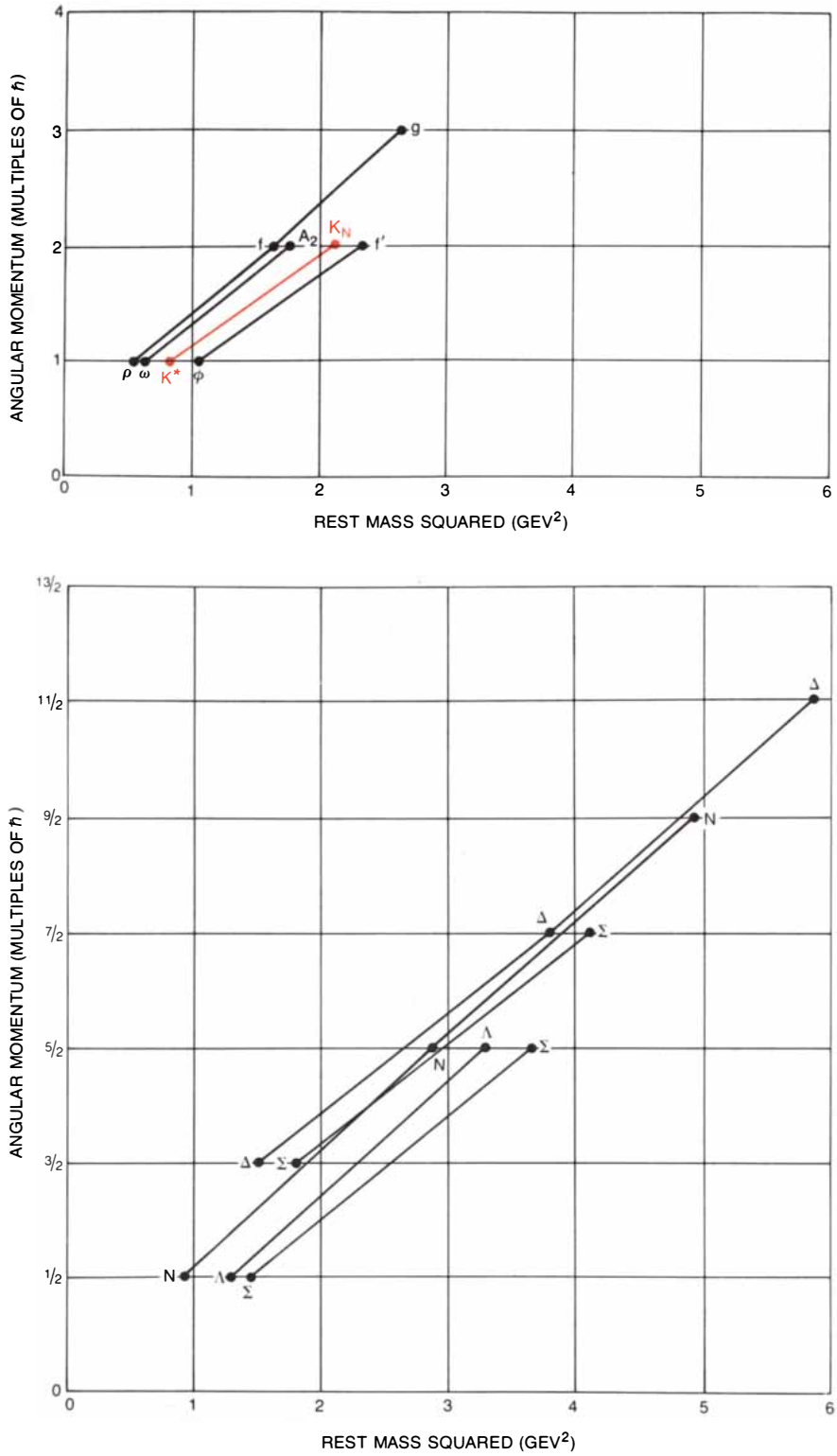
HADRONS	MESONS	PIONS (π^+, π°, π^-) KAONS (K^+, K°, $\bar{K}^\circ$, K^-) ETA (η°) RHOS (ρ^+, ρ°, ρ^-) OMEGA (ω)
	BARYONS	PROTON (p) NEUTRON (n) LAMBDA (Λ°) SIGMAS (Σ^+, Σ°, Σ^-) DELTAS (Δ^{++}, Δ^+, Δ°, Δ^-)
NONHADRONS	LEPTONS	ELECTRON (e^-) POSITRON (e^+) MUONS (μ^+, μ^-) NEUTRINOS (ν_e, $\bar{\nu}_e$, ν_μ, $\bar{\nu}_\mu$)
	OTHERS	PHOTON (γ) GRAVITON—NOT YET OBSERVED

HADRONS AND NONHADRONS, particles that are subject to strong forces as opposed to particles that are not, are broken down into four major subclasses in this partial listing, which includes only the comparatively longer-lived and less massive of the observed elementary particles. Mesons and baryons, the two subclasses of hadrons, are distinguished by various quantum properties such as rotational angular momentum, or spin. A symbol with a bar above it denotes an antiparticle; in a few cases a particle is its own antiparticle. Neutrinos and antineutrinos are of two types: electron-type, ν_e, and muon-type, ν_μ. Hundreds of shorter-lived, heavier hadrons (also known as resonances) have been created in particle accelerators.

onances. Even though there is some ambiguity in the extrapolation procedure, phenomenological duality enables one to derive approximate relations between the parameters describing the high-energy behavior and the masses and coupling strengths of the resonances at lower energies.

Theoretical duality is a property incorporated in the construction of dual-resonance models. It is a precise mathematical statement about the structure of the scattering amplitudes calculated in a first approximation, in which the formulas are relatively simple because the resonances have no width (that is, no spread in energy). Dual-resonance models are calculated as a series of successive approximations, starting from and determined by this first approximation. The entire series constitutes the full theory, but theoretical duality concerns only the first approximation. The precise statement of theoretical duality is that the scattering amplitude calculated by summing the contributions of all resonance "exchanges" equals the one obtained by summing over all "direct channel" resonances [*see bottom illustration on next page*]. This formulation is very similar to the "bootstrap" hypothesis advocated by Geoffrey Chew of the University of California at Berkeley and Steven Frautschi of Cal Tech. According to their hypothesis, the exchange of resonances provides the forces responsible for holding the direct-channel resonances together. One difference is that the duality mechanism does not determine the overall strength of the forces. The new feature of dual-resonance models that makes this possible is that they are based on an infinity of resonances with tightly correlated properties specifically designed to incorporate the bootstrap mechanism in the first approximation.

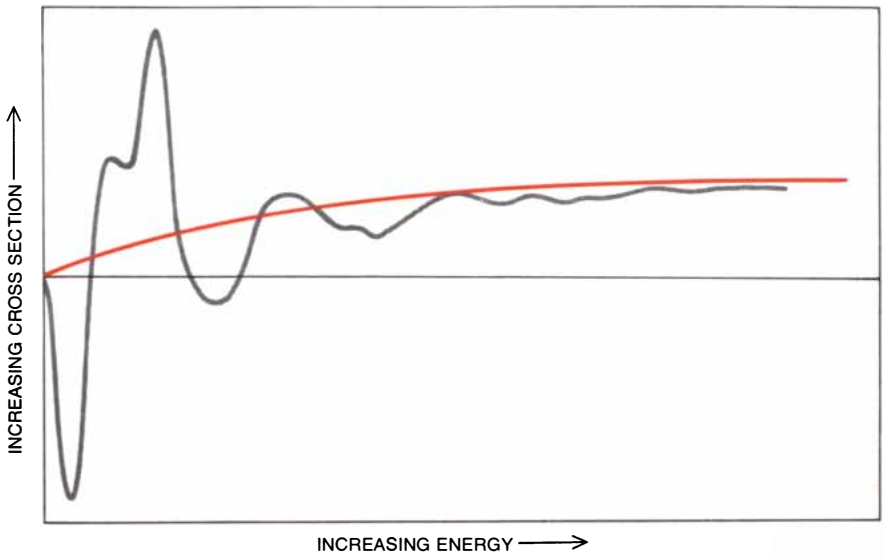
Historically dual-resonance models were developed by inventing formulas for scattering amplitudes in the first approximation. The process was begun by Gabriele Veneziano of the Weizmann Institute of Science and CERN and was carried forward by many others. The theory was then studied to show that it satisfies certain general principles, such as the requirements that interaction probabilities are positive and add up to 1, and that a signal cannot be received before it is sent. The program of understanding how these and other properties are incorporated in dual-resonance theory has been successfully carried out during the past few years. This approach has been encouraged by the recent work of Yoichiro Nambu of the University of Chicago, Leonard Susskind of Yeshiva



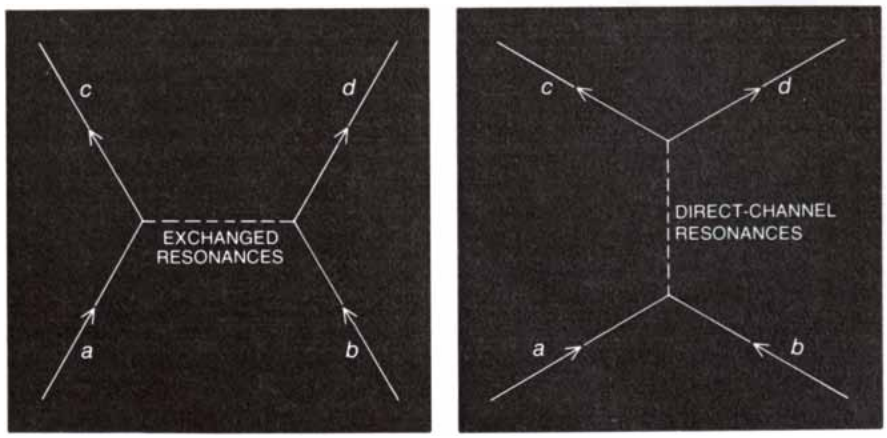
REGGE TRAJECTORIES, hypothetical curves along which families of hadrons with distinct values of spin and mass can be grouped, turn out on observation to be nearly linear and approximately parallel, a finding that is consistent with a prediction of dual-resonance theory. Some of the best-established meson trajectories are indicated by the slanted lines in the graph at top. The symbols denote terminal members of each meson family according to several different naming systems; some of the mesons represented are extremely unstable resonances not listed in the illustration on the opposite page. Particles with the property called strangeness are shown in color. Some of the best-established baryon trajectories are displayed at bottom. Here *N* stands for nucleon, which can be either a proton or a neutron. The unit of angular momentum, $\hbar$, is equal to Planck's constant (designated h) divided by 2π. Mesons have integer values of angular momentum; baryons have half-integer values. Rest mass is given in equivalent energy units; *GeV* stands for gigavolts, or billions of electron volts. All the trajectories in both categories have a slope of slightly less than $1/\text{GeV}^2$.

University, Holger Nielsen of the Niels Bohr Institute and others, showing that there is a simple physical picture of hadrons implied by dual models. This picture provides an elegant way to describe dual-resonance theory, even though it is often not the most convenient for actual calculations.

What Nambu and the others have found is that the hadrons in a dual-resonance model can be identified as the quantum states of a string, albeit a very special type of string. For one thing the string has no intrinsic mass, so that its motion is necessarily relativistic; thus, for example, the ends must always move with the speed of light. Strings with these properties are called light strings. The mass of a hadron arises from the energy due to tension in the string and the kinetic energy of its motion with respect to the center of mass. (The tension, determined by the slope of the Regge trajectories, turns out to be about 13 tons.) Here the difference between dual-resonance models and field-theory models emerges. The elementary particles described by a field-theory model are pointlike, whereas those in a dual-resonance model are spatially extended.



DUALITY OF RESONANCES is defined empirically in terms of the observed behavior of the interaction-probability functions known to particle physicists as scattering cross sections. In this example the gray curve represents the difference between the total cross section for π^+p interactions and the total cross section for π^-p interactions. The bumps and dips in the curve, evident particularly at low energies, are manifestations of resonances. The colored curve is an extrapolation of the smooth high-energy behavior of the cross-section difference back to lower energies. The fact that the colored curve tends to average the bumps and dips of the gray curve is what is referred to as phenomenological duality.

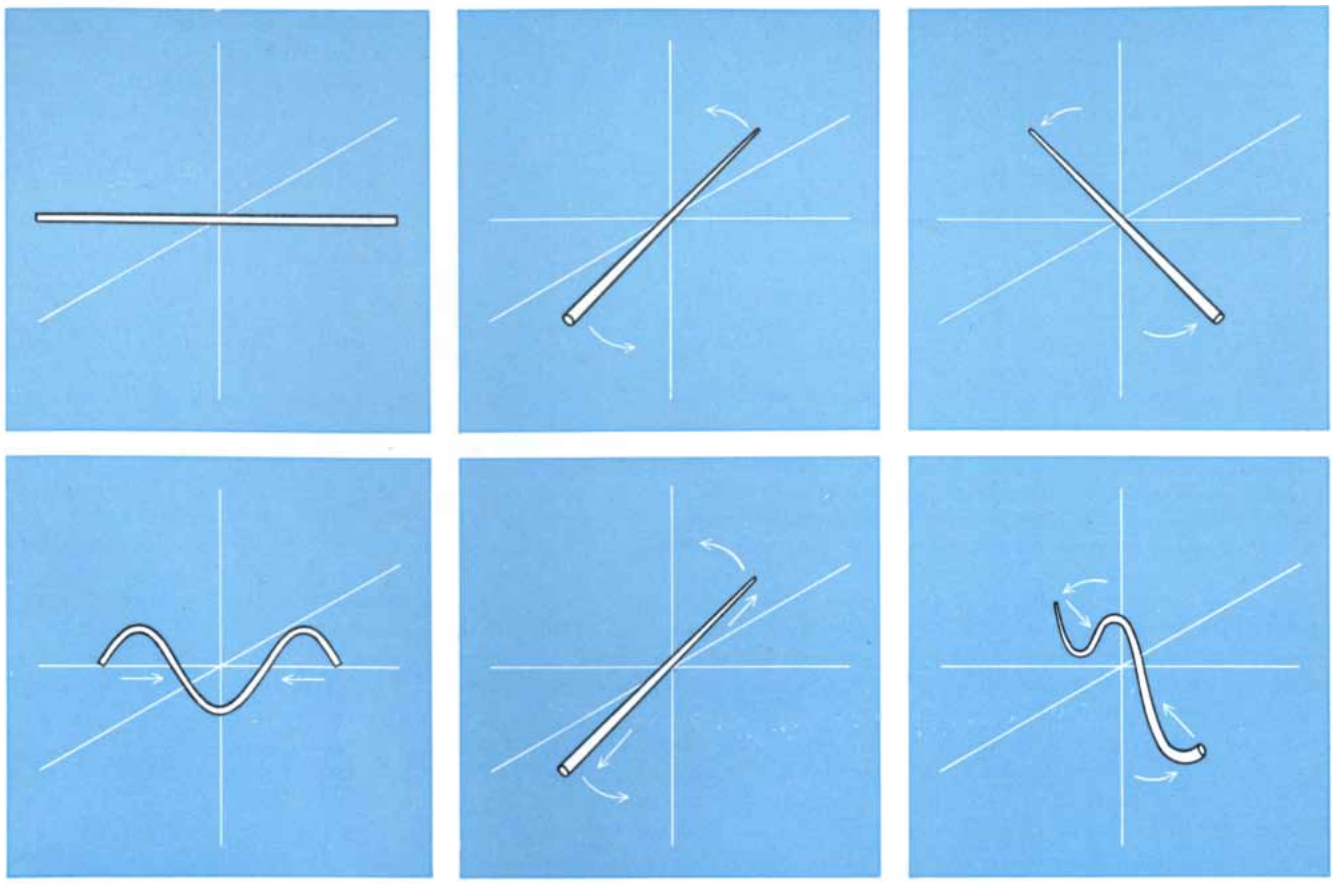


TWO ALTERNATIVE WAYS of understanding the forces that are involved in an interaction in which two incoming elementary particles (*labeled "a" and "b"*) collide to produce two outgoing particles (*labeled "c" and "d"*) are presented here. The forces can be thought of as resulting from an exchange of resonances, as depicted in the schematic diagram at left, or they can be thought of as resulting from the formation of an intermediate resonance that subsequently decays, as in the "direct channel" scattering process depicted at right. Theoretical duality consists of the assertion that these two scattering processes, when considered mathematically at a certain well-defined first level of approximation, are exactly equal.

Once the possibility of spatial extension is accepted it is natural to wonder why the structure of a hadron should be one-dimensional (strings) rather than two- or three-dimensional (blobs). By definition dual-resonance models only allow for a one-dimensional structure. The possibility of developing a theory of blobs is being considered. All we know so far is that any such theory is mathematically much more formidable. In my own opinion the existing schemes based on strings are sufficiently realistic to make it plausible that a correct theory of hadrons can be based on a fancier type of string than any that has been considered to date.

The theory of light strings could have been investigated 40 years ago, but there was little motivation for doing so until the connection with dual models was suspected. This exercise is an interesting and instructive application of the principles of special relativity and quantum mechanics. If one considers only the classical (that is, nonquantum) motions of noninteracting strings, the problem is quite straightforward. The inclusion of the requirements of quantum theory has a profound effect. One striking consequence is that it becomes impossible to satisfy the principles of quantum mechanics and special relativity simultaneously unless the dimension of the space in which the strings move around is chosen to have a particular value. In the real world, of course, space has three dimensions. In the simplest string model it is required to have 25! Fortunately different string models have different critical dimensions, so that this restriction need not represent an insuperable difficulty. Rather, one sees the interesting possibility that the number of dimensions may someday be derived from, or at least related to, other quantities.

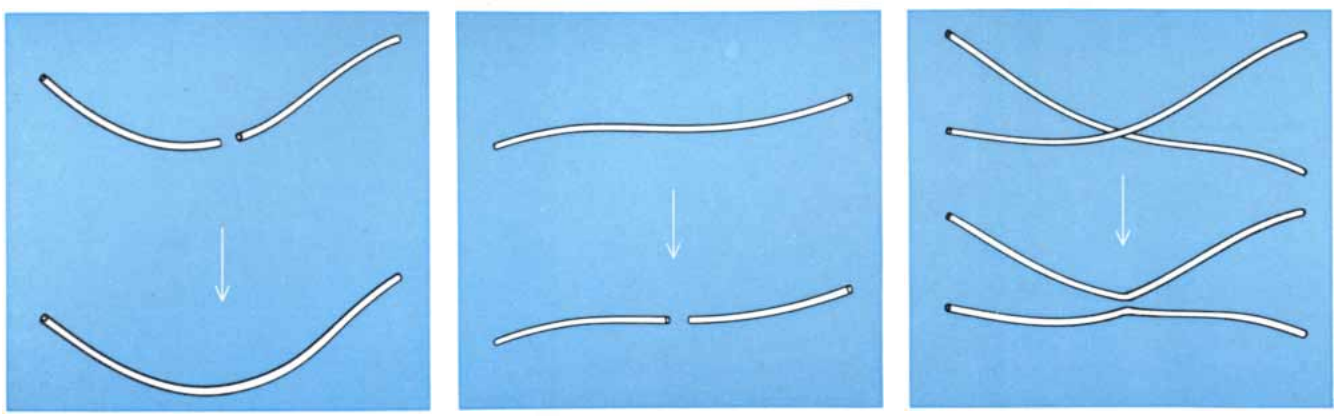
The relation between elementary particles and the states of motion of a hypothetical string deserves some further elaboration. According to the laws of classical physics, a light string has a continuous infinity of possible motions. Quantum theory reduces the number of possibilities in the center-of-mass reference frame to a set of discrete quantum states. The number of quantum states is also infinite, but up to any given mass it is finite. Actually the distribution is precisely of the Hagedorn type. As an ex-



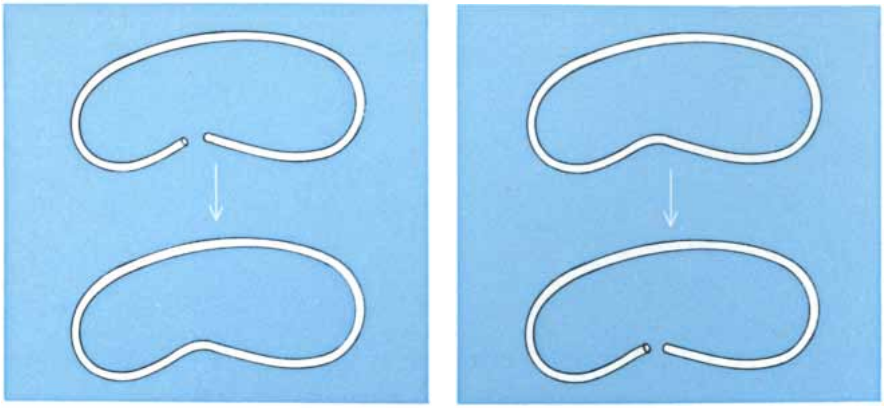
STRING ANALOGY has proved useful in describing the quantum states of hadrons in a dual-resonance model. The analogy calls for a string with no intrinsic mass; as a result the ends of the string must always move with the speed of light. Although the structure of such a "light string" is assumed to be one-dimensional, quantum theory requires that the space in which the string moves around, even in the simplest dual-resonance model, must have 25 dimensions! For a given hadron mass the string has a finite number of possible motions in the center-of-mass reference frame, corresponding to a set of discrete quantum states. For example, the string can rotate around its center while remaining straight (*diagrams at top*), or it can rotate and oscillate simultaneously (*diagrams at bottom*).

ample of a possible motion, suppose the string is rotating around its center point while remaining perfectly straight. The rate of rotation in this case must be arranged so that the angular momentum is a multiple of Planck's constant. The overall length of the string is uniquely determined in each case so that the ends move with the speed of light. This class of states contains the maximum possible amount of angular momentum for a given amount of mass. Therefore the hadron states described in this way correspond to ones on the leading Regge trajectory. There are many other possible motions in which the string undergoes oscillations as well as rotations [*see illustration above*].

The theory of light strings discussed above describes the particle spectrum of a particular dual model in the first approximation. The full story is more complicated, because the strings can interact with one another [*see illustration below*]. The basic interaction was shown by Stanley Mandelstam of the University of California at Berkeley to consist of a string breaking in two or two strings joining together. Another possibility is that when two strings touch at interior



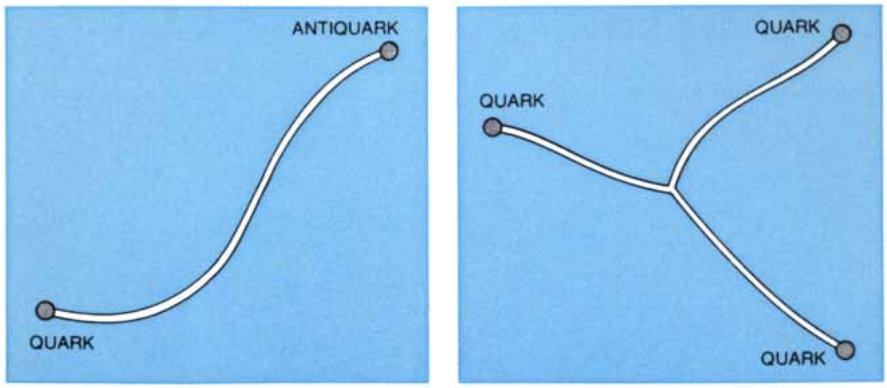
LIGHT STRINGS CAN INTERACT according to dual-resonance theory by joining (*diagram at left*), breaking (*diagram in middle*) or breaking and then rejoining (*diagram at right*). The last phenomenon is analogous to the "crossing over" seen in chromosomes.



NEW TYPE OF LIGHT STRING, required for mathematical consistency, is a closed string, or loop, formed when an open string joins its ends (*diagram at left*). Conversely, the breaking of a closed string to form an open string should be possible (*diagram at right*). Hypothetical particles called pomerons appear to be describable in terms of a closed-string dual-resonance model, but pomeron particles have not been identified so far in the laboratory.

points, they break and rejoin so that each segment of one string joins a segment of the other. (An analogous phenomenon in biology is the "crossing over" of chromosomes.) These rules for the interaction of strings make it possible to calculate the scattering amplitudes for various possible reactions among hadrons. One effect of the interactions is to modify somewhat the properties of the hadrons themselves. For example, the Regge trajectories do not remain exactly linear when one includes corrections beyond the first approximation in considering interactions.

When the effects of the interaction of strings are taken into account, one immediately discovers that a new type of string is required. Since two strings can interact by joining ends, what is to prevent the two ends of a single string from joining, thereby becoming a closed string, or loop? As a matter of fact, mathematical consistency necessitates that closed-string hadrons be included in the theory, even though they are not present in the first approximation [*see illustration above*]. Hadrons corresponding to quantum states of closed strings also lie on parallel linear Regge trajectories (in the first approximation), with a slope exactly half that of the trajectories associated with open strings. The occurrence of these new states in the theory is exciting, because the experimental facts require something beyond the ordinary open-string Regge trajectories. The high-energy behavior of elastic-scattering reactions is seen to involve a diffractive type of interaction. One widely held viewpoint is that diffraction scattering can be explained in terms of single and multiple exchanges of a "pomeron" Regge trajectory. (The name is derived from that of the Russian physicist Isaak Yakovlevich Pomeranchuk.) The closed strings provide excellent candidates for pomeron trajectories. Indeed, this particular theory requires that particles be associated with pomeron trajectories. So far no clear-cut candidates for pomeron particles have been detected in the laboratory, but there are several reasons why such particles should be hard to detect.



QUARKS CAN BE INCORPORATED into dual-resonance models in several possible ways. For example, if one assumes that quarks are pointlike fundamental particles attached to the ends of light strings, a hadron containing one quark and one antiquark (that is, a meson) would be simple to describe (*left*), whereas a hadron with three quarks (that is, a baryon) would require a more complicated topology (*right*). All the schemes that have been suggested so far for relating quarks to string picture present serious mathematical difficulties.

Where do quarks fit into the string picture? One suggestive possibility is that they are pointlike particles (as in any field theory) attached to the ends of strings. If one supposes new quarks always appear at the ends formed by the breaking of strings, then this picture provides a simple explanation of quark confinement: The quarks are held together by strings! At first the picture looks promising for the class of hadrons called mesons, which contain one quark and one antiquark, and for pomeron states, which contain no quarks. The hadrons called baryons, however, are composed of three quarks each and hence require a more complicated construction, such as a string in the shape of a *Y*, in order to have three free ends to which to attach the quarks [*see bottom illustration at left*]. Such a topology gives rise to serious mathematical difficulties. In fact, even the meson construction has problems. Quarks are supposed to carry electric charge and spin, but such particles apparently cannot be attached consistently at the ends of a string moving with the speed of light. One might try to slow the ends down by also attributing mass to the quarks, but this raises other serious problems. In spite of its initial attractiveness, attaching quarks to the ends of strings may be a completely faulty method of incorporating them into dual-resonance models.

One alternative to attaching quarks to the ends of strings would appear to be that the various quantum numbers (such as electric charge) quarks are supposed to carry are distributed along the length of the strings. In this case the baryons need not require a new topology. The suggestion that quark quantum numbers are not localized at points is somewhat unpopular in the theoretical community. The main reason is that recent experiments at the Stanford Linear Accelerator Center (SLAC) involving the high-energy scattering of electrons from protons show the property known as scaling. Scaling is usually interpreted as direct evidence that the electric charge inside a proton at a given instant is localized at one point or a few points, that is, at the positions of the quarks. Another alternative that ought to be compatible with scaling is that the quarks are pointlike particles moving along the string like

beads on a necklace. Unfortunately this approach also leads to very complicated mathematics, so that its consequences cannot be pursued to any degree.

One major shortcoming of the simplest string model, besides its confinement to 25 dimensions, is its failure to account for fermions: particles with half-units of spin. (Spin, like other quantum properties, is confined to certain discrete values.) Among the hadrons these particles are precisely those that are baryons. If the idea of achieving the half-unit spin by attaching three quarks, each with a half-unit of spin, to a string is rejected, then the string itself must somehow give rise to the spin. Noninteracting dual-fermion states were theoretically constructed by Pierre Ramond in 1971. Shortly thereafter a dual model of interacting particles with spin incorporating the Ramond fermions was developed by André Neveu and me. This model satisfies the same general principles as the simpler one, but it contains a more realistic spectrum of particles. A string description of this model has been shown to be consistent with special relativity and quantum mechanics, provided that space has nine dimensions. Even though that is still not the right number, it represents a step in the right direction. Moreover, there seems to be at least a possibility that six of the nine dimensions can be assigned to internal degrees of freedom rather than to ordinary space. If that can be done in a fully consistent manner, then perhaps a modified form of this model could be the right one.

The dual model with spin represents a nontrivial extension of the previous dual model. Whereas the "action" representing the model without spin has a group of mathematical symmetries analogous to that of the general theory of relativity, the model with spin has an even larger group of symmetries. If the elegance of a theory is measured by the amount of symmetry it possesses, then the dual model with spin rates very high. My own belief is that elegance, so defined, is closely correlated with physical relevance.

In summary, a new theory of elementary particles—dual-resonance theory—has been developed over the past six years. It has not yet provided any striking new experimental predictions that have been confirmed (or excluded), but it successfully accounts for many general features of the hadron system. If a few remaining obstacles can be overcome, the chances appear good that this work could eventually result in a complete theory of the strong interactions and the particles that respond to them.

This is a new Hewlett-Packard calculator. The HP-55.

It has 86 preprogrammed functions and operations, and 20 memory registers. You can program it to do repetitive calculations. You can time events with it. It costs $395*

Our new HP-55 not only gives you 38 more mathematical functions and operations on the keyboard than our HP-45. It will also remember an intricate calculation so that you can keep entering variables and see new answers until you're satisfied and turn it off.

You can program an equation with up to 49 steps in it—in addition to which your program has access to all 86 preprogrammed calculations and data manipulations, plus the 20-register memory. You can do direct and conditional-access branching (the calculator will automatically compare values and choose among alternatives). When you've entered your program you can step through it to verify and edit.

You can perform arithmetic into and out of 10 of the registers, store in all 20.

You can do circular trigonometry in degrees, radians, or grads, converting directly from degrees to radians, decimal degrees to D/M/S (and vice versa).

You can do true linear regressions. The special $\hat{y}$ key lets you fit curves.

You get instant true metric conversions without having to multiply by a constant.

You also have the HP four-register stack that lets you enter even the most complicated equation in an easy, natural way without having to keep track of parentheses and brackets. The result of each intermediate calculation is displayed as you perform it, giving you the continuous feedback you need to avoid ending up with a wrong answer.

And whether you need it or not, you get a 100-hour digital timer that displays and records hours, minutes, seconds, tenths, and hundredths—accurate to 0.01 percent. Results are stored in 10 memory registers, so you can clock 10 different events or "splits" from the same starting time—runners, reactions, rats, or whatever—and study the results when things calm down.

As noted above, this extraordinary little machine costs $395*. It doesn't write on and read 100-step magnetic program cards like our incomparable HP-65, but it has more power on its keyboard. If you don't mind explaining a calculation to the HP-55 whenever you want it to start repeating at your command, you can save a lot of money. And so you won't have to think up your own explanations for many common types of problems, we offer two HP-55 applications handbooks—The Math Pac and Stat Pac—at $10* each.

People tend to keep their HP-65's locked in their desks. You can afford to treat the HP-55 more casually. Keep it in your pocket.

How to program without being a programmer.

If you often find yourself repeating a tedious computation over and over, putting in different variables to see a range of answers, you're the person we designed the HP-55 for.

All you have to do is flip a switch to tell the calculator it's going to memorize a computation sequence.

Then you simply key in your calculation (we call it "keystroke programming") as if you were only going to do it once. Except that before and after you enter each variable you press R/S (for Run/Stop). That tells the calculator that whenever you run the program it must stop at that point, wait for a new number, and start computing again after the number is entered.

When you've finished the initial calculation, you press R/S to tell the HP-55 that the sequence is complete. Then you flip the switch to RUN and press BST, whereupon the machine will instantly calculate up to the first variable, stop and wait. When you've entered the new number, press R/S and the machine will zip through the next sequence, stop again and wait for another number. After the last variable is in, press R/S to see a new answer.

The HP-55 will remember your sequence as long as you don't switch it off. So you can iterate and reiterate until you run out of variables.

A better way to measure blood oxygen—non-invasively.

Consider the human ear: external, thin in cross-section, and highly vascularized. Since it metabolizes only a small part of the blood it carries, oxygen saturation of hemoglobin in the ear is high—comparable to that in the aorta.

The ear oximeter makes use of these characteristics of the ear to measure the oxygen saturation of hemoglobin in a patient, without drawing blood. With the ear oximeter, the physician can make the measurement simply by attaching an optoelectronic device to the patient's ear and reading percent oxygen saturation on the display: no blood sample, no delay for lab analysis, no patient discomfort and the opportunity to make continuous measurements and study the dynamics of oxygen saturation.

The oximeter works by analyzing the absorption of light transmitted through the upper part of the ear. Other things being equal, the absorbance spectrum varies directly with the oxygen saturation level in the blood. But other things are seldom equal. There are changes in absorbance from patient to patient that depend on tissue and pigment variations, for example.

The new HP 47201A Oximeter uses digital processor technology to compensate for these patient variations automatically. By specially processing arterial blood tapped from volunteers of both sexes and all ages, races, and oxygen saturation levels, researchers determined the coefficients of absorbance that provide the best fit of oximeter readings to true oxygen saturations. This information is stored in the 47201's digital memory, in effect permanently calibrating it for the full range of patient variables in its intended use. Additionally, the absorbance measurement itself is made at *eight* different wavelengths because different light-interfering substances in the blood and ear have different absorbance effects at different wavelengths. Finally, the 47201's digital processor solves simultaneous equations that take all these calibration and measurement variables into consideration and calculates the percent oxygen saturation. Its accuracy is ± 1.7 percent for saturation levels between 90 and 100 percent, to ± 2.6 percent in the 60 to 70 percent range (1σ).

The 47201 Oximeter provides a reliable indication of respiratory performance on patients with good arterial circulation: it is specifically not recommended when blood flow to the ear is impaired. It is a valuable diagnostic tool especially in pulmonary function labs, for the care and rehabilitation of patients suffering chronic lung diseases, and for the detection and management of hypoxemia. Price is $8200.*

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

*Domestic USA prices only

00541

Mail to: Hewlett-Packard, 1505 Page Mill Road, Palo Alto, CA 94304

( ) HP-55 Programmable Calculator
( ) HP 47201A Oximeter

Name_____ Title_____

Company_____

Address_____

City_____ State_____ Zip_____

HEWLETT hp PACKARD

Sales and service from 172 offices in 65 countries.
Palo Alto, California 94304

BIOLOGICAL CLOCKS OF THE TIDAL ZONE

Endogenous clocks set to the rhythm of the solar day are known throughout the biological world. Many organisms that live along the shore also have a clock set to the rhythm of the lunar day

by John D. Palmer

In the sands between the tide marks on the north shore of Cape Cod lives a microscopic golden brown alga, the diatom *Hantzschia virgata*. The protoplasm of this single-celled plant is encased in an elongated glassy cell wall perforated in places by pores and slits. Through some of the end pores is exuded a mucuslike substance that serves to slowly jet-propel the diatom through its subterranean habitat. During each daytime low tide the tiny motile organism glides up through the interstices between the grains of sand to the surface. There it remains throughout the ebb tide, its photosynthetic machinery bathed in sunlight. In midsummer the diatoms are so abundant that in spite of their microscopic size they form a prominent golden brown carpet over the beach. Moments before they are inundated by the returning tide they move down into the comparative safety of the sand.

A fascinating aspect of this vertical-migration behavior becomes more apparent when sand bearing the diatoms is transferred from the north shore of Cape Cod to the Marine Biological Laboratory at Woods Hole on the south shore. The samples are placed in an incubator where the temperature is held constant and the light is left on continuously. In this new environment, which lacks days, nights and tidal changes, the diatoms continue their periodic excursions up to the surface of the sand in virtual synchrony with the diatoms 27 miles away. Their movements in the laboratory are sufficiently punctual so that when we plan a collecting trip to Cape Cod Bay, we sometimes observe the diatoms in the incubator instead of consulting tide tables. Since the rhythm of the diatoms persists in the absence of the environmental periodicities that would be expected to govern such behavior, it seems that within the plants there is a biological clock that directs the temporal aspects of their lives.

This account is not just another amusing anecdote about a rare occurrence in nature. Clock-controlled rhythms are displayed by most inhabitants of the tidal zone. The rhythms are characterized by the repetition of some behavioral or physiological event, such as a flurry of activity, synchronized with a particular phase of the tide. Since there are two tides each lunar day (a lunar day is 24.8 hours in length, the interval between successive moonrises), the rhythms are called bimodal lunar-day rhythms, in contrast to the unimodal solar-day rhythms of organisms geared to the 24-hour solar day. The biological clocks related to both the lunar-day and the solar-day rhythms are apparently important as an aid to survival in that they give advance warning of the regular changes in certain periodic aspects of the environment, such as nightfall or the return of the flood tide. Under unchanging conditions in the laboratory the clocks continue to function, and thus biological rhythms persist for a considerable length of time.

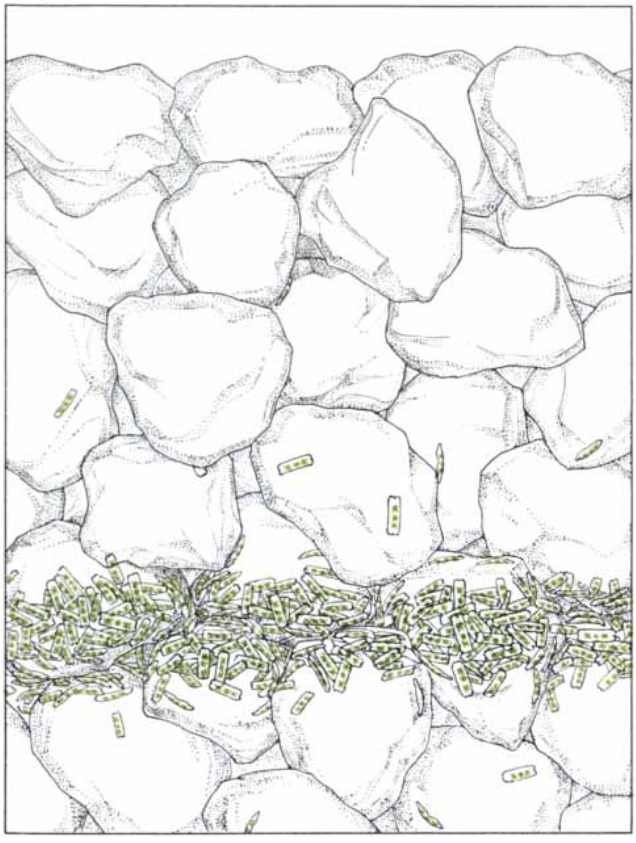
The fiddler crab, a common denizen of mud flats and sand flats on North American coastlines, emerges from its burrow at low tide. It scurries sideways around the flat eating detritus. The males feign battles with one another and try to entice females into their bachelor burrow with awkward beckoning movements of their enormous fiddle claw. With each flood tide all the crabs retreat back into their burrow, where they sit out the deluge.

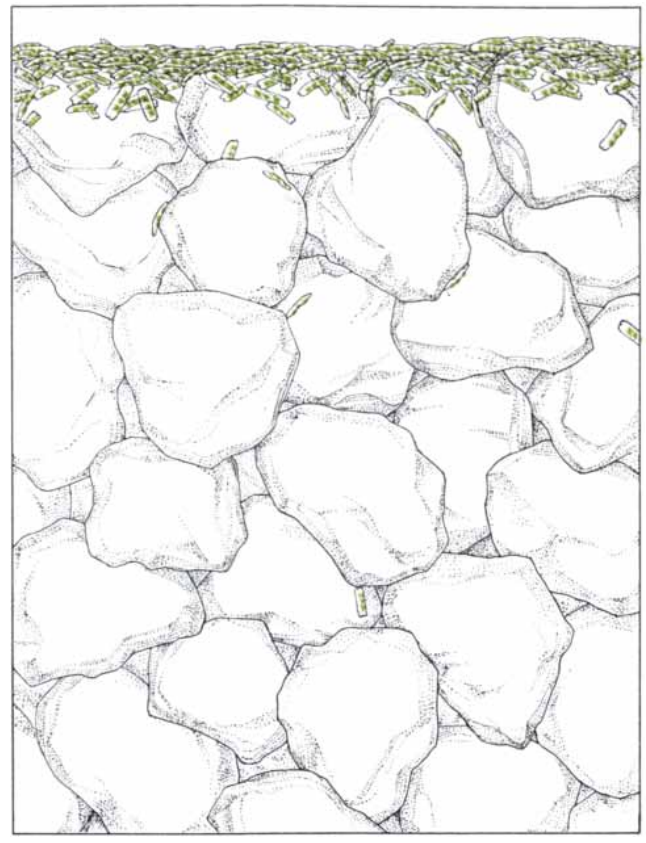
In the laboratory quantifying the locomotor behavior of fiddler crabs is quite simple. Single crabs are placed in plastic boxes (the kind in which fishing lures are bought, and therefore a common commodity in Woods Hole). The boxes are balanced on a knife-edge fulcrum, and as the incarcerated crab moves between ends of this improvised actograph the box teeters, closing a microswitch that causes a deflection of a pen on a chart recorder. The actograph is placed in the unchanging environment of an incubator, and the crab is allowed to perform spontaneously for days. In these monotonous surroundings the crab's clock continues to operate and dictates almost, but not entirely, the same ambulatory pattern found in nature. The difference is slight but significant: in the laboratory the period of the bimodal lunar-day rhythm is slightly longer or slightly shorter than the period displayed in nature. This change in periodicity when an organism is placed in constant conditions is a property of almost all clock-controlled biological rhythms. Since tidal rhythms follow the lunar day, they are called circalunadian (about a lunar day).

The rhythms in some species of fiddler crab will persist for as long as five weeks in the laboratory, but more often they are damped out rather quickly. The crabs must occasionally be exposed to periodic immersion in seawater if their tidal rhythm is to be maintained. Even in nature whenever small populations of fiddler crabs become established along the margins of pools not subject to tides, they lose their tidal rhythm and display only a solar-day rhythm. When the crabs are returned to a tidal flat, they quickly reestablish a lunar-day rhythm that will then persist for some time even after the animals are removed from the tidal location.

Living side by side on the same flats

BEACH ON CAPE COD is inhabited by *Hantzschia virgata*, a species of diatom. These algae live in the damp sand between the high-tide mark and the waterline at low tide. The upper photograph shows a section of the beach that has just been exposed by an outgoing tide. The lower photograph, made a few minutes later, shows golden brown patches formed by diatoms that have migrated to the surface of the sand. When sand containing these diatoms is placed in an incubator where the light and temperature are held constant, the organisms continue to migrate to the surface synchronously with the periods of the daytime low tides at their home beach. The familiar shoreline object seen at the right in both photographs served to mark the location for purposes of comparison.

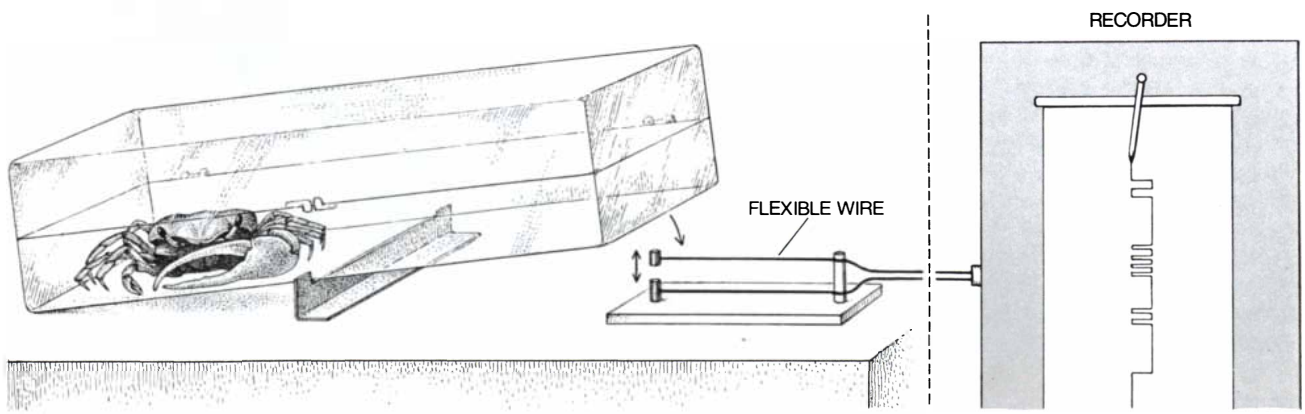




MIGRATORY BEHAVIOR of *Hantzschia virgata* is depicted in vertical section. Diatoms of this species normally reside about a millimeter below the surface of the sand (*left*). Each diatom has two *X*-shaped chloroplasts, which are the site of photosynthesis. During daytime low tides the organisms are propelled upward to the surface by mucus that is forced through pores at the end of their elongated glassy cell wall (*right*). The diatoms remain in the sunlight until moments before sand is inundated by returning tide.

with the fiddler crab are the green crab and the penultimate-hour crab. Both of these crustaceans display tidal rhythms, but they differ from the fiddler crab in that their activity is synchronized with the times of high water. The rhythms of the two species will persist in constant laboratory conditions for about a week before being damped out. In the case of the green crab it has been found that animals that have lost their rhythm in the laboratory need not be subjected to the tides to reestablish the rhythm. Instead it can be reinstated by cooling the crabs to a temperature of four degrees Celsius (39 degrees Fahrenheit) for six hours.

This technique has also been used to demonstrate that tidal rhythms are not learned or otherwise impressed on crabs by the tides themselves. Barbara Williams, working with Ernst Naylor at University College of Swansea, had the perseverance and the rare skill necessary to raise green crabs in the laboratory from eggs through several larval stages to adults. During the entire maturation process the crabs were exposed only to the alternating day-night changes in the laboratory. When the crabs were large

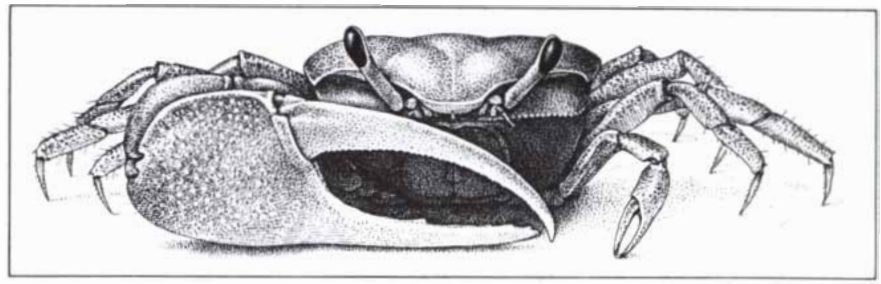


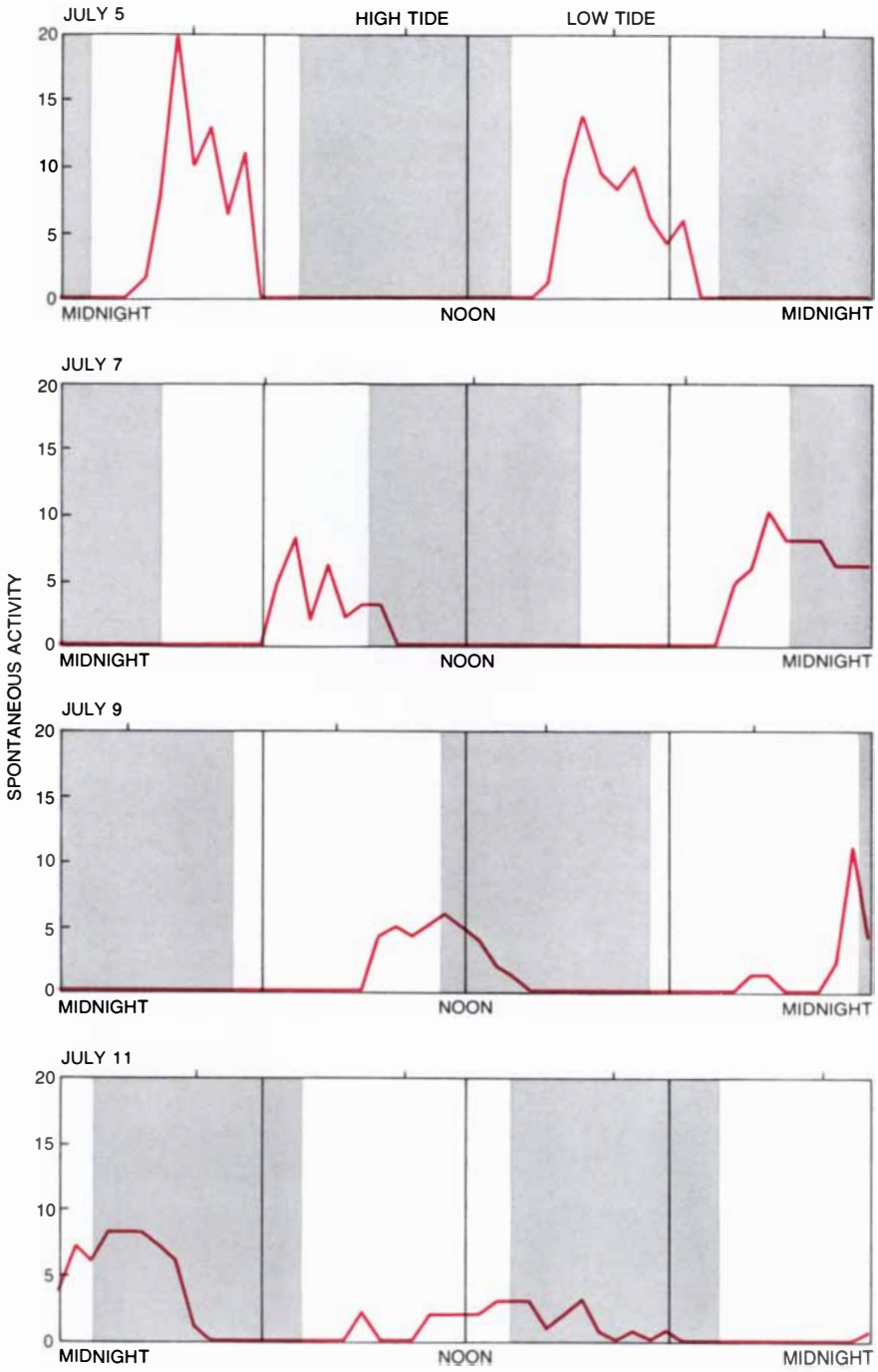
DAILY ACTIVITY of a fiddler crab is recorded by an actograph that consists of a plastic box balanced on a knife-edge fulcrum. When the crab moves to the near end of the box, the box tilts and closes a switch that causes an excursion of a pen on the recorder. The number of daily back-and-forth movements of the crab is determined by counting the number of excursions made by the pen.

enough for their activity to be studied, it was found that their locomotor activity was limited to the daylight hours. Williams then gave the crabs one 15-hour cold treatment and recorded their subsequent locomotor behavior. A distinct tidal component appeared in their activity. Since a single 15-hour cold spell could not have provided the crabs with any information about the 12.4-hour cycle of tides, it is reasonable to conclude that the clock that measures the tidal frequency is innate, and that it merely needs to be activated by some environmental stimulus for its first expression.

Under natural conditions both the penultimate-hour crab and the green crab display in their locomotor activity a clear-cut solar-day rhythm as well as a lunar-day one. In the penultimate-hour crab the solar rhythm appears as a broad peak of activity spanning the hours of darkness. In the green crab the solar rhythm is represented not as an individual peak but as a decrease in the amount of activity at the crest of the daytime tide. The combination of solar-day and lunar-day rhythms is rather common in intertidal organisms, and it raises the question of whether such organisms have a solar-day clock for one rhythmic component and a separate lunar-day clock for the other, or whether a single horologe drives both rhythms. A single-clock mechanism might be regarded as analogous to the kind of wristwatch worn by surf fishermen, in which a single movement is transmitted to present on the dial both the time of day and the time of the tide.

Processes other than locomotor activity are also controlled by the crab's biological clock. Color-change rhythms have been investigated in the fiddler crab, the green crab and the penultimate-hour crab by Frank A. Brown, Jr., Marguerite Webb and Milton Fingerman at the Marine Biological Laboratory and by B. L. Powell of Trinity College in Dublin. Within the hypodermis of these crabs are star-shaped chromatophores that contain granules of dark pigment. When the pigment granules are tightly aggregated in the center of these cells, the coloration of the crabs is light. When the granules are evenly dispersed throughout the extensions of the cells, the coloration is dark. All three species of crab blanch during the night and darken during the daylight hours, even when they are placed in constant conditions in a laboratory. The color-change pattern of the fiddler crab has



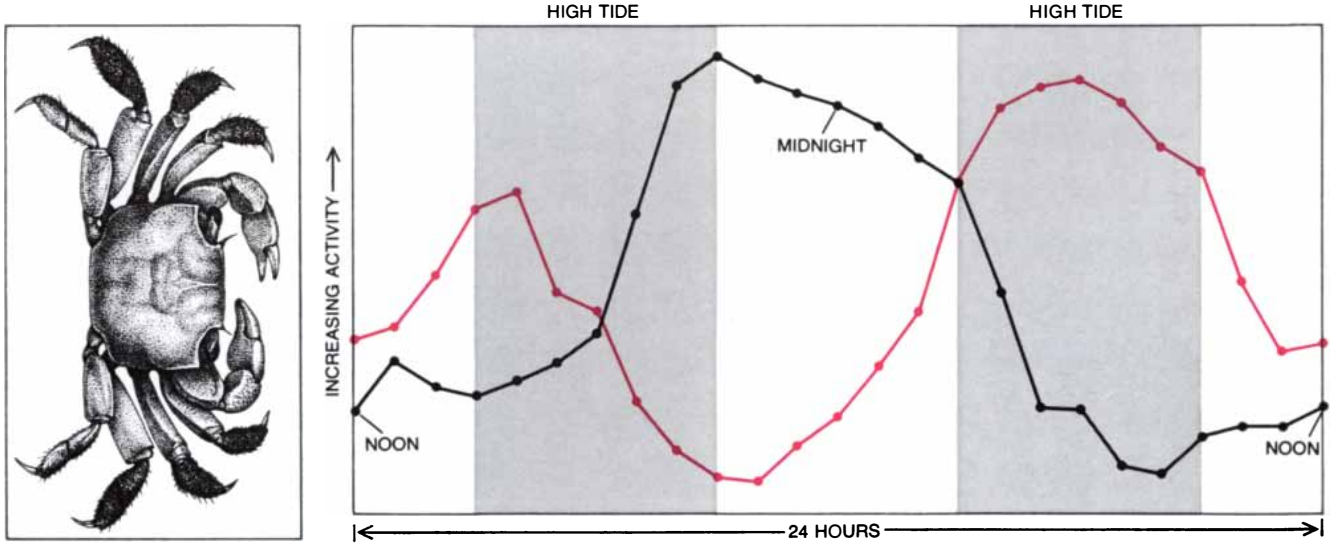


FIDDLER CRAB (*top*) is an inhabitant of tidal mud flats and sand flats. At high tide it remains quiescent in its burrow; at low tide it emerges to look for food. When a fiddler crab is taken from its natural habitat and put in an incubator where the light and temperature are constant, its periods of peak activity initially correspond to the times of low tide at its home location (*top curve*). The period of the crab's rhythm then begins to lengthen, and by the end of a week it is no longer synchronous with the times of the tide (*bottom curve*).

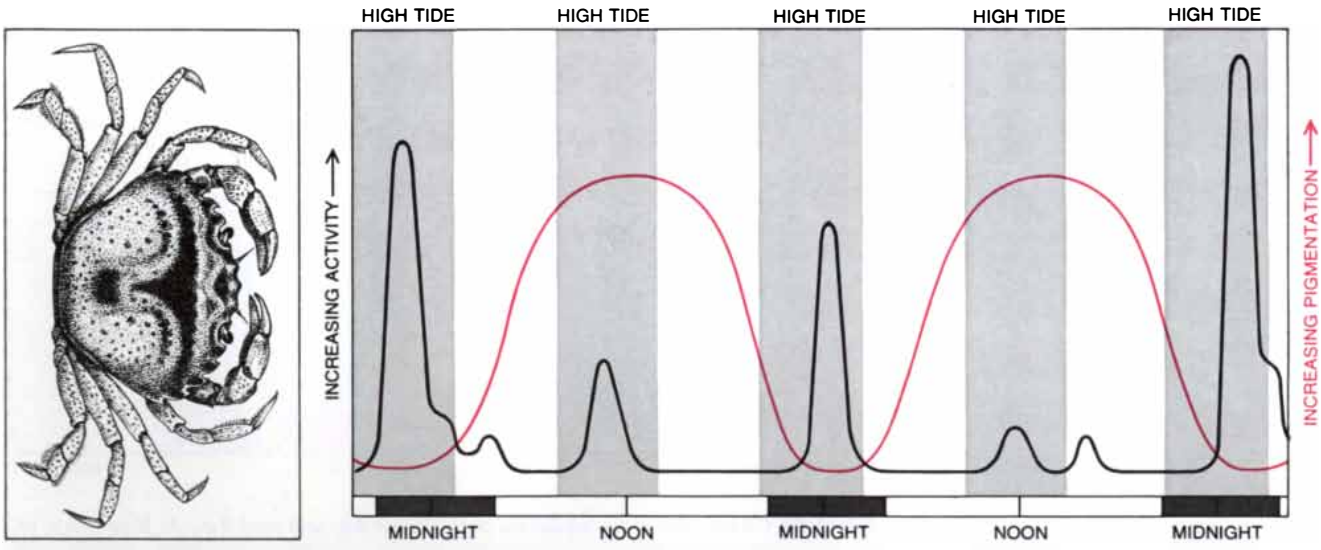
also been found to have a tidal component that gives rise to additional darkening at times of low tide.

The eyes of crabs are mounted on movable stalks. The stalks also house a neuroendocrine unit called the X-organ sinus-gland complex, which secretes a hormone that causes the pigments to disperse within the chromatophores. Powell found that the removal of the eye stalks from the green crab (and thus the X-organ sinus-gland complex as well) destroyed the color-change rhythm of the crab. Furthermore, Powell showed that the rhythm can be restored in a stalkless crab by implanting in it the stalk glands from another crab. These findings strongly suggest that in the green crab the eye stalks are the site of the clock that controls the color-change rhythm. In the penultimate-hour crab and the fiddler crab, however, removal of the eye stalks only reduces the amplitude of the color-change rhythms.

The neuroendocrine system of the eye stalks also exerts some control over the locomotor-activity rhythm of the green crab. When the eye stalks of green crabs were removed in experiments conducted by Naylor and Williams, all locomotor activity ceased, and it returned only gradually over the next six days. Since in green crabs the tidal rhythm normally vanishes after about a week in constant laboratory conditions, it was not surprising to find a lack of rhythm in the stalkless crabs when they resumed their activity. Attempts to reinstate the rhythm, however, by immersing the crabs in cold water were unsuccessful. On the other hand, crabs from which only the retinas were removed, not the stalks, returned to



PENULTIMATE-HOUR CRAB lives in close proximity to the fiddler crab on tidal flats. (The name of the crab is derived from the fact that the activity of a newly caught animal peaks one hour before midnight.) The black curve shows the mean solar-day activity of 30 penultimate-hour crabs over a period of a month, which is expressed as a broad peak of activity during the hours of darkness. The penultimate-hour crab also has a mean lunar-day activity rhythm (*colored curve*) that corresponds to the times of high tide.



GREEN CRAB, which also lives on tidal flats, has a basic pattern of locomotor activity that corresponds to the times of high tide. Its activity greatly decreases when the high tide comes during the hours of daylight. This pattern of activity continues when the crab is placed under constant conditions in the laboratory (*black curve*). In addition the green crab displays a solar-day rhythm in its body color, which blanches at night and darkens during the daylight hours. The color-change rhythm (*colored curve*), which persists when the green crab is kept under constant conditions, is thought to be controlled by a neuroendocrine system in crab's eye stalks.

their normal rhythm when they were immersed in cold water.

In further experiments Naylor and Williams found that subjecting the entire crab to a cold dip was not needed. If an arrhythmic crab is placed in water with a temperature of 15 degrees C. (59 degrees F.) but tethered so that its eye stalks protrude above the water, the tidal rhythm of the crab can be reinstated simply by dripping iced seawater onto its eye stalks for a brief period. Finally, Naylor and Williams made extracts from the eye stalks of green crabs in the quiescent phase of their locomotor-activity rhythm and injected the extracts into active stalkless crabs. The injection caused a significant reduction in the level of activity of the stalkless crabs, showing that there is an inhibitor substance that is periodically liberated from the stalk glands.

I have conducted similar experiments with the penultimate-hour crab. When the eye stalks were removed from these crabs, all locomotor activity stopped. I made eye-stalk extracts from rhythmic crabs during either the active phase or the quiescent phase of their locomotor activity and injected the extracts in various concentrations into crabs that had become arrhythmic because of long-term storage in constant conditions. No consistent alterations in the activity levels of the recipients were observed.

Since the neuroendocrine glands in the eye stalks of the penultimate-hour crab did not appear to be involved in the control of the crab's locomotor-activity rhythm, I carried out an experiment to determine if there was a chemical messenger coming from somewhere else in the crab's body. I joined two crabs, one strongly rhythmic and the other arrhythmic, by cutting small openings in their dorsal exoskeleton and cementing the openings together with sealing wax. In crabs most of the blood is not confined to vessels but flows freely through the spaces between organs; therefore when two crabs are joined, the blood of one mixes freely with that of the other.

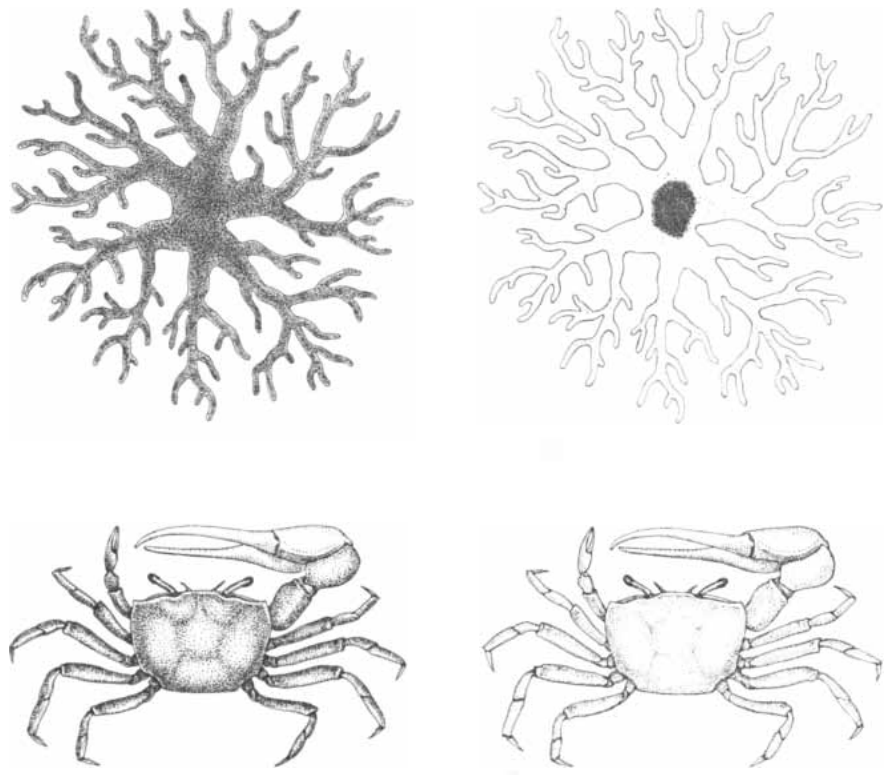
I also capitalized on an anatomical peculiarity of crabs, the process called autotomy. When a crab is attacked, the attacker usually grabs one of the animal's 10 legs. The crab's defense is to cast off the leg and dash away before the predator can grab another. The leg separates from the body at a predetermined breaking point. Excessive loss of blood from the open stump is prevented by a self-sealing mechanism, and the sacrificed leg is regenerated during successive molts.

Taking advantage of this self-amputation mechanism, just before I joined two crabs I made the rhythmic crab cast off all its legs. The joined crabs therefore consisted of an ambulatory arrhythmic crab on the bottom and a rhythmic amputee upside down on the top [*see illustration on next page*]. Any rhythmic locomotor activity recorded thereafter would have been the activity of the legged member, signifying that some substance in the blood of the legless crab had induced the rhythm. In 47 fusion-pair experiments not one rhythm was found. On the other hand, a rhythm was always displayed when two rhythmic crabs were joined in control experiments, indicating that the fusion procedure was not responsible for the lack of expressed rhythmicity.

It is clear, then, that whereas the endocrine system is involved in rhythms of color change and locomotion in the green crab, it is not necessarily involved in such rhythms in other crabs. Nor is it mandatory that an endocrine or a neural mechanism form the basis of any physiological rhythm, since a single-cell level of organization such as that found in the diatom *Hantzschia* is sufficient for the expression of all the known properties of clock-controlled rhythms.

The capacity for rhythmicity is not learned or impressed on organisms by the environment; it is the expression of a genetic potential. Heredity also determines whether the crab will be active at high tide or at low tide. This is not to say, however, that the environment does not play a significant role in the overt manifestation of a rhythm. It is the schedule of the tides on a particular stretch of coastline that determines the hour-to-hour settings of the rhythm. (The relation between a biological clock and the environment is similar to that between a pendulum clock and its owner. The rate at which the pendulum clock runs is determined by the escapement mechanism and the pendulum, but the owner can set the time to any hour by moving the hands on the face of the clock.) Thus a green crab will soon be active at high tide and a fiddler crab at low tide even when they are transported to an unfamiliar beach on a different ocean.

In the sand high on the beaches of southern California lives the sand hopper *Excirolana*. At the peak of each high tide, when the waters flood the habitat of this tiny isopod, it emerges from the sand to swim and feed in the breaking waves. Two or three hours later, when the tide turns, it burrows into the sand



CHANGE IN COLOR of the fiddler crab is the result of aggregation or dispersal of pigment granules in cells in the crab's hypodermis that are called chromatophores (*shown greatly enlarged at top*). In some chromatophores there are only dark pigment granules; in others there are white or orange granules. The fiddler crab blanches at night and darkens in daylight. Tidal component in the rhythm produces additional darkening at times of low tide.

and awaits the return of the next flood tide. James T. Enright of the Scripps Institution of Oceanography found that when he kept sand hoppers in a jar of seawater in constant conditions, they swam actively during the times corresponding to peak tide and remained in repose at the bottom of the jar at other times.

In southern California the tidal pattern changes greatly with the phases of the moon. Over a single month the tides change from one crest per lunar day to two per lunar day. Furthermore, during the transitions from one tide per day to two tides and back to one tide the height of consecutive tidal peaks also changes. L. A. Klapow, who was then working at the University of California at San Diego, showed that the pattern of the tides at the time sand hoppers were collected was reflected in the form of the activity rhythm the animals displayed in the laboratory [*see illustration on opposite page*]. In separate experiments Klapow and Enright also demonstrated that it is the pounding waves and the swirling waters that determine the pattern of the sand hopper's rhythm.

It therefore seems that inhabitants of beaches exposed to the open sea have their activity patterns shaped by the action of the surf. Intertidal organisms that live in protected bays are not normally exposed to a pounding surf and so we must look elsewhere for the elements that help to set their rhythms. The possibilities are numerous, including periodic inundation and periodic changes in temperature, hydrostatic pressure, the chemical composition of the water or the availability of oxygen. Of all these possibilities only two have been shown to play an important role. As a clear-cut example I shall cite another study of the green crab by the prolific team of Naylor and Williams.

One of their most surprising findings was that the principal feature of the tide, the periodic inundation of the shoreline, was not itself an important agent in synchronizing the locomotor-activity rhythm to the tides. This fact was demonstrated by bringing crabs into the laboratory and subjecting them for five days to 6.2 hours of immersion in seawater followed by 6.2 hours of exposure to air. The immersion in seawater was timed to correspond to low tide at the crabs' home beach, in effect reversing the animals' tidal schedule. The temperature of both the water and the air was held constant at 19 degrees C. After this treatment the crabs were placed in actographs, and their locomotor-activity patterns were measured for the next three days at the same constant temperature. The treatment did not rephase the crabs' rhythm. The procedure was repeated, but this time the air temperature was maintained at a level 11 degrees higher than the water temperature. Five days of this treatment did rephase the crabs' rhythm, and the change persisted in constant conditions.

The final version of the same experiment omitted the seawater-immersion portions of the cycle. Crabs were exposed to air at 13 degrees C. for 6.2 hours and then to air at 24 degrees for the same length of time. Complete and persistent synchronization resulted. Recently I have educed the same behavioral changes in fiddler crabs and penultimate-hour crabs. It is therefore the drop in temperature brought by the flood tides that plays an important role in setting the phase of the crabs' rhythm.

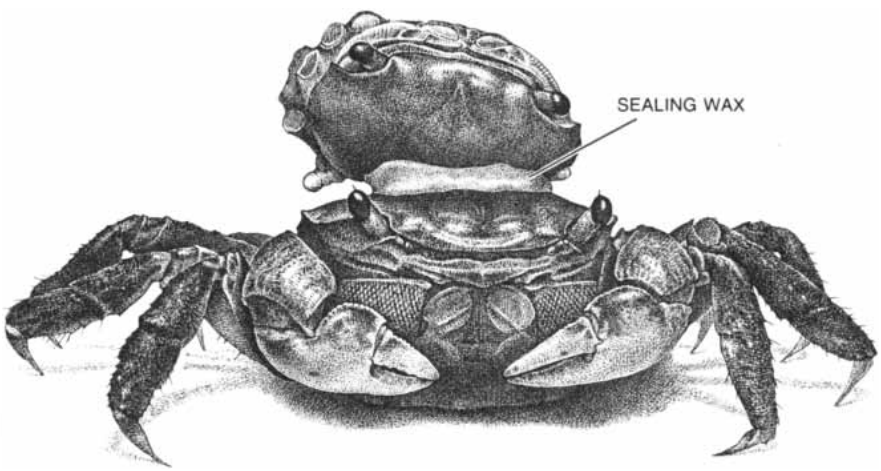
Hydrostatic pressure is the other environmental force that is known to synchronize organismic rhythms to local tides. In one experiment arrhythmic crabs were exposed for five days to a cycle of high pressure (1.6 atmospheres) for 6.2 hours followed by 6.2 hours at normal sea-level pressure. The crabs responded with an increase in activity during the high-pressure periods, and this periodicity persisted when the crabs were kept in constant conditions.

So far we have no firsthand knowledge of how the living horologe actually works. In the search for the elusive timing mechanism, however, several of its properties have been elucidated.

When rhythms in biological processes such as oxidative metabolism, photosynthesis and the like were first discovered, and it was found that these rhythms would persist without external stimuli, the controlling clock was thought to be simply some oscillatory step in the chain of chemical reactions underlying the process. As a result early attempts at locating the clock consisted in dissecting the chain of relevant reactions in the hope that the oscillatory segment could be identified. The rhythmic component was not found, and subsequent experiments showed that it probably does not exist. In fact, the clock is now known to be quite distinct from the process it makes rhythmic.

One of the many observations leading to this conclusion was conducted with the green crab. When the body temperature of the crab was lowered to 10 degrees C., all locomotor activity stopped for the duration of the chilling. When the body temperature was allowed to return to a normal level, activity resumed, and the locomotor rhythm was in exact phase with that of control crabs that had not been chilled. Clearly the crab's clock had continued to run accurately even when no rhythm was being expressed. This finding shows that the clock and the processes it causes to be rhythmic are separate and must be joined to each other in such a way that they can be uncoupled from each other and recoupled.

The disengagement of the coupling between the clock and the driven process may also be responsible for the even-



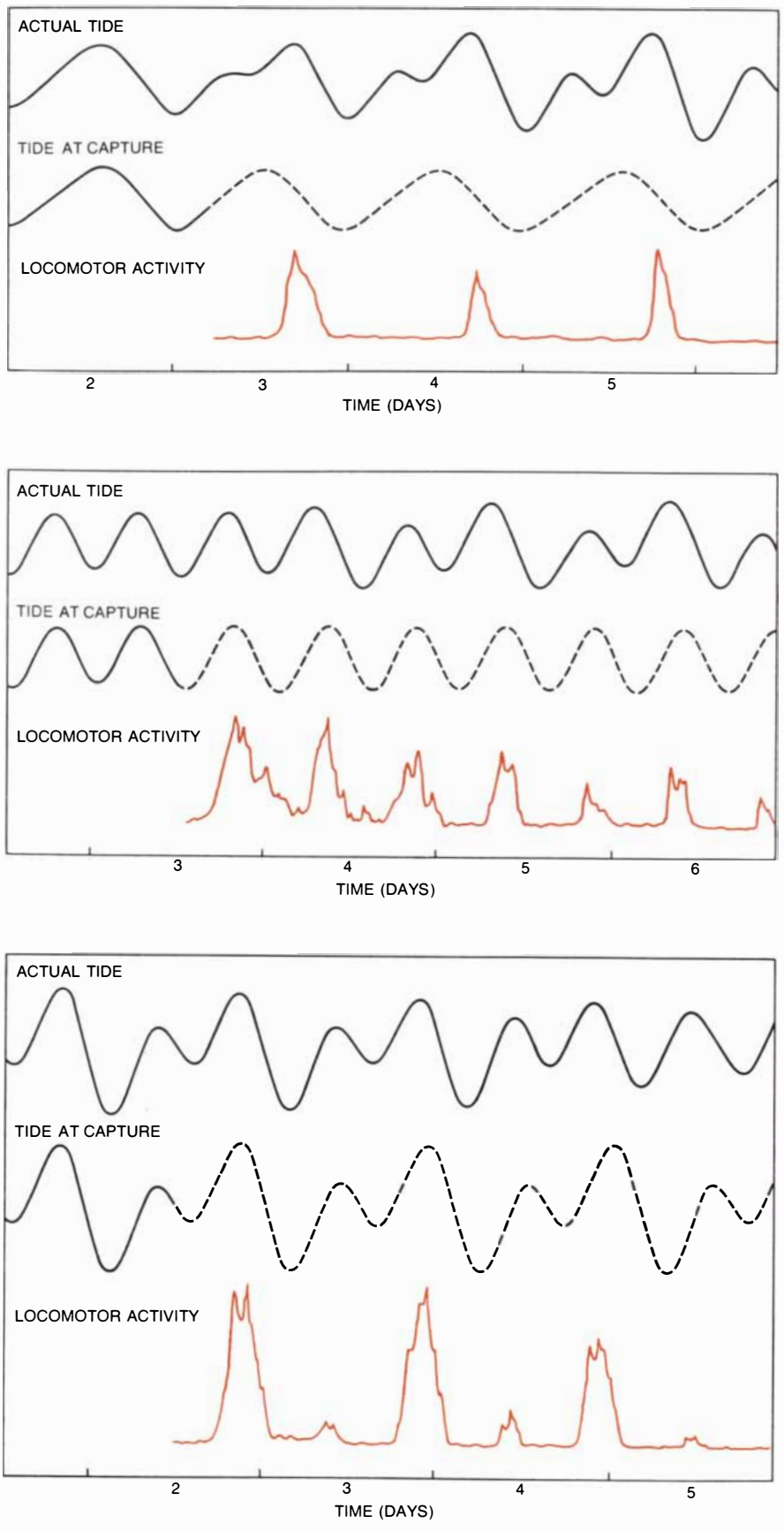
PENULTIMATE-HOUR CRABS ARE JOINED parabiotically so that their blood can continuously mix. Small openings are cut in the dorsal exoskeleton of each crab, and the openings are cemented with sealing wax. The top crab has cast off all its legs through autotomy, the process by which crabs shed a leg when it is seized by a predator. Before the union the locomotor activity of the top crab was synchronous with the tides. The activity of the bottom crab was arrhythmic. In experiments with fused crabs no rhythmic locomotor behavior was found, indicating blood does not contain a chemical messenger that induces such behavior.

tual loss of overt rhythmicity in animals that are maintained in constant conditions in the laboratory. Speaking somewhat teleologically, when intertidal animals are taken away from their tidal environment, there is no longer any pressure for them to maintain a tidal rhythm. Their life processes are emancipated from the clock and become arrhythmic.

The validity of the notion of a coupling between a clock and vital processes is enhanced by the fact that rhythms once lost by crabs, either in the laboratory or in nontidal natural habitats, can be reinstated by a single short-duration stimulus such as being chilled. Since the treatment provides no information about tidal intervals, the simplest interpretation is that the stimulus recouples the clock, which had continued to run, to the processes governing locomotor activity, causing such activity to become rhythmic again.

As we have seen, the vertical migration rhythm of the diatom *Hantzschia* demonstrates that a biological clock needs only the level of organization characteristic of a single cell to express itself. Two other unicellular organisms provide even better examples. The marine dinoflagellate *Gonyaulax* is known to simultaneously display different rhythms in four processes: photosynthesis, luminescence (it glows at night), irritability and cell division. Five different rhythms have been detected in the single-celled green alga *Acetabularia*, and all the rhythms persist even when the nucleus of the cell has been removed by microsurgery. There is evidence that in multicellular plants and animals the clock is also to be found in single cells. When organisms are subdivided and the parts are kept alive in tissue culture, the cells continue their original rhythm. Indeed, the plausible place to look for the living clock is within the single cell, where one would expect to find it in the form of some physiochemical entity. In spite of intensive investigation, however, neither the clock nor any of its components have been located. The search has nonetheless revealed two unusual aspects of the horologe: the rate at which it runs is almost completely insensitive to temperature, and the rate also is not affected by a wide variety of potentially disruptive chemical agents.

In general increasing the temperature increases the rate at which chemical reactions proceed. One would expect that the living horologe, with its chemical clockwork, would be accelerated in a similar way. To test this assumption we subjected groups of crabs to increasingly higher constant temperatures in the lab-



PATTERN OF ACTIVITY OF THE SAND HOPPER (also called the beach flea) found in the beach sands of southern California is adapted to the peculiarities of the tides of the region. The tides alternate every month between one peak per lunar day and two peaks per lunar day. When sand hoppers are kept in constant laboratory conditions, their pattern of activity (*colored curves*) tends to mimic the form of last tidal pattern to which they were exposed (*broken curves*) rather than the form of the actual tidal pattern (*black curves*).

SCIENCE/SCOPE

A new ultra-lightweight radio for tactical field operations, developed by Hughes, employs micro-miniaturized circuits including LSI (Large Scale Integration) to provide high reliability, plug-in modules for easy maintenance, and an AM mode for compatibility with current military systems. Called the HC-191 Manpack, it is a version of the AN/PRC-104 single-side-band transceiver Hughes is building for the U.S. Marine Corps. It has a frequency range of 2 to 30 MHz and 280,000 channels to make enemy jamming difficult. Another significant combat advantage is its completely silent automatic electronic tuning.

The complete Manpack radio weighs only 12½ pounds including a battery pack that gives 16 hours of service before recharging. With its built-in 8-foot whip antenna, the HC-191 has a range of up to 30 miles in the most difficult jungle or mountain terrain. For a copy of the HC-191 brochure, write: Marketing Department, Hughes Aircraft Company, Bldg. 600/C231, P.O. Box 3310, Fullerton, Calif. 92634.

Telephone users in the United Kingdom will benefit from the computer-controlled FACT-II wiring analyzer system recently delivered to Standard Telephones and Cables, Ltd., of International Telephone and Telegraph in Northern Ireland. The 68,000-lb. system can test 34 different products in any of over 5,000 electrical configurations. A special connector developed by Hughes makes it possible to simultaneously access 25,600 circuit terminations in less than 15 seconds.

FACT-II is an adaptation of the system Hughes originally developed to pinpoint and troubleshoot electrical problems in aircraft fire-control systems. Hughes has built 11 FACT systems for European users and scores more for North America, the Middle East, and Japan.

Laser rangefinders for the U.S. Army's M-1 battle tank are being developed by Hughes for prototypes by both Chrysler Corp. and General Motors Corp. Following a competitive evaluation in mid-1976, the Army is expected to select a single contractor. Hughes currently produces laser rangefinders for the Army's M60A2 tank and M551 Sheridan vehicle and is developing a full-solution laser fire control system for an improved version of the M60A1. A tank with a laser rangefinder can fire far more quickly and with a much higher first-round hit probability.

Hughes Ground Systems Group needs Senior Systems Programmer/Analysts and Communication Systems Development Engineers to join the technical staff due to growth of current research and development programs. Applicants must have a BS or MS in electrical engineering or computer science and U.S. citizenship. Qualified applicants should write or send resume to: M. F. Duggins, Hughes Aircraft Company, P.O. Box 3310, Fullerton, CA 92634. An equal opportunity M/F employer.

A solid-state watch module for ladies-size digital watches -- now in production at Hughes -- contains the equivalent of more than 1,500 transistors. It overcomes the size limitations of ladies' watches with a unique time-readout that flashes the hour for about a second, then gives the minutes. The new module supplements the men's watch modules now made for leading name-brand manufacturers by Hughes, one of the largest producers for the watch industry.

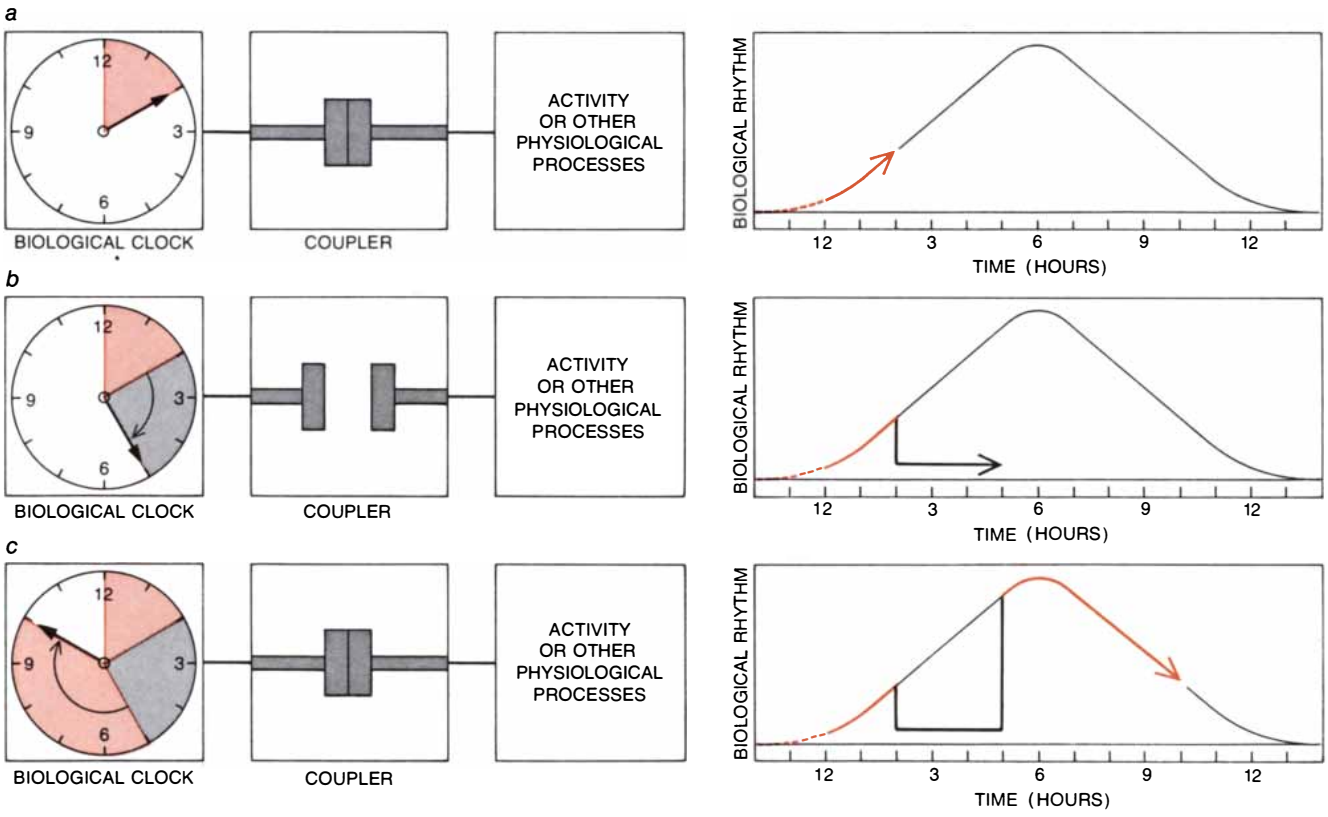
oratory and observed their rhythms over periods of several days. Since the period of an expressed biological rhythm is believed to closely mimic the driving frequency of the clock, a temperature-induced change in the rhythm is assumed to indicate a change in the frequency of the clock. The usual result obtained in such experiments is that there is no change in period at all. If any change is recorded, it is only a fraction of what one would expect from a chemical system.

Attempts to disrupt the rhythms of organisms with chemical substances such as inhibitors of protein synthesis, stimulants, metabolic inhibitors and narcotizing agents have proved to be almost equally futile. Out of hundreds of substances tested only four—deuterium oxide, ethyl alcohol, valinomycin and lithium ions—have been found to alter the period of a rhythm. In view of the variety and number of substances screened, it appears that biological clocks, unlike most other pacemaker systems in organisms, are virtually immune to chemical manipulation.

From a pragmatic point of view these insensitivities might have been predicted. Certainly one of the most important attributes of any clock—living or man-made—is accuracy, and a clock whose rate of running is altered by changes in the temperature or chemistry of its environment would not meet this requirement. In fact, if the clock responded to every change in ambient temperature, it would not be a clock at all but rather a thermometer that signaled ambient temperatures by the rate at which it ran.

The accuracy of biological clocks is even more amazing when one takes into account the fact that precision must be maintained during cell division, when presumably not only the cell but also the clock is replicated. The ease with which this replication is accomplished has been demonstrated in a study of the single-celled protozoan *Paramecium* by Audrey Barnett of the University of Maryland. In the strain of paramecium she worked with the sex of each animal changes from one mating type to another and back again each day. Barnett placed in constant darkness eight paramecia whose clocks had been set to regular day-night cycles. The cells in the population divided 2.2 times per day, and at the end of six days they had given rise to slightly more than 121,000 cells. On the seventh day the sex-reversal behavior of the entire population was examined and was found to be rhythmic. The phase of the rhythm was close to that of control cells that had remained under the regular day-night cycle. Since only the original eight cells had been subjected to a day-night cycle, it appears that each of the original cell clocks had been replicated time and time again with very little loss in accuracy. An alternative interpretation is that each cell contains many clocks, some of which are replicating themselves while others are still coupled to cell processes, causing them to be rhythmic.

We still know very little about the mechanism of living horologes in the tidal zone, and the properties of clocks that have been elucidated in some ways compound the problem. The continued search for such mechanisms is nonetheless a worthwhile endeavor because clock-controlled rhythms are found not only in intertidal organisms but also throughout the kingdom of life. One may hope that the continued effort will eventually lead to discoveries that will enable us to perceive the fundamental principle that underlies the operation of all biological clocks.



COUPLER that has not yet been identified is believed to join biological clocks and biological processes. When the coupler is engaged (*a*), the biological rhythm is expressed. The disengagement of the coupler (*b*) may be responsible for the loss of rhythmic behavior of organisms that are kept under constant laboratory conditions. The clock continues to function, however, and when recoupling occurs (*c*), the rhythm takes up not at the point where it left off but at point that corresponds to "time" determined by the clock.

A CARTHAGINIAN FORTRESS IN SARDINIA

Before her fatal clash with Rome, Carthage dominated the western Mediterranean. A strongpoint on Monte Sirai in Sardinia emerges as part of a network with which she controlled the entire island

by Sabatino Moscati

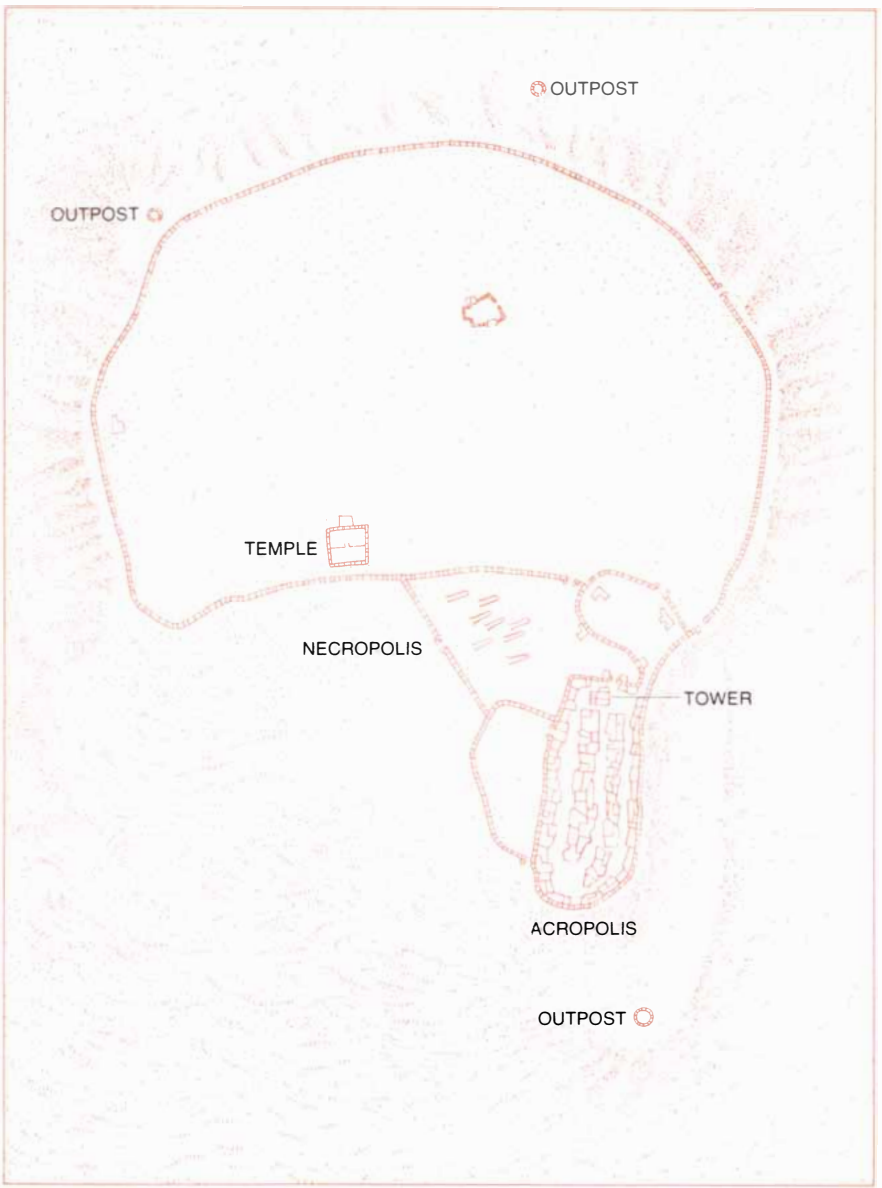
Carthage is best known as the great rival of Rome in the last centuries of the first millennium B.C. What is less generally realized is that centuries before the rise of Rome the Phoenicians who founded this prosperous African city-state wrested control of the western Mediterranean from the Greek colonists of Sicily and southern Italy. At its apogee Carthage held, in addition to its African possessions, Malta, part of Sicily, most of Sardinia, the Balearic Islands off the east coast of Spain and a number of outposts on the Spanish mainland. So great was the naval power of Carthage that in 509 B.C. the young Roman republic was glad to join the Carthaginians in an agreement whereby the western Mediterranean was left to the Carthaginians to exploit.

Sardinia at that time was a rich island:

ENTRANCE TO ACROPOLIS is dominated by the foundations (*center*) of the great tower that was built at the highest point on Monte Sirai. Rectangular in plan, the Carthaginian tower stood on the site of a prehistoric Sardinian nuraghe, or circular tower.

it was a major granary and an important source of ores. It has been traditional to think of the Phoenicians and Carthaginians as being primarily men of commerce, mercantile voyagers whose interests seldom extended beyond the harbors where their trading posts were established. The recent discovery of a Carthaginian military stronghold in southwestern Sardinia, however, has set in motion archaeological investigations that have painted a quite different picture. It is now increasingly evident that not just a few coastal towns but the entire island was under active Carthaginian control during the centuries preceding the Punic Wars between Carthage and Rome.

The Carthaginian site, Monte Sirai, has now been intensively studied for four years. It was discovered in 1962 when a youth from the nearby town of Carbonia went mushroom-hunting in the underbrush that covers the flat summit of the 200-meter hill. To his surprise he came on a weathered slab of stone that bore the figure of a woman carved in relief. His find was reported to the Cagliari superintendent of antiquities, Gennaro Pesce, who sent his inspector, Ferruccio Barreca, to investigate the site. Barreca's reconnaissance revealed the outlines of a small fortress settlement on the hilltop, and the Department of Antiquities asked my institution, the Institute of Near Eastern Studies at the University of Rome, to collaborate in a series of joint excavations. Superintendent Pesce and I undertook the general direction of the project; the excavations were conducted by Barreca, who has now succeeded Pesce as superintendent at Cagliari.

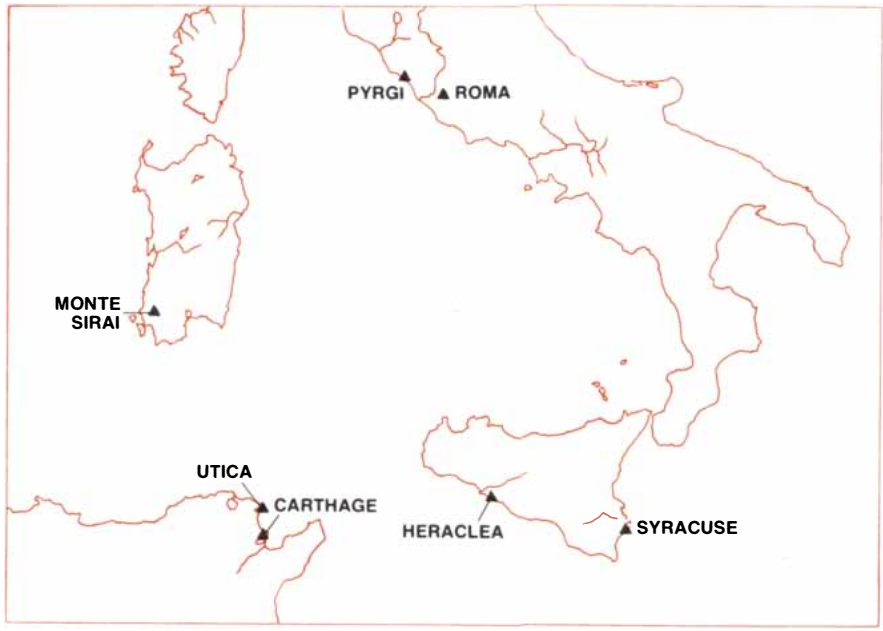


TOP OF MONTE SIRAI was divided into several walled enclosures. Within the acropolis, which was dominated by the rectangular tower, were numerous buildings arranged along streets. Within the necropolis were underground burial chambers and stelae, shafts erected as monuments to the dead. Outside the walls three defensive outposts have been found.

The fieldwork at Monte Sirai is now all but complete, although the restoration and cataloguing of the materials unearthed there are still in progress. The site has proved to be an important one for several reasons. First, although it lies only a few kilometers inland (and quite near the little island of St. Antiochus, the site of the ancient Carthaginian town of Sulcis), Monte Sirai was plainly intended as a stronghold to control travel by land from the southwest coast into the interior of Sardinia, in particular the rich plain of Campidano. Almost all the other known Carthaginian and Phoenician settlements on Sardinia were located at coastal sites; like similar settlements elsewhere in the region, they functioned as entrepôts: combined warehouses and trading posts connecting the sea routes used by trading vessels. The commanding position of Monte Sirai with respect to land travel thus raised a question of considerable historical importance: To what extent did Carthage control the interior of Sardinia during the centuries preceding the wars with Rome?

Unlike stony Corsica to the north, fertile Sardinia had supported a substantial population during the Bronze Age (which in the western Mediterranean roughly spanned the whole of the second millennium B.C.). Persuasive evidence pointing toward this fact is the large number of nuraghi, defensive towers built of rough-hewn stone on a circular plan, usually some 10 meters in diameter, that are found on the island [see "The Proto-Castles of Sardinia," by Giovanni Lilliu; SCIENTIFIC AMERICAN, December, 1959]. Some 6,000 nuraghi have been identified on Sardinia, the ruins of a number of them still rising 10 meters or more above the ground; they bespeak a troubled period when the inhabitants believed local strongpoints were at least desirable and perhaps even essential to survival. Even in the time of the nuraghi the strategic position of Monte Sirai must have been realized; the nucleus of the later Carthaginian fortress, a massive rectangular tower, was built on the foundations of one such round tower.

The tower rose within an acropolis, the highest and most heavily fortified part of the Carthaginian outpost. Before I take up the various components of the



WESTERN MEDITERRANEAN ABOUT 500 B.C. was an arena of conflict pitting various Greek communities on Sicily and in southern Italy against the Carthaginians, who held much of Sardinia and Sicily and had a foothold in the Balearic Islands and in Spain. At this time Rome, although only one of several independent states in Italy, was on the rise.

settlement and the artifacts uncovered in each of them, let me present a general sketch of the hilltop fortress. To begin with, the level ground at the top of Monte Sirai was entirely enclosed by a massive stone wall that rose from the bedrock. To the west, north and east the wall ran parallel to the steep edge of the hilltop; to the south the wall ran along a gentler slope that gives easier access to the summit [*see illustration on preceding page*]. Defensive posts were distributed outside the perimeter wall; the remains of several small towers have been found on the hillside.

The acropolis stood on the south side of the hilltop, enclosed partly by the perimeter wall and partly by a wall of its own. At the north end of this oblong enclosure was a gate flanked by two small towers, and just within the gate rose the large rectangular tower whose foundations incorporated the nuraghe. In a depression north of the acropolis, enclosed in part by the perimeter wall and in part by its own inner bounds, was a necropolis where the settlement's tombs had been cut into the bedrock. Just northwest of the necropolis stood the settlement's temple and in close association with it the open-air tophet, or burning ground, that was a fixture of Phoenician religious ritual. The rest of the fortified enclosure on the hilltop was at least 20 times larger than the acropolis and occupied the remainder of the hilltop; in this parklike area only two building foundations have been excavated, although traces of others are apparent.

The rectangular tower is only one of the structures within the acropolis. Roughly parallel rows of modest but readily identifiable foundations that extend southward from the tower seem to have been separated by streets that are not yet traced in detail. The foundations probably represent living quarters reserved for the officers of the garrison and their families. It was the Carthaginian tradition to hire mercenary soldiers and sailors but to place them under Carthaginian officers. Both the mercenaries stationed at Monte Sirai and whatever native Sardinians were employed by the garrison would presumably have been housed outside the acropolis, and except when they were on duty, they may have preferred the advantages of town life in nearby Sulcis.

The acropolis tower shows the traces of successive architectural additions. One of these is a sacellum, or small sanctuary. Excavation of the sanctuary courtyard has yielded a number of votive objects. Perhaps the most important is a female figure sculptured in stone; stone sculpture is rare in Carthaginian art. The body of the figure was blocked out roughly; the torso is legless and the arms are spindly. The head, however, was carefully rendered. The dangling curls are stylized, and so are the ears, which are represented as concentric circles. The overall style suggests direct contact with Near Eastern sculptural traditions.

Another rare object from the tower sanctuary is a representation of a man's face in the form of a mask. The hair is rendered in stylized curls, and the long, narrow beard is marked by a vertical groove [*see illustration on page 84*]. Nothing like this sculpture has been found in Sardinia before, although two similar masks are known: one from Carthage and one from Utica, another Phoenician colony in Africa. It has been suggested that the Monte Sirai mask was an import from Africa. The elegance and fine craftsmanship of the mask do not, however, preclude its having been made in Sardinia and even by native Sardinians. The Carthaginian art traditions were strong enough to make it difficult to draw a distinction between articles produced in one or another sector of the Carthaginian sphere of influence. Moreover, as a number of other finds at Monte Sirai attest, the level of local workmanship was high.

An example is provided by three small bronze figures unearthed at the site. The Sardinians began to make bronze statuettes during the eighth century B.C., that is, at about the same time the first Phoenician entrepôts were established on the island. Thus it seems probable that, whereas Sardinian work in bronze eventually evolved independently, the Phoenicians and their Carthaginian successors must have exercised some influence on the rise of the technology. This influence appears to be what we see in the Monte Sirai bronzes. One is a small quadruped mounted on a ring; another is a seated musician playing a lyre. The most elaborate of the three is a figure, also seated, pouring from a pitcher held in his left hand into a cup held in his right hand [*see bottom illustration on page 85*]. Stylistically all three figures are certainly connected with art traditions that flourished in Phoenicia and neighboring Near Eastern areas during that period. Nothing, however, precludes their having been manufactured by native Sardinian bronzesmiths.

Several of the underground tombs in the necropolis are quite well preserved; in architecture they follow the Phoenician and Carthaginian tradition. A dromos, or corridor, leads down to an entrance that in most cases is sealed off with a slab of stone. The rock-hewn room within the entrance is square or rectangular, with the recesses that served as beds for the dead cut into the tomb walls. Although the architectural tradition is imported, local touches (not

to say misunderstandings) are evident. Consider one of the Monte Sirai tombs where the masons left a pillar of bedrock standing in the center of the tomb chamber. Cut in bold relief on the pillar is the sign of Tanith, a common Carthaginian religious symbol consisting of a circle resting on a horizontal bar that is balanced on the apex of a triangle. The Sardinian mason who was commissioned to execute the symbol at Monte Sirai knew the components of the sign but evidently did not understand their meaning: he put the circle at the bottom and the triangle at the top.

Evidence in the form of fresh chisel marks indicated that recent tomb robbers removed a relief sculpture from this same chamber. Fortunately the stolen work was quickly recovered. It is a human head, so extremely stylized that it most closely resembles a tall inverted triangle. Most of the facial details appear in negative relief. The eyes are mere slits; the nose is defined by a pair of deep gouges and the mouth by a wider gash. The stone bears traces of red coloring. Although the sculpture is stylistically related to other representations of female heads found at Monte Sirai, it is strikingly individualistic in its appearance [*see illustration on cover of this issue*].

The burial chambers yielded a number of pottery vases, some still intact, made in shapes characteristic of Phoenician and Carthaginian ware. The necks of some vases are decorated with boldly modeled relief representations of animal and human heads. Also found in the tombs were amulets made of glass paste and scarabs. (The Phoenicians bowed to Egyptian suzerainty more than a few times during their long history.) One of the scarabs depicts the figure of a warrior in an attitude that seems to have been derived from the Greek artistic tradition.

Just northeast of the necropolis stood the temple, a rectangular structure with several rooms. The chief feature of the innermost room, which was perhaps the most sacred, was an oven where the excavators found ashes and fragments of bone. The oven had originally been rectangular, but successive rebuildings had altered the plan to a circular one. This alteration I take to be the result of the Carthaginian settlers' eventually being replaced by the local Sardinians; it is an exact reversal of the sequence at the tower, where a circular Sardinian nuraghe served as the foundation of the rectangular Carthaginian structure.

Among the artifacts recovered in the temple area was a male figurine modeled in terra-cotta. The body of the figurine, only partially preserved, is cylindrical at the top and bell-shaped at the bottom. The facial features include rounded, bulging eyes, protruding lips and a prominent pointed beard. Figurines of this type are known from both Phoenicia and Cyprus; once more it is evident that there were direct ties between the art of Monte Sirai and the art of the Near East.

As I have noted, the tophet, the open-air burning ground where human sacrifices were made to the gods, is directly attached to the temple at Monte Sirai. The juxtaposition is unique; up to now no such association has been found at

TRADING POSTS established first by Phoenicians and later by Carthaginians at coastal sites on Sardinia (*triangles*) have long been known to historians. Discovery of the fortress at Monte Sirai, inland from Sulcis (*bottom left*), has led to further reconnaissance in Sardinia and to identification of many previously unknown Carthaginian sites (*diamonds*). New evidence indicates that Carthaginians held firm control of the island for centuries.

a Carthaginian site. Moreover, the ash-filled oven within the temple indicates that human sacrifices were performed not only on open-air altars, as they were at other Carthaginian tophets, but also inside a religious building. Buried in the precincts of the tophet were urns containing the bones of the sacrificed. They are all the bones of children. A number of sculptured stelae, or stone pillars, that recorded the sacrifices also came to light. Similar evidence of sacrifices is usually found within this kind of sacred enclosure; indeed, the number of sacrifices at Monte Sirai seems to have been lower than usual. This can be understood if we take into consideration the fact that Monte Sirai was a fortress and not a town.

Most of the stelae bear the same design, beginning at the top with an Egyptian motif: a frieze of uraei, or cobralike Egyptian asps, above a winged disk of the sun. Supporting this sculpture is a pair of pillars, their capitals decorated with spirals, and standing between the pillars is a female figure dressed in a long, plain garment, staring straight ahead, her folded arms cradling a large disk against her breast. The design is an almost exact imitation of the one found on stelae at the nearby settlement of Sulcis; the parallel adds weight to the assumption that it was from this trade center, first Phoenician and later Carthaginian, that the garrison of Monte Sirai was drawn.

Other stelae, however, are considerably more simplified and probably represent the work of local craftsmen. The niche in which the figure stands loses its surrounding decorations or disappears altogether. For example, one of these stelae has no niche at all. The figure of the woman, carved in relief, is extremely stylized, and rather than clasping a disk she is apparently supporting a child on her hip. At first the simple design was thought to be a crude representation of the act of offering a child for sacrifice. In my opinion it more likely represents a common art motif borrowed from the Greek world: a woman holding an infant in her arms.

TERRA-COTTA MASK representing a bearded man shows the head covered with a series of stylized curls. Masks like this one have been found both at Carthage and at nearby Utica.

So far, although I have mentioned early Phoenician influences and later Carthaginian ones at Monte Sirai, I have not put the findings at the site in any precise chronological framework. Stratigraphy and pottery show that it was first garrisoned by the people who settled Sulcis in the seventh century B.C. Therefore the initial occupation would have been Phoenician; the Carthaginians did not replace their Near Eastern predecessors in Sardinia for another century or so. Some centuries later, at the conclusion of the Second Punic War (218–201 B.C.), Rome formally took possession of Sardinia after having had effective control of the island for some 40 years. There is evidence at Monte Sirai that the hilltop stronghold was briefly occupied by a Roman garrison before it was finally abandoned, probably during the first century B.C.

In summary, Monte Sirai during the centuries before the Punic Wars was the seat of a Carthaginian garrison that controlled roads giving the Carthaginians access to the interior of Sardinia. At the same time the fortress acted as a defensive outpost guarding the coastal entrepôt of Sulcis. As we have seen, the existence of such an inland garrison has naturally raised questions about the extent and character of Carthaginian rule in Sardinia. The series of archaeological reconnaissances and soundings undertaken in order to find answers to these questions has proved enlightening.

To begin in the vicinity of Sulcis, it appears that Monte Sirai was by no means the only Carthaginian fortress in the interior of southwestern Sardinia. There were others at Monte Crobu, Corona Arrubia and Pani Loriga; moreover, the military settlements were all much the same in size and layout. A systematic campaign of excavation has been started at Pani Loriga, and already it seems reasonable to say that the Carthaginians maintained a well-organized network of military garrisons in this part of the island.

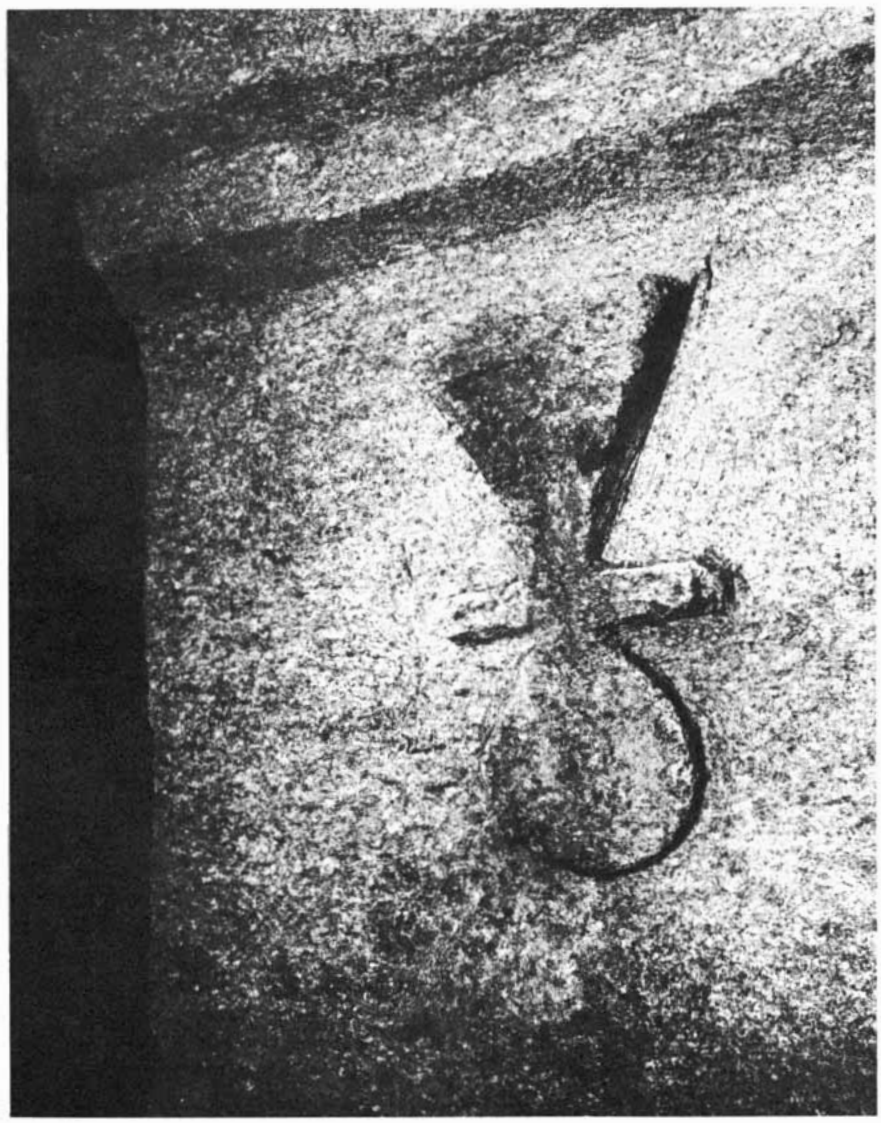
It has also been possible to trace a line of fortresses in the mountains inland. The Carthaginians maintained defensive positions at San Simeone di Bonorva, Sedilo, Neoneli, Fordongianus,

PAIR OF STELAE erected in the tophet, or burning ground, on Monte Sirai to commemorate sacrifices exemplify the contrast between Carthaginian design and native Sardinian art. The woman cradling a disk in her arms (*left*) stands under a frieze made up of Egyptian motifs: a row of snakes' heads above a winged disk of the sun. This traditional design is also found on stelae at Sulcis, a nearby Carthaginian trading settlement. The other stele (***right***) shows a much simpler figure of a woman; stylized rectangle at her side evidently represents a child resting on her hip. The iconography is probably that of a Greek art theme: mother with child.

THREE BRONZE FIGURES found in a small sanctuary attached to the acropolis tower are (*from left*) a seated lyre player, the image of an unidentified quadruped mounted on a ring and another human figure filling a cup from a pitcher. The Sardinians knew the art of bronzesmithing by the time the first Phoenicians visited the island, so that native metallurgists may well have cast the three figures. Stylistically, however, all three represent a Near Eastern art tradition that implies foreign supervision of the work.

Genoni and Ballao [*see illustration on page 83*]. Like the fortresses near Sulcis, these outposts have features in common with Monte Sirai. All are built on hills with a flat summit, are enclosed by a wall and have an acropolis and a tower at the highest point of the summit. Moreover, each location commands an adjacent valley and a fordable stretch of river. The six mountain-chain sites were evidently established between the fifth century B.C. and the time during the third century when the First Punic War began (265 B.C.).

Concurrently with the reconnaissance of the Sardinian interior a reevaluation has been made of the Carthaginian presence along the west coast of the island. There a number of Carthaginian centers had long been recognized, but the entrepôts had appeared to be isolated from one another. Now other Carthaginian buildings and harbor works have been found from the city of Bithia northward. They are present at Cape Malfatano and Cape Teulada. Newly identified Carthaginian ruins in the Sulcis area include buildings on the island of San Pietro, in the vicinity of Porto Botte and at Porto Pino Bay, where a wide canal runs inland from the sea. Farther north, in the Iglesias region, ruins that are apparently of Phoenician origin have been found at Cape Frasca. When we add to this list the Carthaginian cities already known along the west coast of Sardinia, such as Tharros and Bosa, and the others brought to light by earlier reconnaissances in the vicinity of Alghero Bay and in the Sassari region, we can only con-



SIGN OF TANITH, a common Carthaginian religious symbol, was carved in relief on one of the pillars left standing in an underground burial chamber in the Monte Sirai necropolis. The carving appears to have been made by a native Sardinian unfamiliar with the foreign garrison's religious traditions: the circle, which is normally the uppermost element in the sign, was placed at the bottom and the triangle, normally at the bottom, was at the top.

clude that the Carthaginians once colonized all western Sardinia.

If this can be said for the west and the interior of Sardinia, what about the eastern part of the island? At Cape Carbonara the remains of Carthaginian waterfront construction have emerged. Farther north, at Villaputzu and San Giovanni di Saralà, well-preserved walls of Carthaginian buildings have been located. From the Colostrai Valley comes a milestone engraved with Carthaginian characters, and the ruins of a settlement with important architectural remains have been discovered at Cala Gonone. A famous Carthaginian entrepôt, Olbia, is the northernmost settlement now known on the island's east coast. When all these traces are taken together, a pattern of settlement emerges that, although it is not to be compared to the more numerous settlements along the west coast, is nevertheless impressive.

The answer to our question, then, is generally as follows. Starting from modest Phoenician coastal posts established in the ninth and eighth centuries B.C., the Carthaginians came to control the entire island of Sardinia by the fourth century B.C. The extent of Carthaginian influence around the Tyrrhenian Sea is independently indicated by the discovery at the site of Pyrgi, on the west coast of the Italian peninsula, of the famous gold plates made for Thefarie Velianas, the ruler of Etruscan Caere, about 500 B.C. The long inscription on the plates is written as a double text in Etruscan and Phoenician characters. Both the findings at Monte Sirai and the further archaeological work in Sardinia they inspired enable us to paint a much more detailed picture of circumstances in Sardinia preceding the fateful clash between Carthage and Rome.

The significance of these recent discoveries in terms of political history must not, however, be allowed to overshadow their equal importance in terms of art history and the history of civilization. The contact between the Phoenicians and Carthaginians on the one hand and the indigenous population of Sardinia on the other gave rise to a striking folk art; the Phoenician contact in particular left a legacy of Near Eastern art motifs that could not have been imagined by any scholar of ancient Italian art only a few years ago. The events that excluded the Greeks from this corner of the Mediterranean made Sardinia unique as the only part of what is now Italy where the culture and civilization of Greece did not overwhelm all that had existed earlier.

Alfred Wegener and the Hypothesis of Continental Drift

Sixty years ago a German scientist argued that the continents move, and he proposed a history of their migrations. The validity of his theory was not recognized until new evidence emerged in the 1960's

by A. Hallam

The jigsaw-puzzle fit of the coastlines on each side of the Atlantic Ocean must have been noticed almost as soon as the first reliable maps of the New World were prepared. The complementary shapes of the continents soon provoked speculation about their origin and history, and a number of early theories suggested that the shapes were not the product of mere coincidence. In 1620 Francis Bacon called attention to the similarities of the continental outlines, although he did not go on to suggest that they might once have formed a unified land mass. In the succeeding centuries several other proposals attempted to account for the correspondence, usually as a result of some postulated catastrophe, such as the sinking of the mythical Atlantis. The first to suggest that the continents had actually moved across the surface of the earth was Antonio Snider-Pellegrini in 1858, but he too attributed the event to a supernatural agency: the Great Flood. Today, of course, the migration of the continents is an essential feature of the theory of the earth's structure that is all but universally accepted by geologists.

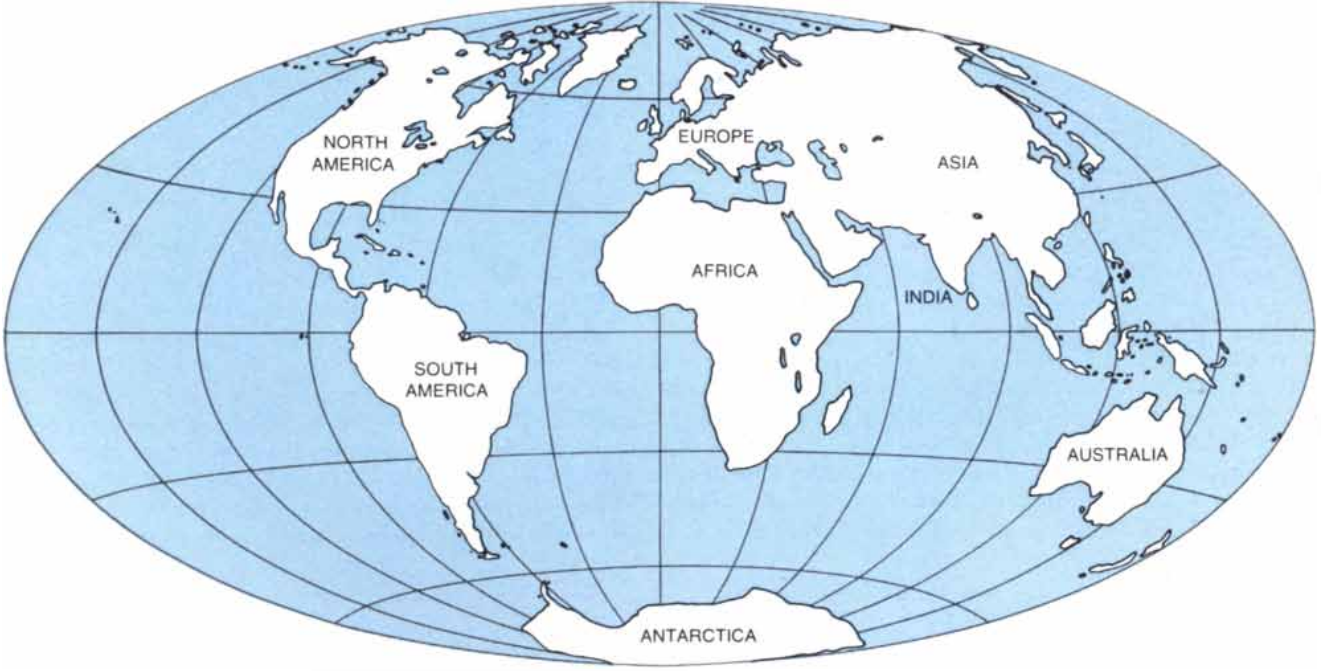
Of the various hypotheses that preceded the modern theory of plate tectonics, one version stands out: the one propounded by Alfred Wegener in the early years of the 20th century. Wegener had access to only a small part of the information available today, yet his theory anticipates much that is now fundamental to our conception of the earth, including not only the movement of the continents but also the wandering of the poles, with consequent changes in climate, and the significance of the distribution of ancient plants and animals. There are also parts of his work that have turned out to be wrong, but the most important points in his arguments have been substantiated. His hypothesis stands not merely as a forerunner of the concept that now prevails but as its true ancestor.

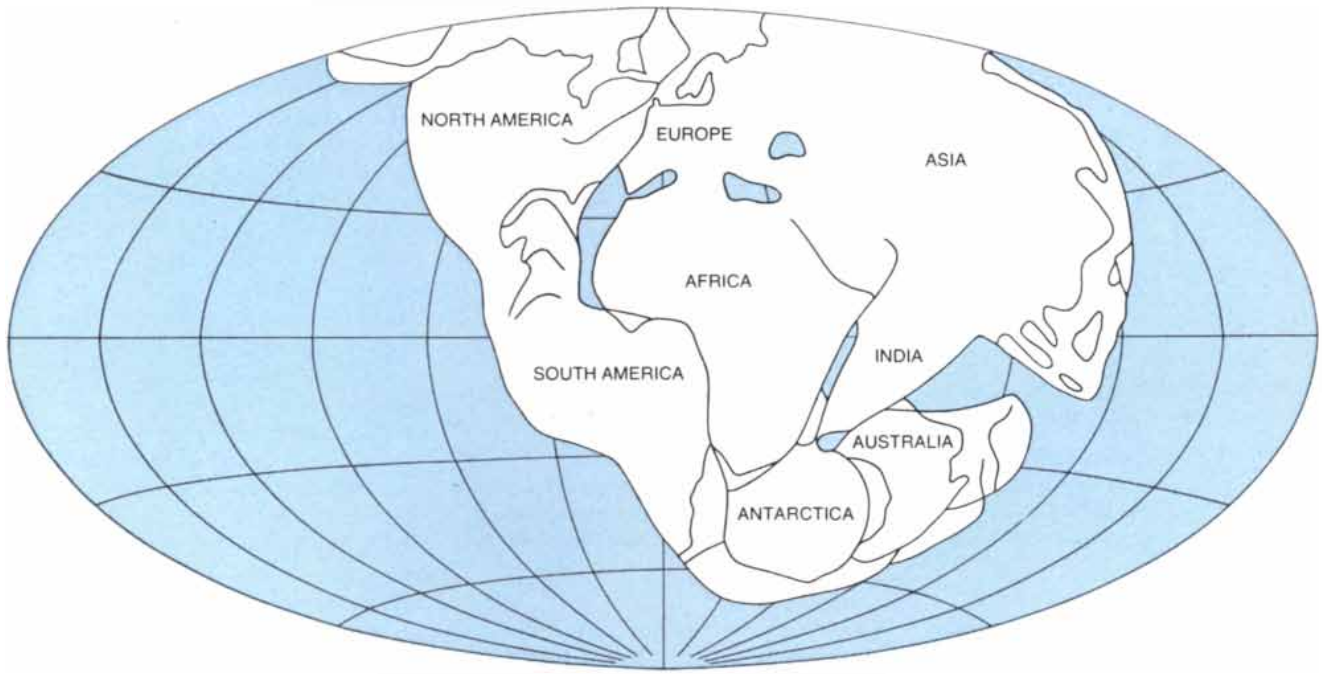
Wegener first presented his ideas to the scientific community in 1912, but it was not until 50 years later that they gained general currency. When his view of the earth did replace the older model (in the 1960's), the change represented a radical revision of a well-established doctrine, and it took place only because new evidence, derived from discoveries in geophysics and oceanography, compelled it. In the interim Wegener's theory had at best been neglected, and it had often been scorned. At the nadir proponents of continental drift were dismissed contemptuously as cranks. To understand that reaction we must examine both Wegener's work and the attitudes of his contemporaries. Did Wegener derive his conclusions from reliable evidence, and did he support them with coherent argument? Or did he merely guess, and happen to guess correctly? If his reasoning was plausible, why was his work opposed with such determination and persistence? I shall attempt to answer questions such as these, and in addition to gauge what kind of man Wegener was and to estimate his rank as a scientist.

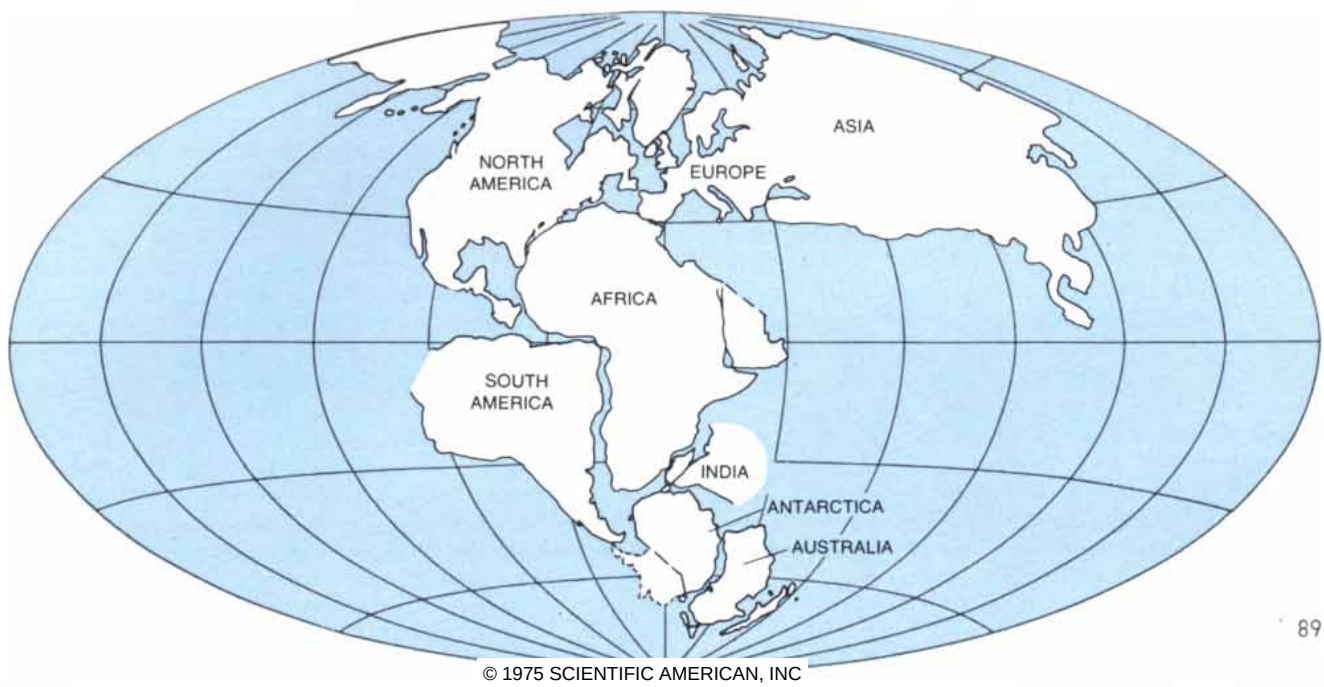
Wegener had no credentials as a geologist. Born in Berlin in 1880, the son of an evangelical minister, he studied at the universities of Heidelberg, Innsbruck and Berlin and took his doctorate in astronomy. (His most useful accomplishment in the field of his degree was a paper on the Alphonsine tables of planetary motion.) From his early days as a student he had cherished an ambition to explore in Greenland, and he had also become fascinated by the comparatively new science of meteorology. In preparation for expeditions to the Arctic he undertook a program of long walks and learned to skate and ski; in the pursuit of his other avocation he mastered the use of kites and balloons for making weather observations. He was so successful as a balloonist that in 1906, with his brother Kurt, he established a world record with an uninterrupted flight of 52 hours.

Wegener's preparations were rewarded when he was selected as meteorologist to a Danish expedition to northeastern Greenland. On his return to Germany he accepted a junior teaching position in meteorology at the University of Marburg and within a few years had written an important text on the thermodynamics of the atmosphere. A second expedition to Greenland, with J. P. Koch

ANCIENT SUPERCONTINENT incorporated all the earth's large land masses. Wegener's reconstruction of the supercontinent, which he called Pangaea, is shown in the middle illustration on the opposite page. A more recent version, shown at the bottom, differs in details of placement and orientation but preserves the major features of Wegener's proposal. Both maps are based not on the coastlines of the continents but on the edges of the continental shelves. For comparison a map of the world as it appears today is shown at top.







ALFRED WEGENER was by profession a meteorologist; he also participated in three exploratory expeditions to Greenland. In geology and geophysics he was virtually without credentials; nevertheless, it was through his work in these fields that he made his most significant contribution. This drawing of him is based on a photograph made in the 1920's.

of Denmark, followed in 1912. It included the longest crossing of the ice cap ever undertaken; the published glaciological and meteorological findings fill many volumes.

In 1913 Wegener married Else Köppen, the daughter of the meteorologist Wladimir Peter Köppen. After World War I (in which Wegener served as a junior officer) he succeeded his father-in-law as director of the Meteorological Research Department of the Marine Observatory at Hamburg. In 1924 he accepted a chair of meteorology and geophysics at the University of Graz in Austria, where he found that his colleagues were more sympathetic to his research interests than his colleagues in Hamburg had been.

Wegener died while leading a third expedition to Greenland in 1930, probably as a result of a heart attack induced by overexertion. His laudatory obituaries suggest that he had achieved considerable distinction both as a meteorologist and as an Arctic explorer; other sources suggest that in addition he had been a capable organizer and administrator and a lucid and stimulating teacher. His work on continental drift, which will surely be his permanent legacy, had remained a peripheral interest, albeit one that had absorbed him deeply.

Just how Wegener first conceived the idea that the continents could move is not certain. One unauthenticated account has it that he was inspired during a trip to Greenland while watching the calving of glacier ice (the process by which icebergs are born). From his own writings and those of his contemporaries, however, it seems more likely that he came to the theory in the same way that his predecessors had: by noting on a map the complementarity of the Atlantic coastlines. By Wegener's own account the notion first occurred to him in 1910, but a contemporary who knew Wegener as a student maintains that he had shown interest in the matter as early as 1903. Whether the idea had been maturing for a decade or for only two years, it was first presented publicly in January, 1912, in a lecture before the German Geological Association in Frankfurt am Main. The first published reports appeared later that year in two German journals.

The prevailing theory of the structure and evolution of the earth in 1912 could not accommodate drifting continents. Geologists and geophysicists then believed that the earth had been formed in a molten state and that it was still solidifying and contracting. During the process the heavy elements, such as iron, had sunk to the core and lighter ones, such as silicon and aluminum, had risen to the surface to form a rigid crust.

To most geologists of the time the model seemed quite successful in accounting for the more prominent features of the earth's surface. Mountain ranges were produced by compression of the surface during contraction, much as wrinkles develop in the skin of a drying, shrinking apple. On a larger scale the pressure generated by contraction, applied through great arches, caused some regions of the surface to collapse and subside, creating the ocean basins, while other areas remained emergent as continents. Vertical movements of the crust were considered entirely plausible, although movements parallel to the surface were excluded. Thus the continents and the ocean basins were in the long run interchangeable; some continental areas sank faster than the adjacent land and were inundated by the sea; at the same time parts of the ocean floor emerged to form dry land.

The similarity or identity of numerous fossil plants and animals on distant continents was explained by postulating land bridges that had once connected the land masses but had since sunk to the level of the ocean floor. The stratification of sedimentary deposits suggested successive marine transgressions onto the continents and regressions from them. The regressions could be attributed to the subsidence of the ocean basins and the transgressions to the partial filling of the basins with sediment eroded from the continents. At about the time Wegener was devising his hypothesis of continental drift a refinement of the traditional view was proposed in which the

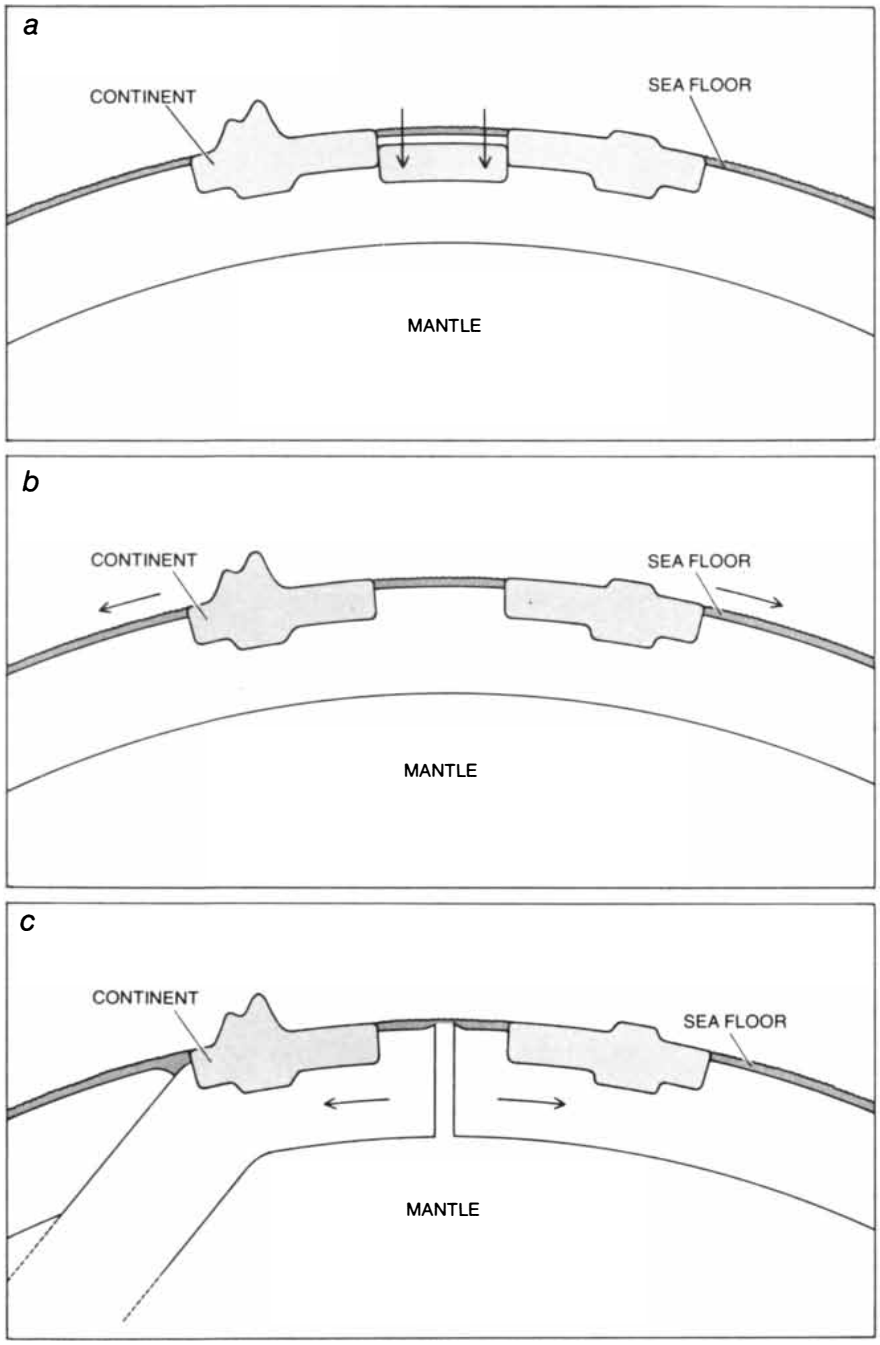
vertical movements of the crust are governed by isostasy: the concept that all elements of the system are in hydrodynamic equilibrium. Thus the continents, being less dense than the layer under them, float above the ocean floor.

Wegener detected a number of flaws and contradictions in the contracting-earth model. Moreover, many distinctive features of the earth's surface could not be explained at all by that model, unless they were to be considered the result of coincidence. The most obvious of these features is the correspondence between the Atlantic coasts of Africa and South America. (In plotting the correspondence Wegener employed not the coastline itself but the edge of the continental shelf, which is a more meaningful boundary. The same practice is followed in modern reconstructions.) Another anomaly is the distribution of mountain ranges, which are mainly confined to narrow, curvilinear belts; if they had been produced by the contraction of the globe, they should have been spread uniformly over the surface, as the wrinkles on a dried apple are.

Still another peculiarity was discovered in a statistical analysis of the earth's topography. From calculations of the total area of the earth's surface at each of many land elevations and ocean depths, Wegener found that a large fraction of the earth's crust is at two distinct levels. One corresponds to the surface of the continents, the other to the abyssal sea floor [*see illustrations on pages 94 and 95*]. Such a distribution would be expected in a crust made up of two layers, the upper one consisting of lighter rock, such as granite, and the substratum consisting of basalt, gabbro or peridotite, which would also form the ocean floor. This interpretation is supported by measurements of local variations in the earth's gravity. It is not consistent with a model of the crust in which variations in elevation are the result of random uplift and subsidence; in that case one would expect a Gaussian, or bell-shaped, distribution of elevations around a single median level.

Wegener also found support for his arguments in fossils and distinctive geological features that seemed to cross continental boundaries. In the fossil record an excellent example is provided by the reptiles. Fossils of *Mesosaurus*, a small reptile that lived late in the Paleozoic era, about 270 million years ago, are found in Brazil and in South Africa and nowhere else in the world [see "Continental Drift and the Fossil Record," by A. Hallam; SCIENTIFIC AMERICAN, November, 1972]. The peculiar distribution was traditionally explained by the sinking of a land bridge that was assumed to have connected the continents. On geophysical grounds Wegener rejected this explanation; it violated the principle of isostasy, since the material of the bridge would be less dense than that of the sea floor and could not sink into it. The only reasonable alternative was that the continents had once been joined and had since drifted apart.

The geological evidence is of a similar

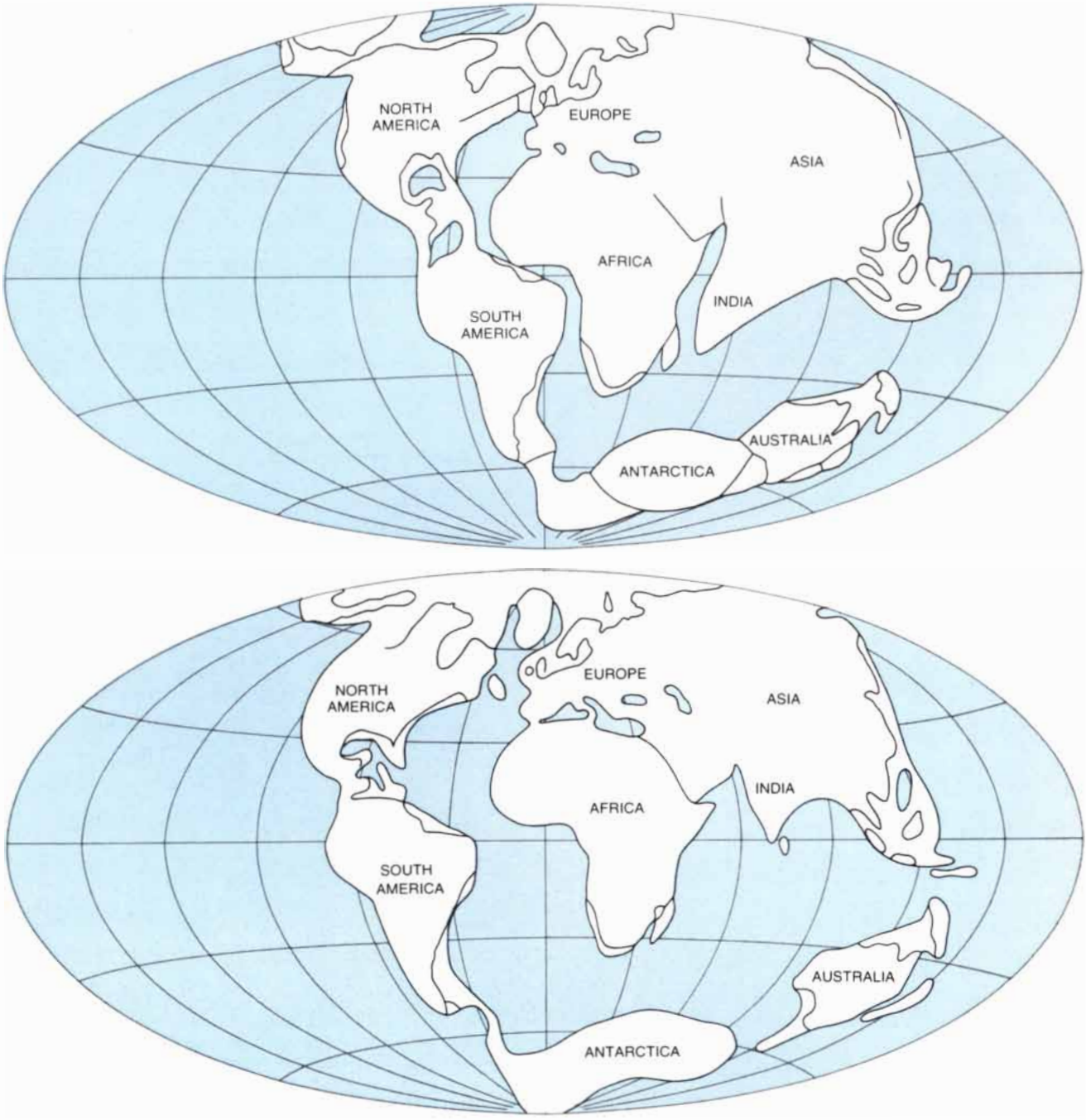


THEORY OF THE EARTH'S STRUCTURE held by most geologists at the beginning of the 20th century was challenged by Wegener. In the traditional view (*a*) the continents were fixed in place and lateral movement was impossible. Vertical motion of the crust, as in the sinking of land bridges, was allowed in some versions of the theory but objected to in others as a violation of the principle of isostasy, which holds that the continents float in hydrodynamic equilibrium on a substratum of denser material. In Wegener's hypothesis (*b*) the continents migrate through the substratum, which acts as a viscous fluid. He suggested that they were propelled by forces related to the earth's rotation, but that idea was soon discredited. In the modern theory (*c*) the continents are carried as passengers on large rigid plates. The plates are forced apart where material wells up to form new sea floor.

nature. For example, large blocks of particularly ancient rock are found both in Africa and across the Atlantic in South America; if the continents are brought together in the proper orientation, the blocks line up precisely [*see illustration on opposite page*]. Wegener himself recognized and described the power of this discovery: "It is just as if we were to refit the torn pieces of a newspaper by matching their edges, then check whether or not the lines of print run smoothly across. If they do, there is nothing left but to conclude that the pieces were in fact joined this way. If only one line was available for the test, we would still have found a high probability for the accuracy of fit, but if we have n lines, this probability is raised to the nth power."

One further line of evidence on which Wegener relied should be mentioned. Geodetic observations made early in the 20th century seemed to indicate that Greenland was moving westward, separating from Europe at a measurable rate. Such a movement might constitute a direct validation of continental drift, but it has not been confirmed in recent measurements employing more accurate techniques.

In order to resolve these contradictions Wegener formulated a comprehensive theory of the origin of the conti-



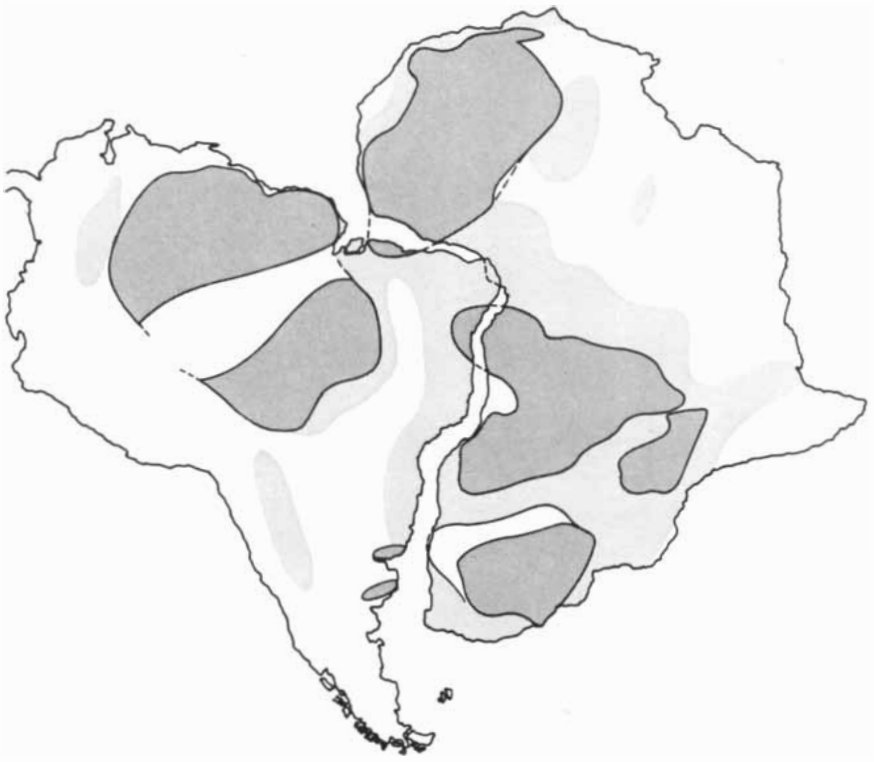
BREAKUP OF PANGAEA and the formation of the modern continents were described by Wegener in two stages. In his account all the continents remained contiguous as late as the Eocene epoch, about 50 million years ago (*upper map*). Even in the early Quaternary period, about two million years ago, South America and Antarctica were connected by an isthmus (*lower map*). Modern workers have considerably revised Wegener's sequence and his dating of events. The breakup actually began about 200 million years ago.

nents. In his reconstruction of earth history all the world's land area was originally united in a single primordial supercontinent, which he named Pangaea from the Greek meaning "all land" [*see illustration on page 89*]. He published his conclusions and evidence in 1915 in a book titled *Die Entstehung der Kontinente und Ozeane* (*The Origin of Continents and Oceans*).

The geophysical basis of Wegener's theory was closely related to the principle of isostasy. Both assume that the substratum underlying the continents acts as a highly viscous fluid; Wegener assumed further that if a land mass could move vertically through this fluid, it should also be able to move horizontally, provided only that a sufficiently powerful motive force was supplied. As evidence that such forces exist he cited the horizontal compression of folded strata in mountain ranges. There is also elegant evidence of the fluid nature of the underlying material: the earth is an oblate sphere, bulging slightly at the Equator, and the size of the bulge is exactly what would be expected for a sphere of a perfect fluid spinning at the same rate. It is a fluid of a special nature. Under short-term stresses, such as those of an earthquake, it acts as an elastic solid; only over the much longer periods of geologic time can its fluid characteristics be observed. Its behavior is analogous to that of pitch, a material that shatters under a hammer blow but flows plastically under its own weight, that is, under the milder but persistently imposed force of gravity.

After World War I Wegener merged his two principal research interests by investigating, with Köppen, changes in world climate through geologic time. By mapping the distribution of certain kinds of sedimentary rock he was able to infer the position of the poles and the Equator in ancient times. His most impressive results were obtained for the Carboniferous and Permian periods, about 300 million years ago [*see illustration on page 96*]. The position of the South Pole was determined from the disposition of boulder beds called tillites, which are formed during the movements of glaciers. In Wegener's reconstruction of Pangaea the pole was just east of what is today South Africa and within ancient Antarctica.

Ninety degrees from the pole Wegener found abundant evidence of a humid equatorial zone. The evidence consists of the vast deposits of coal that stretch from the eastern U.S. to China; fossil plants identifiable within the coals are of a tropical type. Other climatic indicators in sedimentary rock are salt and wind-deposited sand, which suggest the presence of ancient deserts. In the past, as today, deserts were formed mainly in the trade-wind belts on each side of the Equator. In the present Northern Hemisphere the Carboniferous coals are replaced in more recent strata by salt and sand-dune deposits that Wegener interpreted as signifying a southeastward shift in the position of the Equator. The movement was confirmed by a corresponding southeastward shift in the main center of tillite deposits, implying that the pole also shifted in that direction.



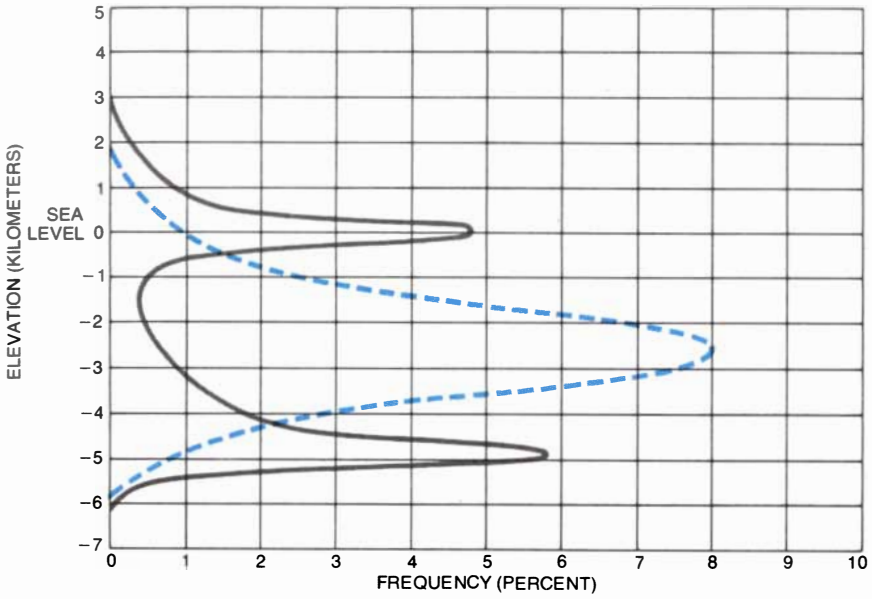
GEOLOGICAL EVIDENCE that the continents once formed a single land mass is provided by distinctive rock formations that can be assembled into continuous belts when Africa and South America are lined up next to each other. Areas shown in dark gray are ancient blocks called cratons; light gray areas represent regions of somewhat younger rock. Wegener considered continuities such as these good evidence for the continental-drift hypothesis.

The last edition of *The Origin of Continents and Oceans* devoted a chapter to ancient climate and contained an extensive discussion of the wandering of the poles. Polar movement is a phenomenon quite distinct from continental drift, but unless the continents are reassembled more or less as Wegener proposed, the distribution of tillites and coal and salt deposits cannot be interpreted coherently.

The last edition also contains more extensive documentation than the earlier ones of similarities in the geology of the southern continents, the outcome of a productive exchange of ideas between Wegener and the South African geologist Alexander L. du Toit. The essential geophysical arguments, however, remain remarkably similar to those proposed in Wegener's first paper, written almost two decades earlier.

The initial reaction of the scientific community to Wegener's hypothesis was not uniformly hostile, but at best it was mixed. At the first lecture in Frankfurt am Main some geologists were provoked to indignation; at Marburg a few days later, however, the audience seems to have been more sympathetic. Following the early publications several prominent German geologists announced their opposition to "continental displacement" (a more accurate rendering of Wegener's term, *Verschiebung*, than "drift"). A number of geophysicists, on the other hand, expressed approval of the concept. Indeed, in 1921 Wegener was able to say that he knew of no geophysicist who opposed the drift hypothesis.

The early publications, including the first edition of *The Origin of Continents and Oceans*, do not seem to have been read much outside Germany; it was not until the third edition was published in 1922 and translated into several other



STATISTICAL ANALYSIS of the elevations of the earth's surface provided Wegener with an argument in support of his hypothesis. If topographical features were formed by random uplift and subsidence, a random distribution centered on the mean elevation (*colored line*) would be expected. Actually two levels predominate (*gray line*). One, near sea level, represents average elevation of the continental platforms, the other that of the abyssal sea floor.

languages (including English) two years later that Wegener's hypothesis attracted an international audience. As in Germany, his work was initially given a fair hearing, or at the least it was not dismissed out of hand. At a meeting of the British Association for the Advancement of Science in 1922 discussion of continental drift was "lively but inconclusive," according to the published report. As one would expect, many were skeptical, but the general reception accorded the hypothesis was sympathetic. At about that time several leading geologists on each side of the Atlantic declared themselves in favor of the theory.

The stubborn antagonism to Wegener's ideas that was to be the orthodoxy of geophysics until the 1960's began to develop in the mid-1920's. Two events were instrumental in this hardening of resistance. One was the publication of a treatise titled *The Earth* by Harold Jeffreys of the University of Cambridge; the other was a symposium held in 1928 by the American Association of Petroleum Geologists.

Jeffreys attacked Wegener's theory at what was perhaps its weakest point: the nature of the forces to which Wegener attributed the movement of the continents. Wegener proposed that the westward drift of the Americas could be explained as a consequence of tides in the earth's crust; the postulated force responsible for the northward migration of India and the compression of the crust in the Alpine and Himalayan mountain belts he called *Polflucht,* or "flight from the poles." Jeffreys was able to demonstrate by simple calculations that the earth is far too strong to be even slightly deformed by such forces. If it were not, mountain ranges would collapse under their own weight and the sea floor would be perfectly flat. If the tidal force were strong enough to shift the continents westward, it would also be strong enough to halt the earth's rotation within a year.

These objections are cogent ones, and the mechanism Wegener proposed has long since been abandoned. (The modern theory attributes the movement to the spreading of the sea floor along a system of mid-ocean ridges, where molten rock from the earth's interior wells up.) We can now see, however, that Jeffreys had not refuted the theory merely by demonstrating the inadequacies of the hypothetical motive force. For the most part he simply ignored the empirical evidence, which was the most substantial part of Wegener's argument. Jeffreys dismissed the wandering of the poles as being geophysically impossible.

Of the participants in the American Association of Petroleum Geologists symposium most were hostile, in varying degrees, to Wegener's theory; only one was strongly sympathetic. The proceedings of the symposium are in the main a chorus of criticism. The supposed jigsaw-puzzle fit of the Atlantic continents was inaccurate, the contributors contended, and it did not allow for vertical movements of the crust. The similar rock formations on opposite sides of oceans were not that closely related after all; in any case present similarity did not necessarily imply former contiguity. Ancient animals could have migrated across land bridges. The Carboniferous and Permian tillites of South Africa and other areas were probably not glacial and the Northern Hemisphere coals were probably not tropical. The evidence for the movement of Greenland was inconclusive.

Some of the contributors also demanded that the hypothesis resolve paradoxes that have been approached only in the more recent versions of the theory. For example, they asked why, if the American continents could move laterally by displacing the ocean floor, they crumpled on their western edge, forming the Cordilleran mountain ranges. Did not the compressive force that formed these mountains suggest considerable resistance from the supposedly fluid sea floor? Moreover, why did Pangaea remain intact for most of the earth's history, then abruptly split apart in a few tens of millions of years?

Finally, in addition to questioning Wegener's interpretations and conclusions, some participants in the symposium assailed his credentials and his methods. He was a mere advocate, they protested, selecting for presentation only those facts that would favor his hypothesis. He "took liberties with our globe" and "played a game in which there are no restrictive rules and no sharply drawn code of conduct."

Wegener did not campaign to defend his theory from these criticisms. A middle-aged man with only limited time for research, he felt that he was unable to keep up with the swelling volume of literature and was content to leave the field to younger workers. He did, however, gently rebuke his critics for their partiality. "We are like a judge confronted by a defendant who declines to answer," he said of scientists in relation to the earth, "and we must determine the truth from circumstantial evidence. All the proof we can muster has the deceptive character of this type of evidence. How would we assess a judge who based his decision on only part of the available data?"

After Wegener's death geologists and geophysicists became even more hostile to his account of earth history. In the U.S. the reaction was particularly strong; for an American geologist to ex-

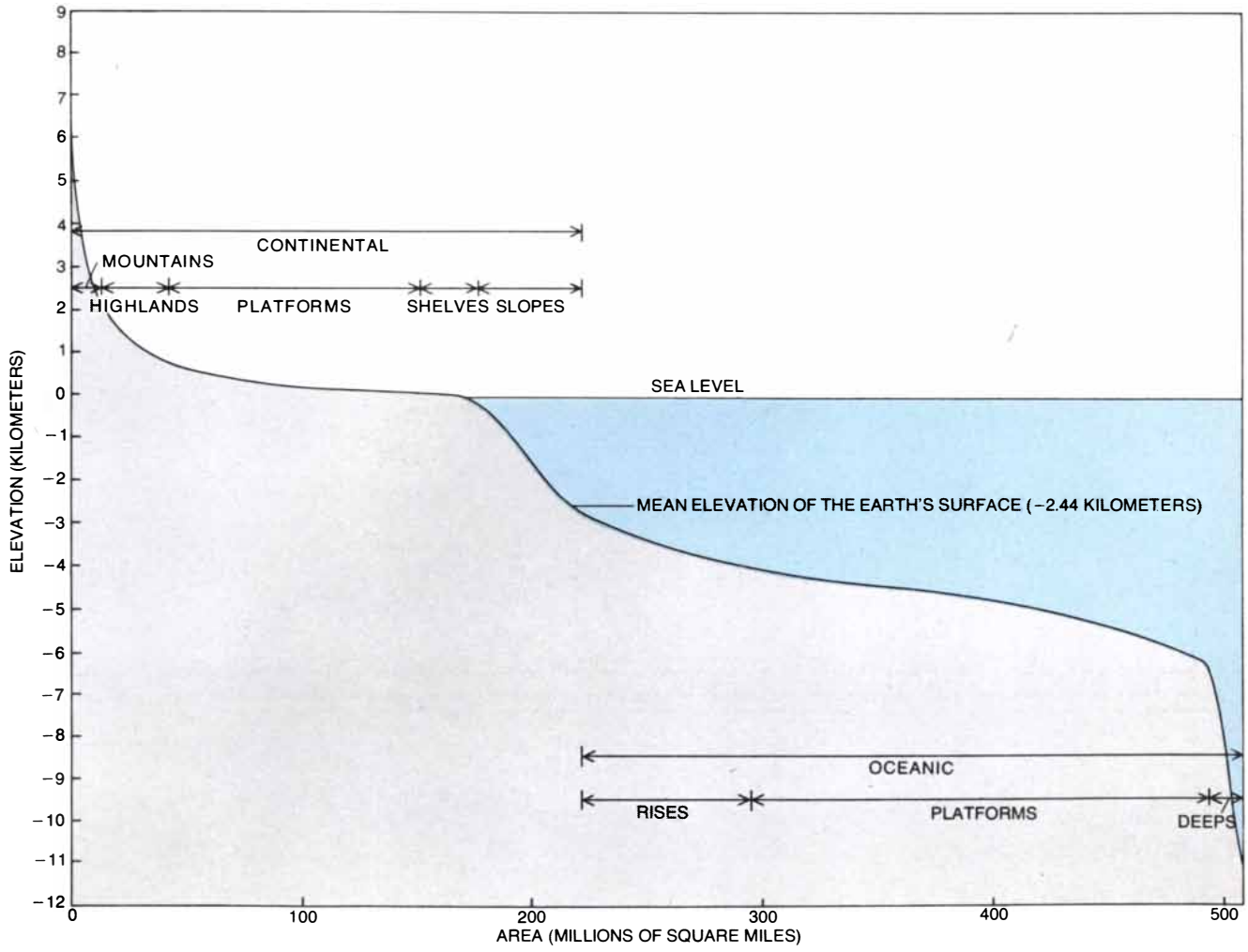
press sympathy for the idea of continental drift was for him to risk his career.

Ironically, just as the condemnatory verdict was reached, the theory was significantly strengthened by the contributions of du Toit and of Arthur Holmes of the University of Edinburgh. They eliminated some of Wegener's weaker arguments, marshaled more evidence in support of the hypothesis and provided a more plausible motive mechanism. Holmes suggested that the continents are moved by convection currents in the earth's mantle. His hypothesis was not entirely satisfactory, but it successfully circumvented the criticisms of Jeffreys and his followers. Moreover, it anticipated the modern explanation of why continents move.

In hindsight, and with the knowledge that many aspects of Wegener's theory have been confirmed in the past two decades, we can readily appreciate his accomplishments. He examined critically a model of the earth that was then almost universally accepted. When he discovered weaknesses and inconsistencies, he was bold enough and independent enough to embrace a radical alternative. Furthermore, he had sufficient breadth of knowledge to seek out and perceptively evaluate supporting evidence from a variety of disciplines. The same qualities of mind were applied to the explication of ancient climate and the wandering of the poles.

The intellectual rigor that Wegener brought to his work is illustrated in his own writings and attested to in the words of those who knew him well. In a letter to Köppen written in 1911 Wegener defended his views on continental drift. The letter is reproduced in the biography of Wegener written by his widow. The following passage, which I have translated, has not to my knowledge been published previously in English.

"You consider my primordial continent to be a figment of my imagination, but it is only a question of the interpretation of observations. I came to the idea on the grounds of the matching coastlines, but the proof must come from geological observations. These compel us to infer, for example, a land connection between South America and Africa. This can be explained in two ways: the sinking of a connecting continent or separation. Previously, because of the unproved concept of permanence, people have considered only the former and have ignored the latter possibility. But the modern teaching of isostasy and more generally our current geophysical ideas oppose the sinking of a continent because it is lighter than the material on which it rests. Thus we are forced to consider the alternative interpretation. And if we now find many surprising simplifications and can begin at last to make real sense of an entire mass of geological data, why should we delay in throwing

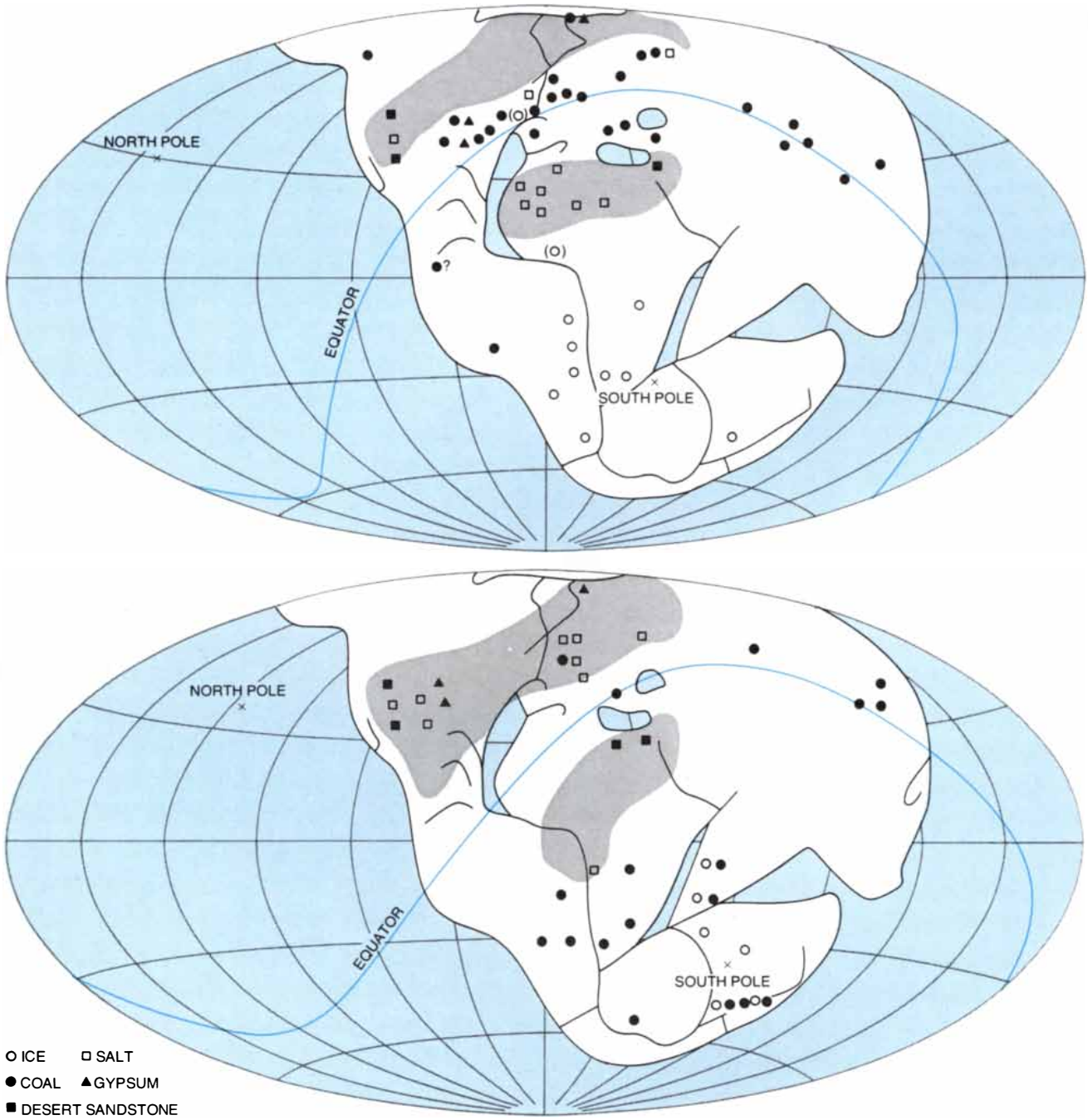


DISTRIBUTION OF ELEVATIONS, when correlated with the area of the earth's surface, confirms that the crust has two fundamental levels. Approximately 300 million square miles of the 510 million square miles of the earth's surface is continental platform or oceanic platform. The distribution conforms to Wegener's model of the earth, in which the continents float in a dense substratum.

the old concept overboard? Is this revolutionary? I don't believe that the old ideas have more than a decade to live. At present the notion of isostasy is not yet thoroughly worked out; when it is, the contradictions involved in the old ideas will be fully exposed." It evidently seemed quite obvious to Wegener, but it is clear that he underestimated the ardor of those committed to the "old ideas."

Wegener's success in constructing from scattered and seemingly unrelated observations a systematic theory of earth history could be attributed to his broad attack on the problem, and perhaps even to his status as a nonspecialist. A glimpse of his approach to scientific problems is provided by an obituary written by Hans Benndorf, a professor of physics and a colleague of Wegener's at Graz. This passage too was translated by me and has not appeared before in English.

"Wegener acquired his knowledge mainly by intuitive means, never or only quite rarely by deduction from a formula, and when that was the case, it needed



WANDERING OF THE POLES was proposed by Wegener to account for the distribution of ancient climates. Evidence of tropical climate was provided by certain kinds of coal, of polar climate by the tillites that signal glaciation, and of arid climate by deposits of salt, gypsum and desert sandstone. The symbols employed in the maps are identified in the key at left; in addition arid zones, which are characteristic of trade-wind latitudes, are indicated by gray areas. The upper map shows Wegener's reconstruction for the Carboniferous period, about 300 million years ago; the lower map is for the Permian period, about 230 million years ago. (The distortion of the Equator is caused by the projection employed.) The movement of the continents and the movement of the poles are unrelated, but a coherent account of ancient climate can be devised only by rearranging the continents approximately as Wegener did.

only to be quite simple. Also, if matters concerning physics were involved, that is, in a field distant from his own field of expertise, I was often astonished by the soundness of his judgment. With what ease he found his way through the most complicated work of the theoreticians, with what feeling for the important point! He would often, after a long pause for reflection, say, 'I believe such and such,' and most times he was right, as we would establish several days later after rigorous analysis. Wegener possessed a sense for the significant that seldom erred."

Benndorf's assessment of Wegener's method is supported by the remarks of Wilhelm Max Wundt, who knew Wegener as a student in Berlin: "Alfred Wegener started out to tackle his scientific problems with only quite ordinary gifts in mathematics, physics and the other natural sciences. He was never, throughout his life, in any way reluctant to admit that fact. He had, however, the ability to apply those gifts with great purpose and conscious aim. He had an extraordinary talent for observation and for knowing what is at the same time simple and important, and what can be expected to give a result. Added to this was a rigorous logic, which enabled him to assemble rightly everything relevant to his ideas."

If Wegener was in fact the talented and perceptive scientist his contemporaries describe, and if his conclusions were well rooted in evidence and argument, an obvious question arises: Why was the opposition to his ideas so strong, widespread and persistent?

One possible explanation is that Wegener's theory was "premature" at the time he presented it. Gunther S. Stent of the University of California at Berkeley has argued that an idea must be considered premature if it cannot be connected by a series of simple, logical steps to the canonical, or generally accepted, knowledge of the time [see "Prematurity and Uniqueness in Scientific Discovery," by Gunther S. Stent; SCIENTIFIC AMERICAN, December, 1972]. A related principle, formulated by Michael Polanyi, a British writer on the philosophy and sociology of science, holds that in science there must always be a prevailing opinion of the nature of things, against which the truth of all assertions is tested. Any observation that seems to contradict the established view of the world must be presumed to be invalid and set aside in the hope that it will eventually turn out to be false or irrelevant. This interpretation of how science works suggests that geologists and geophysicists had to be overwhelmed by evidence, as they were in the 1960's, before they could abandon the established doctrine of stationary continents. Wegener's innovations, precisely because they were innovative, had to be held in abeyance until a new orthodoxy could be created from them.

There is almost certainly truth in this analysis, and reassurance that the rejection of Wegener's work was a necessary circumstance in the orderly progress of science; nevertheless, the analysis is not entirely convincing. It could account for indifference to the hypothesis of continental drift but not for the attitude of the many scientists who relegated the theory to the realm of fantasy. It also fails to explain why the traditional model of the earth was retained even after Wegener demonstrated that there were contradictions in it and even though those contradictions were never resolved. Paleontologists continued to rely on vanishing land bridges, for example, at the same time that geophysicists, who had adopted the principle of isostasy, insisted that the sinking of such bridges was impossible.

It has been suggested that the principal impediment to the acceptance of continental drift was the lack of a plausible motive force after Jeffreys had refuted Wegener's initial proposals. If that is the case, however, why was Holmes's convection-current theory given so little consideration? Furthermore, even today the nature of the motor that moves the continents remains uncertain, yet plate tectonics is so well established that those who reject its basic tenets are commonly dismissed as reactionaries.

Perhaps the long travail of Wegener's hypothesis can be best explained as a consequence of inertia. A geologist at the 1928 symposium of the American Association of Petroleum Geologists is reported to have said: "If we are to believe Wegener's hypothesis, we must forget everything that has been learned in the past 70 years and start all over again." It should also be remembered that to the geologists of the time Wegener was an outsider; they must have regarded him as an amateur. Today, of course, we can see that his position was an advantage because he had no stake in preserving the conventional viewpoint. Moreover, we can see that he was not an amateur after all but an interdisciplinary investigator of talent and vision who surely qualifies for a niche in the pantheon of great scientists.

MATHEMATICAL GAMES

How the absence of anything leads to thoughts of nothing

by Martin Gardner

"Nobody seems to know how to deal with it. (He would, of course.)"
—P. L. HEATH
in *The Encyclopedia of Philosophy*

Our topic is nothing. By definition nothing does not exist, but the concepts we have of it certainly exist as concepts. In mathematics, science, philosophy and everyday life it turns out to be enormously useful to have words and symbols for such concepts.

The closest a mathematician can get to nothing is by way of the null (or empty) set. It is not the same thing as nothing because it has whatever kind of existence a set has, although it is unlike all other sets. It is the only set that has no members and the only set that is a subset of every other set. From a basket of three apples you can take one apple, two apples, three apples or no apples. To an empty basket you can, if you like, add nothing. The null set denotes, even though it does not denote anything. For example, it denotes such things as the set of all square circles, the set of all even primes other than 2 and the set of all readers of this department that are chimpanzees. In general it denotes the set of all x's that satisfy any statement about x that is false for all values of x. Anything you say about a member of the null set is true, because it lacks a single member for which a statement can be false.

The null set is symbolized by Ø. It must not be confused with 0, the symbol for zero. Zero is (usually) a number that denotes the number of members of Ø. The null set denotes nothing, but 0 denotes the number of members of such sets, for example the set of apples in an empty basket. The set of these nonexisting apples is Ø, but the number of apples is 0.

A way to construct the counting numbers, discovered by Gottlob Frege and rediscovered by Bertrand Russell, is to start with the null set and apply a few simple rules and axioms. Zero is defined as the cardinal number of elements in all sets that are equivalent to (can be put in correspondence with) the members of the null set. After creating 0, 1 is defined as the number of members in all sets equivalent to 0, 2 as the number of members in all sets equivalent to the set of 0 and 1, 3 as the number of members in all sets equivalent to the set of 0, 1, 2, and so on.

A few years ago John Horton Conway of the University of Cambridge hit on a remarkable new way to construct numbers that also starts with the null set. He first described his technique in a photocopied typescript of 13 pages, "All Numbers, Great and Small." It begins: "We wish to construct all numbers. Let us see how those who were good at constructing numbers have approached the problem in the past." It ends with 10 open questions, of which the last is: "Is the whole structure of any use?"

Conway explained his new system to Donald E. Knuth, a computer scientist at Stanford University, when they happened to meet at lunch one day in 1972. Knuth was immediately fascinated by its possibilities and its revolutionary content. In 1973 during a week of relaxation in Oslo, Knuth wrote an introduction to Conway's method in the form of a novelette. It was issued in paperback last year by Addison-Wesley, which also publishes Knuth's well-known books titled *The Art of Computer Programming*. I believe it is the only time a major mathematical discovery has been published first in a work of fiction. Conway is reportedly working on his own book, *On Numbers and Games*, which will present a fuller account of his method.

Knuth's novelette, *Surreal Numbers*, is subtitled "How Two Ex-Students Turned on to Pure Mathematics and Found Total Happiness." The book's primary aim, Knuth explains in a postscript, is not so much to teach Conway's theory as "to teach how one might go about developing such a theory." He continues: "Therefore, as the two characters in this book gradually explore and build up Conway's number system, I have recorded their false starts and frustrations as well as their good ideas. I wanted to give a reasonably faithful portrayal of the important principles, techniques, joys, passions, and philosophy of mathematics, so I wrote the story as I was actually doing the research myself."

Knuth's two ex-mathematics-students, Alice and Bill (*A* and *B*), have fled from the "system" to a haven on the coast of the Indian Ocean. There they unearth a half-buried black rock carved with ancient Hebrew writing. Bill, who knows Hebrew, manages to translate the opening sentence: "In the beginning everything was void, and J. H. W. H. Conway began to create numbers." JHWH is a transliteration of how the ancient Hebrews wrote the name Jehovah. "Conway" also appears without vowels, but it was the most common English name Bill could think of that fitted the consonants.

Translation of the "Conway stone" continues: "Conway said, 'Let there be two rules which bring forth all numbers large and small. This shall be the first rule: Every number corresponds to two sets of previously created numbers, such that no member of the left set is greater than or equal to any member of the right set. And the second rule shall be this: One number is less than or equal to another number if and only if no member of the first number's left set is greater than or equal to the second number, and no member of the second number's right set is less than or equal to the first number.' And Conway examined these two rules he had made, and behold! they were very good."

The stone's text goes on to explain how on the zero day Conway created zero. He did it by placing the null set on the left and also on the right. In symbolic notation $0 = \emptyset \mid \emptyset$, where the vertical line divides the left and right sets. No member of the left Ø is equal to or greater than a member of the right Ø because Ø has no members, so that Conway's first rule is satisfied. Applying the second rule, it is easy to show that 0 is less than or equal to 0.

On the next day, the stone reveals, Conway created the first two nonzero integers, 1 and −1. The method is simply to combine the null set with 0 in the two possible ways: $1 = 0 \mid \emptyset$ and $-1 = \emptyset \mid 0$. It checks out. Minus 1 is less than but not equal to 0, and 0 is less than but not equal to 1. Now, of course, 1 and −1 and all subsequently created numbers can be plugged back into the left-right formula, and in this way all the integers

are constructed. With 0 and 1 forming the left set and Ø on the right, 2 is created. With 0, 1 and 2 on the left and Ø on the right, 3 is created, and so on.

At this point readers might enjoy exploring a bit on their own. Jill C. Knuth's illustration for the front cover of *Surreal Numbers* shows some huge boulders shaped to symbolize $0 \mid 1$. What number does this define? And can the reader prove that $-1 \mid 1 = 0$?

"Be fruitful and multiply," Conway tells the integers. By combining them, first into finite sets, then into infinite sets, the "copulation" of left-right sets continues, aided by no more than Conway's ridiculously simple rules. Out pour all the rest of the real numbers: first the integral fractions, then the irrationals. At the end of aleph-null days a big bang occurs and the universe springs into being. That, however, is not all. Before Conway has finished, his method has produced all of Georg Cantor's transfinite numbers, all infinitesimal numbers (they are reciprocals of infinite numbers) and infinite sets of queer new quantities such as the roots of transfinites and infinitesimals!

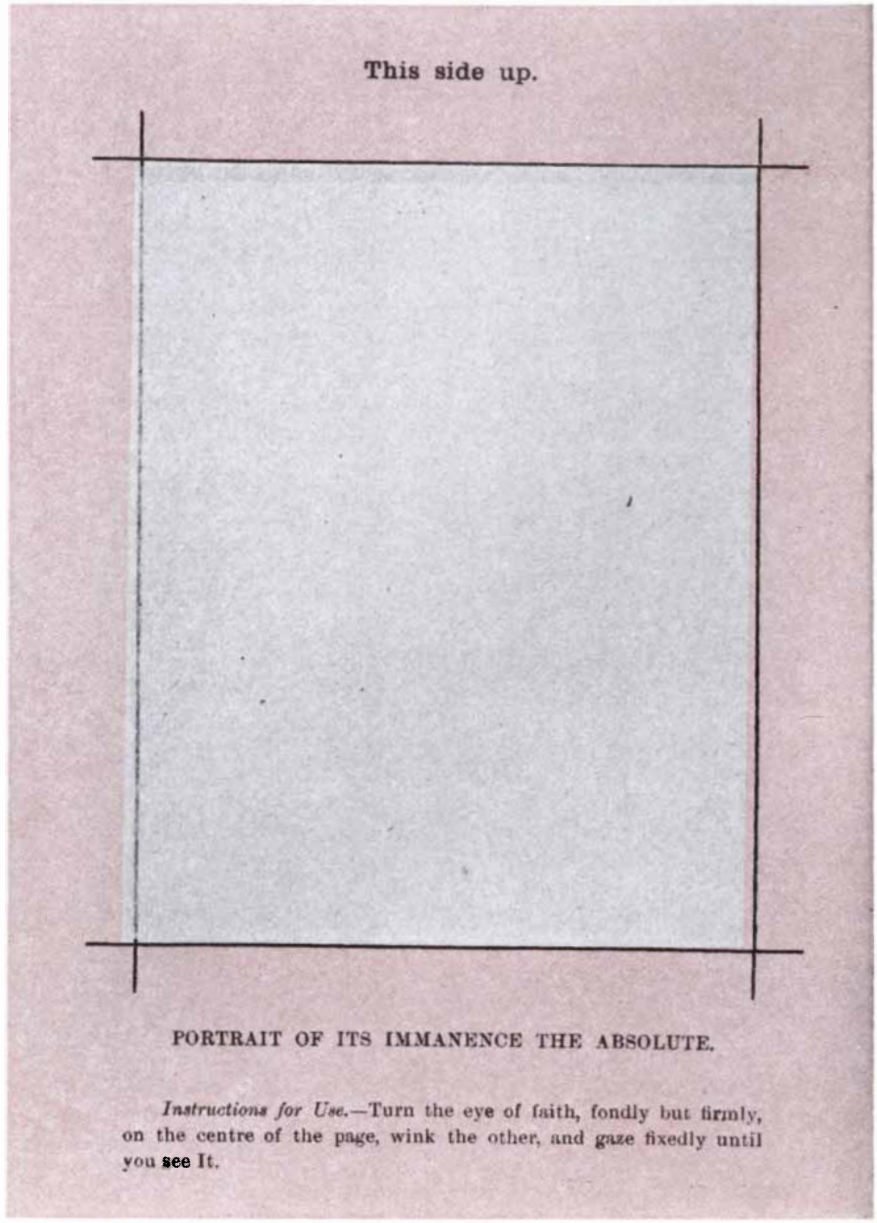
It is an astonishing feat of legerdemain. An empty hat rests on a table made of a few axioms of standard set theory. Conway waves two simple rules in the air, then reaches into almost nothing and pulls out an infinitely rich tapestry of numbers that form a real and closed field. Every real number is surrounded by a host of new numbers that lie closer to it than any other "real" value does. The system is truly "surreal."

"Man, that empty set sure gets around!" exclaims Bill. "I think I'll write a book called *Properties of the Empty Set.*" This notion that nothing has properties is, of course, commonplace in philosophy, science and ordinary language. Lewis Carroll's Alice may think it nonsense when the March Hare offers her nonexistent wine, or when the White King admires her ability to see nobody on the road and wonders why nobody did not arrive ahead of the March Hare because nobody goes faster than the hare. It is easy, however, to think of instances in which nothing actually does enter human experience in a positive way.

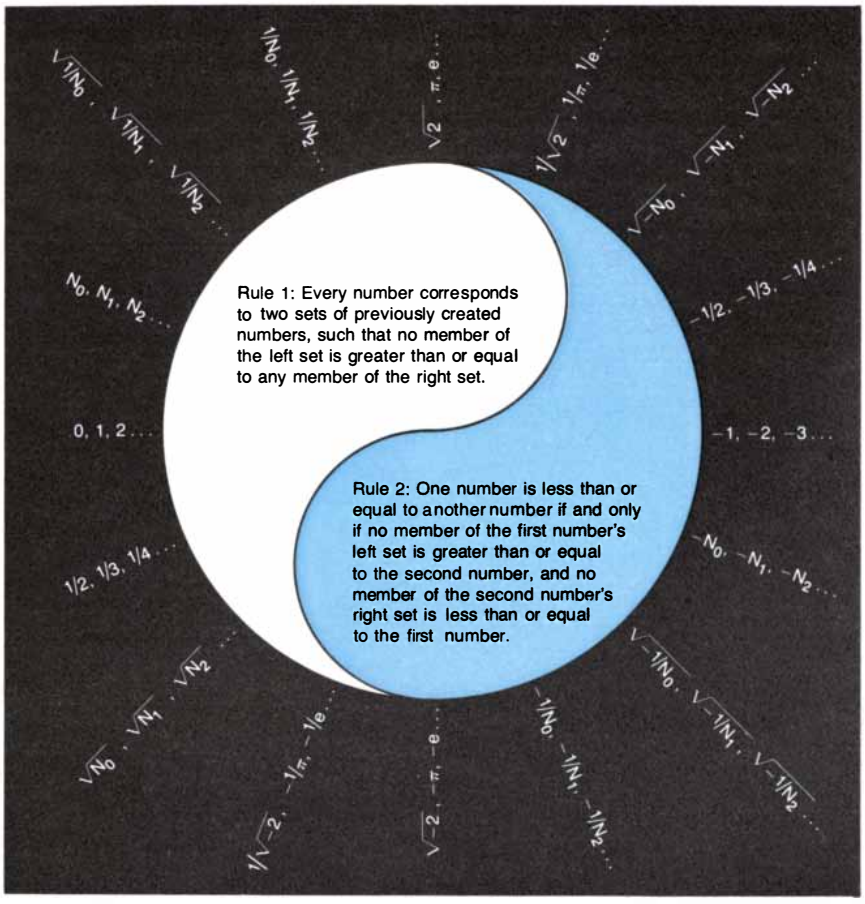
Consider holes. An old riddle asks how much dirt is in a rectangular hole of certain dimensions. Although the hole has all the properties of a rectangular parallelepiped (corners, edges, faces with areas, volume and so on), the answer is that there is no dirt in the hole. The various holes of our body are certainly essential to our health, sensory awareness and pleasure. In *Dorothy and the Wizard in Oz,* the braided man, who lives on Pyramid Mountain in the earth's interior, tells Dorothy how he got there. He had been a manufacturer of holes for Swiss cheese, doughnuts, buttons, porous plasters and other things. One day he decided to store a vast quantity of adjustable postholes by placing them end to end in the ground, making a deep vertical shaft into which he accidentally tumbled.

The mathematical theory behind Sam Loyd's sliding-block puzzle (15 unit cubes inside a 4-by-4 box) is best explained by regarding the hole as a moving cube. It is analogous to what happens when a gold atom diffuses through lead. Bubbles of nothing in liquids, from the size of a molecule on up, can move around, rotate, collide and rebound just like things. Negative currents are the result of free electrons jostling one another along a conductor, but holes caused by an absence of free electrons can do the same thing, producing a positive "hole current" that goes the other way.

Osborne Reynolds, a British engineer who died in 1912, invented an elaborate theory in which matter consists of microparticles of nothing moving through the ether the way bubbles move through liquids. His two books about the theory, *On an Inversion of Ideas as to the Structure of the Universe* and *The Sub-Mechanics of the Universe,* both published by the Cambridge University Press, were taken so seriously that W. W. Rouse Ball, writing in early editions of his *Mathematical Recreations and Essays,*



Frontispiece of* Mind, *Special Christmas Number, 1901



John Horton Conway's rules for generating numbers from nothing

called the theory "more plausible than the electron hypothesis."

Reynolds' inverted idea is less crazy than it sounds. P. A. M. Dirac, in his famous theory that predicted the existence of antiparticles, viewed the positron (the antielectron) as a hole in a negative-charge continuum. When an electron and a positron collide, the electron falls into the positron hole, causing both particles to vanish.

It is true that the old concept of a "stagnant ether" has been abandoned by physicists, but in its place is not nothing. The "new ether" consists of the metric fields responsible for the basic forces of nature, perhaps also for all the particles. John Archibald Wheeler proposes a substratum, called superspace, of infinitely many dimensions. It is empty of everything except mathematical structure. Occasionally a portion of it twists in such a peculiar way that it explodes, creating a universe of finite dimensions, with its own set of laws and within which the field gets tied into little knots that we call "matter." There still is a difference between something and nothing, but it is purely geometrical and there is nothing behind the geometry.

Empty space is like a straight line of zero curvature. Bend the line and you have a universe. Add little bumps that ripple back and forth and you create matter and energy. Outside the utmost reaches of our now expanding cosmos are vast regions unpenetrated by light and gravity, beyond which there may be other universes. Shall we say that these regions contain absolutely nothing, or are they still saturated with a metric of zero curvature? Ancient Greek and medieval arguments about being and nonbeing, one world or many and whether a perfect vacuum can "exist" seem quaint on first reading. Change the terminology a bit and they turn out to be equivalent to present controversies.

There are many examples from the arts—some jokes, some not—of nothing admired as something. A nonobjective painting consists of an all-black canvas. *The Nothing Book,* on sale this past Christmas in regular and deluxe bindings, has nothing but blank pages. Chapters 18 and 19 of the final volume of *Tristram Shandy* are without content. John Cage's piano composition, "4′33″," calls for four minutes 33 seconds of total silence as the player sits frozen on the piano stool. I have not heard it performed, but friends tell me it is Cage's finest composition. I know of no piece of "minimal sculpture" that is reduced to the absolute minimum of nothing, but Henry Moore has certainly exploited the aesthetics of holes. Nor have I heard of a play or motion picture that consists from beginning to end of an empty stage or screen, although some of the eccentric films produced by Andy Warhol come close to it.

Events can occur in which nothing is as startling as a thunderclap. An old joke tells of a man who slept in a lighthouse under a foghorn that boomed regularly every 10 minutes. One night at 3:20 A.M., when the mechanism failed, the man leaped out of bed shouting, "What was that?" As a prank all the members of a large orchestra once stopped playing suddenly in the middle of a strident symphony, causing the conductor to fall off the podium. One afternoon in a rural section of North Dakota where the wind blew constantly there was a sudden cessation of wind. It is said that all the chickens fell over.

There are many examples that are not jokes. An absence of water can cause death. The loss of a loved one, of money or of a reputation can push someone to suicide. The law recognizes innumerable occasions on which a failure to act is a crime. Grave consequences will follow when a man on a railroad track, in front of an approaching train and unable to decide whether to jump to the left or to the right, makes no decision.

These are all examples of little pockets in which there is an absence of something. What about that monstrous dichotomy between all being—everything there is—and nothing? From the earliest times the most eminent thinkers have meditated on this ultimate split. It seems unlikely that the universe is going to vanish (although I myself once wrote a story, *Oom,* about how God, weary of existing, abolished everything, including himself), but the fact that we ourselves will soon vanish is real enough. In medieval times the fear of death was mixed with a fear of eternal suffering, but since the fading of hell (albeit it is now enjoying a renaissance) this fear has been replaced by what Sören Kierkegaard called an "anguish" or "dread" over the possibility of becoming nothing.

This brings us abruptly to what Paul Edwards has called the "superultimate question." "Why," asked Leibniz, Schelling, Schopenhauer and a hundred other philosophers, "should something exist rather than nothing?"

Obviously it is a curious question, not like any other. Large numbers of people, perhaps the majority, live out their

lives without ever considering it. If someone asks them the question, they may fail to understand it and believe the questioner is crazy. Among those who understand the question, there are varied responses. Thinkers of a mystical turn of mind, Martin Heidegger for instance, consider it the deepest, most fundamental of all metaphysical questions, and they look with contempt on all philosophers who are not equally disturbed by it. Those of a positivistic, pragmatic turn of mind consider it trivial. Since everyone agrees there is no way to answer it empirically or rationally, it is a question without cognitive content, as meaningless as asking if the number 2 is red or green. Indeed, a famous paper by Rudolf Carnap on the meaning of questions heaps scorn on a passage in which Heidegger pontificates about being and nothingness.

A third group of philosophers, including Milton K. Munitz, who wrote an entire book titled *The Mystery of Existence,* regards the question as being meaningful but insists that its significance lies solely in our inability to answer it. It may or may not have an answer, argues Munitz, but in any case the answer lies totally outside the limits of science and philosophy.

Whatever their metaphysics, those who have puzzled most over the superultimate question have left much eloquent testimony about those unexpected moments, fortunately short-lived, in which one is suddenly caught up in an overwhelming awareness of the utter mystery of why anything is. That is the terrifying emotion at the heart of Jean-Paul Sartre's great philosophic novel *Nausea.* Its red-haired protagonist, Antoine Roquentin, is haunted by the superultimate mystery. "A circle is not absurd," he reflects. "It is clearly explained by the rotation of a straight segment around one of its extremities. But neither does a circle exist." Things that do exist, such as stones and trees and himself, exist without any reason. They are just insanely there, bloated, obscene, gelatinous, unable not to exist. When the mood is on him, Roquentin calls it the nausea. William James had earlier called it an "ontological wonder sickness." The monotonous days come and go, all cities look alike, nothing happens that means anything.

G. K. Chesterton is as good an example as any of the theist who, stunned by the absurdity of being, reacts in opposite fashion. Not that shifting to God the responsibility for the world's existence answers the superultimate question; far from it! One immediately wonders why God exists rather than nothing. But although none of the awe is lessened by hanging the universe on a transcendent peg, the shift can give rise to feelings of gratitude and hope that relieve the anxiety. Chesterton's existential novel *Manalive* is a splendid complement to Sartre's *Nausea.* Its protagonist, Innocent Smith, is so exhilarated by the privilege of existing that he goes about inventing whimsical ways of shocking himself into realizing that both he and the world are not nothing.

Let P. L. Heath, who had the first word in this article, also have the last. "If nothing whatsoever existed," he writes at the end of his article on nothing in *The Encyclopedia of Philosophy,* "there would be no problem and no answer, and the anxieties even of existential philosophers would be permanently laid to rest. Since they are not, there is evidently *nothing to worry about.* But that itself should be enough to keep an existentialist happy. Unless the solution be, as some have suspected, that it is not nothing that has been worrying them, but they who have been worrying it."

The Sam Loyd puzzle picture that was reproduced last month showed a square Washington pie divided into 16 smaller squares. The problem was to slice it into just six squares, not necessarily the same size. The 16 small cells are pure misdirection. The only way to

"Nuclear Energy," sculpture by Henry Moore

do it is to rule the pie into nine squares, like a ticktacktoe board. The pie is now divided easily into one 2-by-2 square and five unit squares.

Readers too numerous to mention pointed out that Barry Wolk's rule ["Mathematical Games," October, 1974] for picking the best triplet *B* to beat triplet *A* is equivalent to putting in front of *A* the complement of its next-to-last symbol, then dropping the last symbol. More than half of the correspondents found that the same method also works for quadruplets except for the two in which *H* and *T* alternate throughout. In those two cases the symbol put in front is the same as the next to last.

Several readers searched for a general algorithm that would pick the best *n*-tuplet for all values of *n*. The only conjecture that did not fail for 5-tuplets was by Michael J. Connolly of Madison Township, N.J. His procedure is:

1. Form two *n*-tuplets by prefixing either *H* or *T* to *A*'s tuplet and drop its final symbol.

2. If one of the two new tuplets is all heads or all tails, eliminate it and choose the other.

3. If neither tuplet has all symbols alike, place each directly below the original tuplet. Slide the lower tuplet to the right until each of its symbols matches the symbol directly above. Choose the tuplet with minimum or zero overlap.

For example, we want to know the best 7-tuplet for beating *HHHTHHH*. The two new tuplets are *HHHHTHH* and *THHHTHH*. Since neither has all symbols alike, we compare its head-tail overlap with the original.

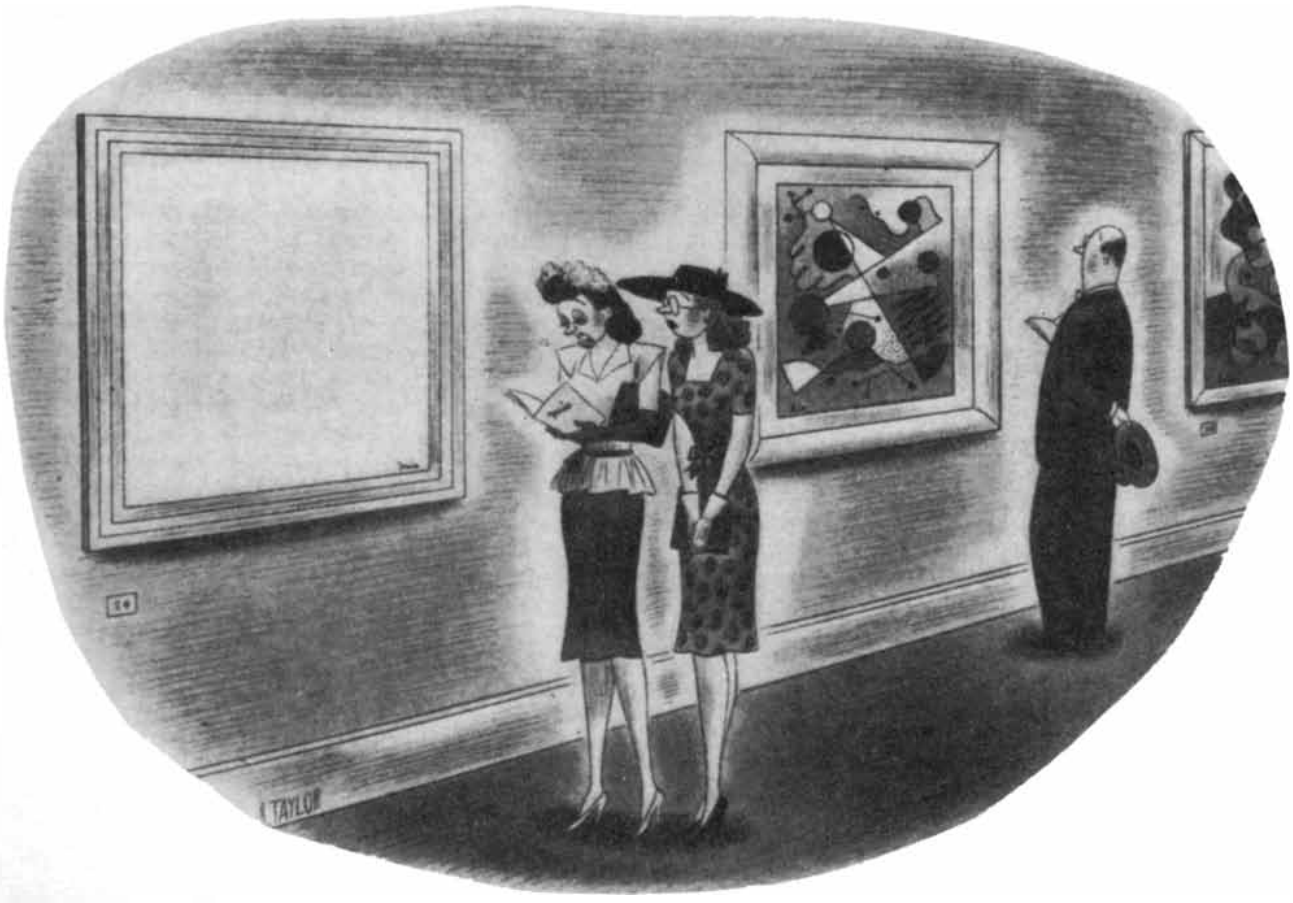
HHHTHHH *HHHTHHH*
HHHHTHH *THHHTHH*

The overlap of three on the left is smaller than the overlap of four on the right, so that the best 7-tuplet is *HHHHTHH*.

Connolly wrote a computer program that confirmed his procedure for all *n*-tuplets through $n = 7$ and for selected *n*-tuplets of the next three higher orders. I would appreciate receiving any validation of the algorithm or a counterexample.

Proofs of John Horton Conway's algorithm for computing the winning odds for two different *n*-tuplets were supplied by David Sachs and Bryce Hurst, but both are beyond my competence to evaluate. Sachs and Dennis Clayton each noted that Conway's "leading number," when an *n*-tuplet is compared with itself, automatically provides that tuplet's waiting time. Simply double the leading number. William W. Hsieh described his method of calculating winning odds by using directed graphs, which produces the same results as Conway's method.

It was my error, not Conway's, to state that his algorithm applied to all pairs of tuplets of different lengths. That is true only when the smaller tuplet is not entirely contained within the larger. If the smaller tuplet is within, not at the end of, the larger (for example *THTH* and *HT*), it obviously wins with a probability of 1. If it is at the end of the larger (for example *HHT* and *T*), it obviously wins or ties with a probability of 1.



"It says, 'During the Barcelona period he became enamored of the possibilities inherent in virgin space. With a courage born of the most profound respect for the enigma of the imponderable, he produced, at this time, a series of canvases in which there exists solely an expanse of pregnant white.'"

Cartoon from The New Yorker, *September 23, 1944. © 1944, 1972 The New Yorker Magazine, Inc.*

From the Fundamentals to the Frontiers: Freeman Books in Biology

Introduction to Biophysical Plant Physiology

PARK S. NOBEL, University of California, Los Angeles

In this expansion of his earlier *Plant Cell Physiology*, the author relates biophysical models of cell interaction to the whole plant, discusses in some detail the physicochemical areas of thermodynamics and photochemistry, and includes an extensive chapter on environmental physiology. The mathematical arguments may be followed by readers who have had a brief exposure to elementary calculus. *An Introduction to Biophysical Plant Physiology* can serve as a text for advanced undergraduate and graduate courses in plant physiology as well as a reference for the professional.

"I plan to use this book along with three others to teach an advanced undergraduate-graduate course in plant physiology. . . . It is a required text as it brings plant physiology from a vague qualitative approach to a quantitative approach!"
—Govindjee, University of Illinois

1974, 488 pp., 68 illus., $13.50

The Strategy of Life

CLIFFORD GROBSTEIN, University of California, San Diego

The *Quarterly Review of Biology* said of the first edition: "The book is brief, and says much in little space. It is beautifully conceived and written. As an outstanding statement of what modern biology thinks about life, it is mandatory reading for all would-be and for practicing biologists, as well as for non-biologists who like to read wonderful books."

The second edition—revised to include greater emphasis on man—maintains these high standards. "It is," writes the author, "a statement about life—where it came from, what its nature is, and what premonition we have of its future. . . . Anyone contemplating the issues of our time needs to understand the strategy that life has followed to reach Man and his present state. Without that understanding, Man appears more fragile than he will prove to be in the terrible test of the coming century."

Second Edition, 1974, 174 pp., 90 illus., cloth $4.95, paper $2.95

Plant Science

An Introduction to World Crops

JULES JANICK, Purdue University,
ROBERT W. SCHERY, The Lawn Institute,
FRANK W. WOODS, The University of Tennessee,
and VERNON W. RUTTAN, Agricultural Development Council

"I am using the book as a required text for a course known as Plant Science 200—Ecology of Crop Production. The course is required by students majoring in Environmental Resource Management (renewable resources). The text provides sufficient depth about environmental factors and plant growth plus a very fine discussion of the technology of crop production for those students interested in land use from an environmental viewpoint."
—John W. Mastalerz, The Pennsylvania State University

"I used the book when I was developing our course, World Crops and Food Supply. . . . The book goes beyond what we need in this particular course, but it is so readable. I am sure it is our best choice for our course. I am pleased with the new chapters on pollution and the environment and on agricultural research."
—Neri A. Clark, University of Maryland

Second Edition, 1974, 740 pp., 355 illus., $14.50

DNA Synthesis

ARTHUR KORNBERG, Stanford University

". . . Genetics and DNA have become a branch of chemistry. Despite its chemical complexity, DNA is being modified, dissected, analyzed, and synthesized in the test tube. There are insights into a metabolic dynamism of DNA that had not been anticipated. DNA suffers lesions and is repaired. DNA molecules are specifically modified and degraded, twisted and relaxed, transcribed in reverse from RNA as well as directly into RNA. DNA functions not only in the nucleus but also in mitochondria and chloroplasts. There is now a stimulus to determine the total base sequence of DNA and to resynthesize it. There is also a confidence that the metabolic gyrations of DNA in the cell can be understood in as explicit detail as those of, say, glucose or glutamate."—from *DNA Synthesis*

Written by a Nobel Laureate in medicine, this is the first comprehensive treatment of DNA synthesis emphasizing its biochemical aspects and recent develoments.

1974, 399 pp., 194 illus., $18.00

Comparative Morphology of Vascular Plants

ADRIANCE S. FOSTER, late of University of California, Berkeley, and ERNEST M. GIFFORD, JR., University of California, Davis

"Since its introduction in 1959, this book has become a classic in its field. The new edition is excellent, truly outstanding, far better than the already excellent first edition. It has no rivals."
—Rudolf Schmid, University of California, Berkeley

"The revised edition is exactly the type of text needed for senior students in Plant Morphology since the authors have retained the excellent organization of the first edition and yet have updated the content without becoming verbose. The illustrations are excellent."
—R. L. Peterson, University of Guelph, Ontario

"I have used the first edition every year except one since its publication. I will be using the second edition this fall for my one-semester course in the morphology of vascular plants. . . . It is the clearest presentation of the subject on today's market."
—J. M. Herr, Jr., University of South Carolina

Second Edition, 1974, 751 pp., 440 illus., $17.00

The Vertebrate Retina

Principles of Structure and Function

R. W. RODIECK, The University of Sydney

"This book is an attempt to explain to myself what is known about how the vertebrate retina is organized and how it functions. What I have tried to present the reader with is a simple and straightforward account, which starts with the fundamentals and leads to the frontier."—R. W. Rodieck

The Vertebrate Retina—the most comprehensive volume now available on the subject—draws from a wide range of methodological approaches. The book contains 615 illustrations (plus a twelve-page full-color section) and includes extensive citations to key original literature. One of the appendixes is an English translation of Ramón y Cajal's classic monograph, *La Rétine des Vertébrés*.

1973, 1044 pp., 615 illus., $39.50

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

THE AMATEUR SCIENTIST

The pleasures and problems of raising snails in the home

Conducted by C. L. Stong

Last summer I put half an inch of mud from a nearby pond in a glass jar, added water from the pond and set the jar on a window ledge. The water soon teemed with microorganisms, each species seemingly bent on tipping the ecological balance in its own favor. Then a pair of pinhead-sized specks appeared. They clung to the inner wall of the jar, and when they moved, they left clear trails on the smudged glass. In a remarkably short time the specks put an end to my fun by maturing into a pair of drab and voracious snails that apparently either ate or poisoned all the other organisms.

I complained about the snails to Robert A. Brown, formerly an associate curator of the New York Zoological Society and now managing director of the Anti-Cruelty Society (157 West Grand Avenue, Chicago, Ill. 60610). "But snails are remarkable animals!" he exclaimed. "Don't put them down until you get to know one personally." Subsequently he wrote the following procedure for getting to know snails.

"Snails conduct themselves in bizarre and intriguing ways. For example, they become dormant under a number of unfavorable environmental conditions, such as when they are too hot, too cold or too dry. Land snails are particularly easy to observe and care for. They represent one of the most important groups of invertebrate animals, the mollusks. Some 80,000 species of mollusk have evolved, of which about half survive.

"Of the several thousand kinds of land snail, perhaps the easiest species to maintain and observe is *Otala lactea.* It is shipped in wicker baskets as food to New York and other metropolitan centers from North Africa, principally Morocco. It is a close relative of the escargot, the prized food snail of French gourmets. Paris reportedly consumes more than 200 million escargots per year.

"*O. lactea* is somewhat smaller than the escargot and is dark brown inside the shell opening. It can be found on sale as *babaluce* in the fish markets of Italian communities. If you cannot find a supply in your locality, have a friend who lives in a big city send a few. Alternatively, you can order land species from biological supply houses such as Ward's Natural Science Establishment, Inc. (P.O. Box 1712, Rochester, N.Y. 14603), or Turtox/Cambosco, Macmillan Science Co., Inc. (8200 South Hoyne Street, Chicago, Ill. 60620).

"The distributor may not have *O. lactea* in stock when you place your order. The inventory varies with the season. The procedure to follow, however, applies to land snails of any species. Three individuals are sufficient. They will multiply. Moreover, the labor required to keep snail cages clean is not negligible, and it increases in proportion to the number of animals.

"The snails will doubtless arrive in the state of dormancy known as estivation. They enter this state when they are warm and dry. Estivation is the counterpart of hibernation, the period of dormancy some animals enter during cold winter months. In Europe the escargot may estivate in summer and hibernate in winter.

"Observers have reported that various land snails have survived in estivation for four or five years without emerging from the shell. When the animal is in estivation, it will not become active unless it is either exposed to moisture or disturbed. When you receive the animals, keep them in a dry place until their cage is ready.

"Examine each specimen carefully. Mark its shell distinctively with a colored lacquer such as nail polish. I usually enter the identifying mark at the top of a page in a notebook. I weigh each individual on a postage scale and record its weight in the notebook along with its coloring and any distinctive markings.

"Observe in particular the opening of the shell. No doubt it will be closed by dried mucus. This thin, leathery film helps the animal to conserve its moisture. Even so, snails can lose up to 50 percent of their body weight during estivation. The weight will be regained quickly after the animal recovers from the dormant state, as you can observe in detail by weighing each specimen hourly during its first active day.

"Although snails are easy to keep, their cage should be designed to avoid three problems that frequently arise. Unless the environment is right, snails that have been roused may reenter the state of estivation. The animals may also tend to become inactive during the daylight hours, primarily because they are nocturnal by nature. Finally, snails quickly foul their enclosures with trails of slime and excrement. The cage must be cleaned frequently, a matter that should be kept in mind when it is bought or being designed.

"The cage or housing can be of the type I refer to as the 'hungry man's aquarium.' It is a clear plastic hatbox or shirt box of the kind found in stores that specialize in housewares. They are currently priced at about $2. A box has a volume of two or three liters and a loose-fitting lid that provides adequate ventilation. Moreover, there are no metal parts to rust or otherwise corrode. If the lid is airtight, melt a few holes in the plastic with a heated skewer or nail.

"Small glass aquariums that are available from distributors of novelties and dealers in tropical fish are equally satisfactory. The new container should be washed with soap and water and rinsed thoroughly. Place a thin layer of fallen leaves, a lettuce leaf, bits of carrot and a twig or two on the bottom along with a shallow container filled with water to a depth of from three to five millimeters and a second container of good soil of the kind sold for potting houseplants. These materials appear to stimulate the snails by providing a diversified environment in terms of smell and taste.

"The container for the soil can be a

small flowerpot or a rectangular plastic box. The soil should be at least two centimeters deep. Keep the soil moist but not flooded. The animals will doubtless lay their eggs in it.

"Normally the animals will be most active at night. It is possible, however, to reverse their cycle of behavior for observation during daylight hours. Flood the aquarium all night with light from a bright incandescent lamp. In the morning cover the aquarium with an opaque cloth. To observe the animals during the day pull down the window shades and remove the cover. Snails may adjust to reversed day and night within a week or so. If this technique fails to arouse the snails during the daytime, try placing them in a shallow pan of water.

"Do not overpopulate the aquarium. Three snails are enough. If you want six, set up a second aquarium. Do not put into the container any material that cannot be washed or discarded, except the soil. Remove the snails and all the materials twice a week and then wash and rinse the aquarium thoroughly to remove slime and excrement. Discard the uneaten foraging materials and replace them with fresh food and natural plant litter.

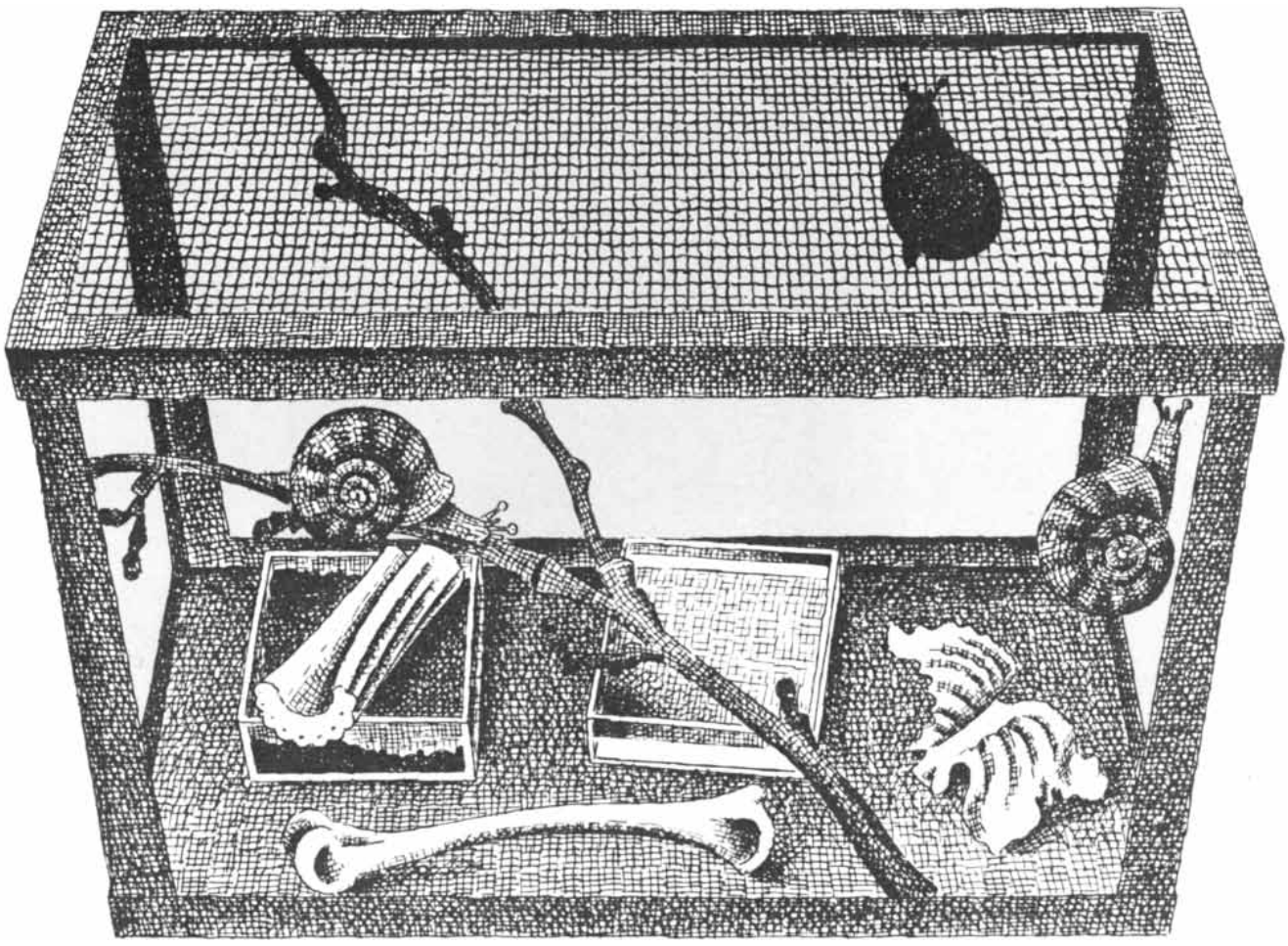
"Natural objects, such as stones and pieces of broken flowerpot, serve well as ornaments in the aquarium. A clamshell or an old bone should always be included. It will gradually be eaten because a snail requires a relatively large amount of calcium. A box of fallen leaves and sticks can be collected on nature walks and stored for use during the winter.

"Rouse the snails only after the aquarium is ready. All you have to do is put them on a moist surface in their new home. Keep in mind that you are working with land snails, not the aquatic kind. Do not put more than three millimeters of water into the aquarium or the snails may drown. Within about 10 minutes one snail or more will stir and crawl up any nearby surface. If some of the animals have not moved at the end of a day or so, they may be dead. Dead snails can be identified after a few days by their strong unpleasant odor!

"Observe the animals frequently during the first day after they have come out of estivation. If the air inside the aquarium is humid, as it ought to be, the snails will crawl up the sides and move upside down onto the bottom of the lid. You should be able to see at least one of them lift the forepart of its foot, head and tentacles off the lid, twist around and work its mouthparts over the opening of its shell. The animal is eating its first meal after fasting for many days or weeks; it consists of the remains of the dried mucus film that covers the body during estivation.

"Land snails are not fussy eaters. Although they will eat a wide variety of plant and animal matter, it is fun to investigate their food preferences. Once I offered four snails a bit of tuna. Three of them spurned it, but the fourth ate it all. Evidently the animals develop individual preferences.

"The diet should include cereal such as dry oatmeal flakes, bran and packaged breakfast foods. As I have mentioned, calcium is essential. Sprinkle a few pinches of ground limestone or bone meal on the vegetables in addition to keeping a clamshell or a piece of bone in

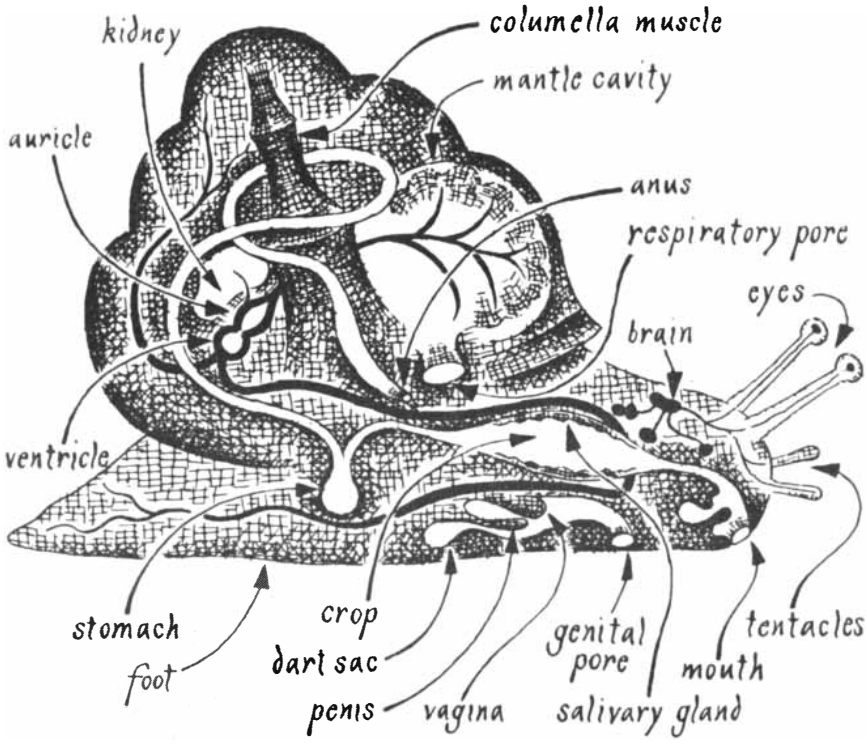


Arrangements recommended by Robert A. Brown for keeping snails

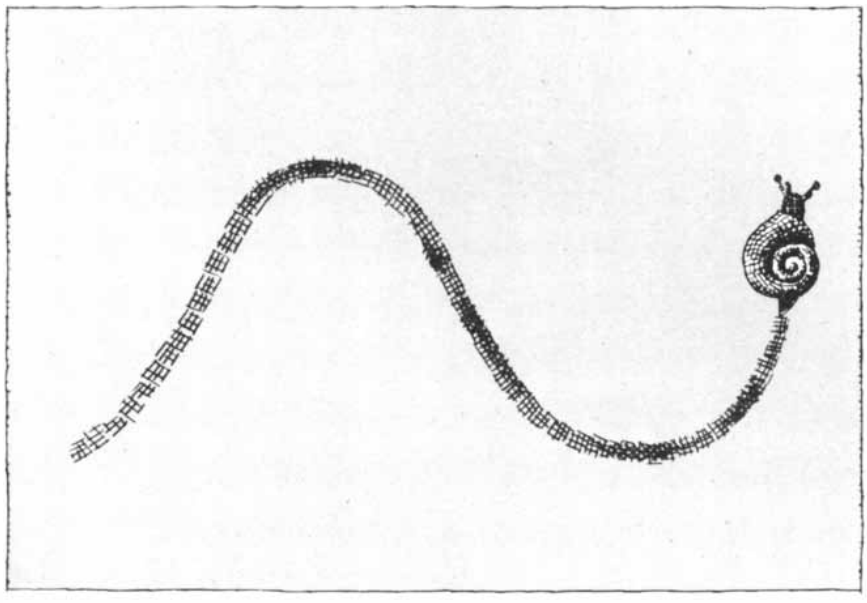
the aquarium. Limestone can be obtained in the houseplant sections of supermarkets; bone meal is available in pet shops.

"Calcium is the building material of the snail's shell, which grows throughout the life of the animal at the lip of the shell opening. Land snails in the wild get calcium from a variety of organic materials. Much as man ingests calcium for his bones from milk, so snails consume the remains of other animals. In the absence of adequate calcium land snails may become dormant.

"The two most obvious features of a snail are its shell and its foot, which is the long extension of the body on which the animal glides. The mouth is at the front of the foot and can be observed through the glass or plastic when the animal moves. The mouth is a round hole when it is open and a *T*-shaped slit when it is closed. The mouth is surrounded by sensory flaps called lappets.



Anatomy of a snail



Tendency of snail to move upward as position of vertical sheet of glass is shifted

"The snail combines the functions of teeth and tongue in a single organ: the radula, a toothed, filelike muscle inside the mouth. The snail employs it as a scraper to masticate food. The radula can be seen when the mouth is open or when one directs a strong light through the head from the top and observes from the bottom.

"Keep quiet and listen carefully when a snail is eating hard materials such as crisp lettuce and carrots. You will hear a distinct scraping noise as the radula works its way into the food. You can also feel the radula by putting a thin film of honey or jam on the tip of your finger and offering it to a snail. The animal will clean off the film with the radula. It is interesting to record in the notebook the food preferences of the individual snails in an aquarium.

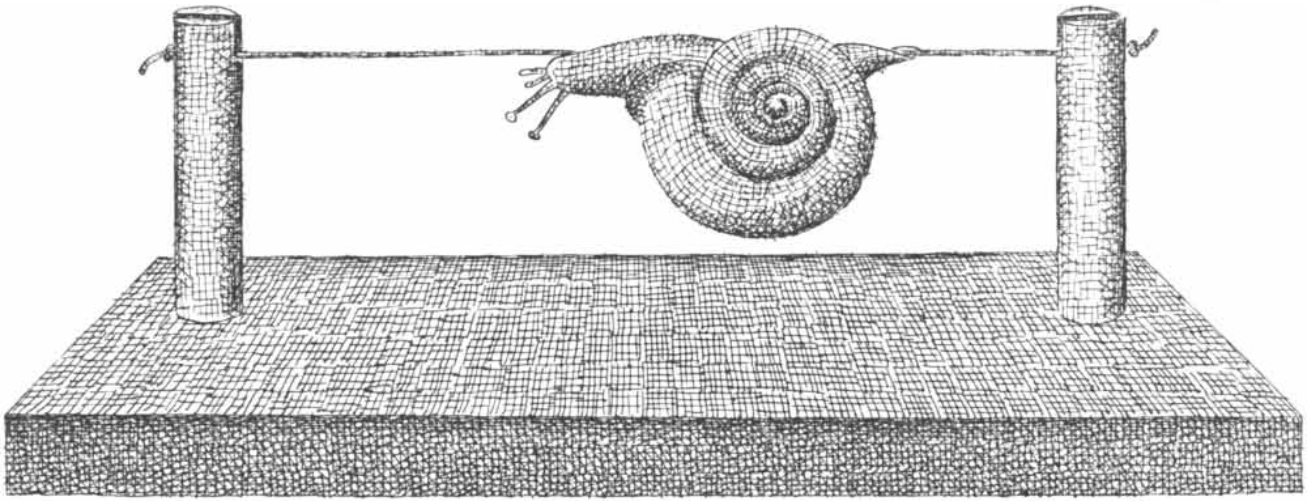
"The anus appears on the right side of the body when the animal comes out of its shell. It can be detected easily when excrement emerges, particularly if a highly colored food has been eaten. Observe in particular the sensory organs of the head. They consist of front and rear tentacles. The rear tentacles are longer. At the tip of each one is an eye. When the tentacles are touched, they withdraw by turning inside out. They can be seen in detail with a bright light and a magnifying glass. Retracted tentacles emerge by the reverse procedure of turning right side out.

"Note the dark, fleshy tissue around the edge of the shell opening. It is known as the mantle. Among other functions it secretes the shell. The mantle also forms a cavity that, in the land snail, serves as a lung. Look upward toward the shell opening when the snail is fully emerged. You can see the large air hole that leads to the lung cavity. It is usually open.

"Snails are hermaphrodites. Each individual is both male and female. The animals engage in an elaborate and easily observed courtship that culminates in cross-fertilization. Mating snails fondle each other for a long time and then shoot each other with a calcareous dart before mating. Mating takes place when a large sex organ emerges from the right side of the head. The sex organ of each snail penetrates the sex organ of its partner, and sperm from each animal fertilizes the eggs of the other.

"Speck-sized eggs are laid in soil by both snails. Whenever you clean the aquarium, look carefully for the eggs. Disturb them as little as possible. Although the eggs are fragile, they may hatch. If they do, you can set up a separate aquarium for the young.

"To investigate how snails move, put



Acrobatics of a snail on a string

a snail on a sheet of clean glass or clear plastic. Hold the sheet up vertically. The snail will move upward. Rotate the sheet slowly back and forth, keeping it vertical. The animal will sense the change in the direction of gravity and turn so as to maintain its upward course.

"Examine the motion of the foot through the glass. Waves of muscular contraction will appear to travel from the rear of the foot to the front. The waves seem to move contrary to the snail's forward motion. Actually the forward-moving part of the foot is not in contact with the surface of the glass and is therefore less conspicuous.

"The prominent trail of mucus that appears in the wake of the snail is secreted by a gland you can see just behind the mouth. The mucus makes possible the snail's gliding motion. It is easy to demonstrate the acrobatic talents of a snail. For example, put a snail on a string or a fine thread stretched horizontally between two supports. The animal will move along the line as readily as it moves along a flat surface. It can also glide across prickly surfaces, such as the barbs of a cocklebur, without injury or apparent discomfort.

"One can determine by observation and a few experiments that all parts of the snail's body sense light, even though the animal has eyelike organs at the tip of its longest tentacles. The snail's reaction to light can be tested in various ways. For example, install a small incandescent lamp on each side of the aquarium. Light the lamps alternately, simultaneously shading the unlighted side. The snail will move to the darker side.

"The animal may react to the abrupt approach of a small shadow by retreating into its shell. Pass the shadow of your hand over a snail. The animal appears to interpret the shadow as a threat.

"Investigators who have removed the eyes on the longer tentacles have concluded that they are of minor importance, since the behavior of animals so deprived does not change significantly. The eyes appear to fix on objects only at very close range. You may occasionally see a snail bend its eye tentacles downward enough to fix on the supporting surface.

"How does a snail react to odors? You can test the sense of smell by placing warm, cooked foods on a moistened surface. Note the seeking movements the animal makes, elevating the front part of its foot and waving its tentacles.

"Are the foods sensed by smell or by sight? Put two or more opaque screens on the surface with the snail between them. The screens can be strips of sheet metal bent into an *L* shape. Observe the behavior of the animal with respect to the screens. Then put the warm food behind the screens and note the altered behavior.

"Can the animals distinguish between foods of various kinds? Do they exhibit preferences? To learn the answer put a drop of vinegar in the path of a snail. It will turn aside. (Do not put vinegar on the animal.) Occasionally a snail will merely retract the edge of its foot that is nearest the drop, thus forming a pocket. The pocket will move from the front to the rear along the edge of the foot as the animal glides forward.

"Does a snail hear? The answer is doubtful. The animal will react to loud, abrupt noises, perhaps because the air vibrates the shell. A snail does have an extremely acute tactile sense. The tentacles retract at the slightest touch. For example, a single filament of cobweb that drifts in contact with a tentacle causes the organ to retract. On the other hand, as I have mentioned, the foot is not injured by sharp objects. A snail can crawl over the edge of a razor blade without cutting itself.

"As one might expect, the snail reacts strongly to moisture. After cleaning the aquarium put a shallow divider across the bottom with a layer of dry aquarium gravel on one side and moist gravel on the other. To which side does the snail move?

"The animal's reaction to gravity is of course related to its sense of touch. You can base an amusing apparatus on this reaction. Make up a miniature seesaw by cementing the middle of a glass microscope slide across a length of string that is stretched about two centimeters above a tabletop. Put the snail on the slide at the bottom. It will crawl up the slide, which will tip downward when the snail crosses the balance point. The animal will then turn around and proceed in the opposite direction until the seesaw tips again. Limit the experiment to a few cycles.

"What is a snail's pace? The maximum velocity at which a snail moves appears to be influenced by several factors. Snails tend to move uphill toward shadow, favored foods and moisture. Certain snails may respond more strongly than others to various combinations of these stimuli with the result that the pace varies with both individuals and circumstances. I find it interesting to time each animal's progress over a measured distance of, say, 20 centimeters when it moves away from the light, uphill and toward water. By recording the preferences of individual snails one can calculate odds for handicapping a snail race."

Discover

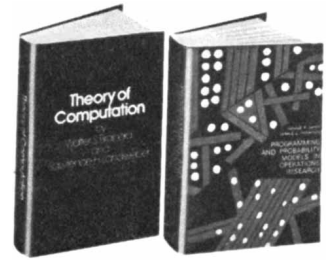
the one organized (and inexpensive) way to keep up with what's new in computer techniques and management—

The Library of Computer and Information Sciences

Take any 3 books for only $1 each

(values to $53.90)

You simply agree to buy 3 more books—all at handsome discount—over the next two years.



(Retail prices shown)

60290. MANAGEMENT PLANNING FOR DATA PROCESSING. *Dick H. Brandon.* Accents the quantitative information necessary for anyone concerned with the operation of or planning for computers **$12.50**

59840. MAN-MACHINE COMMUNICATION. *Charles T. Meadow.* Human beings are slow, noisy, narrow-band, and inventive; computers are fast, accurate, wide-band . . . and unimaginative. Ways and means whereby the two can "talk" to each other. **$15.95**

52250. HANDBOOK OF SUCCESSFUL DATA PROCESSING APPLICATIONS. *Jerome K. Pescow, Jack Horn and Marvin Bachman.* How to save time and dollars by obtaining the maximum value from a computer. *Counts as 2 of your 3 books.* **$24.95**

42500. DOCUMENTATION STANDARDS (Second Edition). *Keith R. London.* Complete guide to the development, design, and implementation of a standard system of documentation. **$12.50**

44030. ELECTRONIC COMPUTER MEMORY TECHNOLOGY. *Edited by Wallace B. Riley.* Complete "state of the art." **$17.00**

42045. DESIGN OF MAN-COMPUTER DIALOGUES. *James Martin.* Thoroughgoing study of how effective interaction can be achieved between man and computer. **$16.50**

56191. INTEGER PROGRAMMING. *Robert S. Garfinkel and George L. Nemhauser.* One of the most exciting and rapidly developing areas of operations research **$18.95**

55370. INFORMATION PROCESSING MANAGEMENT. *Ralph A. Szweda.* How today's data processing manager can be not only technically proficient but also managerially skillful. **$12.50**

70730. PROGRAMMING AND PROBABILITY MODELS IN OPERATIONS RESEARCH. *Donald P. Gaver and Gerald L. Thompson.* The application of mathematical quantitative methods to the solution of practical systems problems. **$17.95**

42261. DIGITAL COMPUTING AND NUMERICAL METHODS. *Brice Carnahan and James O. Wilkes.* A brilliant presentation of methods with practical application to digital computers. **$13.95**

53720. HOW TO BUILD AND USE ELECTRONIC DEVICES. *Stuart A. Hoenig and F. Leland Payne.* Delightful volume shows you how to have fun building numerous useful gadgets—even an analog computer. **$9.95.**

40061. COMPUTER SECURITY MANAGEMENT. *Dennis Van Tassel.* "Must" reading for all computer professionals. **$10.50**

79230. SOFTWARE SELECTION. *Susan Wooldridge.* Everything you need to know to make wise software selections. **$9.95**

41620. DATA BASE MANAGEMENT. *William C. House.* Surveys every important aspect of data processing management. *Counts as 2 of your 3 books.* **$19.95**

81880. SYSTEM ANALYSIS TECHNIQUES. *Edited by J. Daniel Couger and Robert W. Knapp.* 25 carefully selected papers by the leaders in the field. **$14.95**

83760. THEORY OF COMPUTATION. *Walter S. Brainerd and Lawrence H. Landweber.* A totally new presentation of the theory of computation in a mathematically rigorous fashion. *Counts as 2 of your 3 books.* **$19.95**

56370. INTRODUCTION TO PROGRAMMING LANGUAGES. *Harry Katzan, Jr.* A clear introduction to today's three most widely used languages. **$13.75**

53900. J.K. LASSER'S HOW TO RUN A SMALL BUSINESS: Fourth Edition. *J.K. Lasser Tax Institute.* A completely revised and updated 4th edition of the classic. **$9.95**

The Library of Computer and Information Sciences
Riverside, New Jersey 08075 7-3AA

Please accept my application for membership and send me the three volumes indicated, billing me only $1.00 each. I agree to purchase at least three additional Selections or Alternates during the first two years 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 anytime 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 $1.00 each Indicate by number the 3 books you want.

Some expensive books count as 2 selections

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.

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 three to four 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 book at all, simply indicate your decision on the reply form always enclosed with the **News** and mail it so 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. • After completing your trial membership, you will be entitled to take advantage of our fabulous bonus plan

BOOKS

A vision of an architecture that is not only visually pleasing but also humane

by Philip Morrison

AMERICAN BUILDING 2: THE ENVIRONMENTAL FORCES THAT SHAPE IT, by James Marston Fitch. Houghton Mifflin Company ($15). Two paired photographs eloquently represent the polemic side of this clarifying essay on the bases of architecture. The pictures show a beautiful piece "of cantilevered terrace and roof," "hovering planes" designed by the great formal architect Mies van der Rohe. The first photograph is calmly beautiful, but in the second one the inhabitants of this Illinois house have been at work. The elegant clear spans are shaded in by an insect screen and blocked with chairs, because when the summer heat drove the owner to his porch, the insects rendered his hours impossible. The owner made the space habitable at the price of "radically altering its esthetic impact."

The book, a revised and enlarged edition of *American Building: The Forces that Shape It,* elevates to an articulate philosophy what most Americans must feel: Formal beauty—celebrated primarily in photography, that purely visual record far from the polydimensional reality of noises, leaks and flies—is no substitute for honest design based on need. "For architecture, the solution of experiential problems is the only source of valid form." Professor Fitch, a distinguished Columbia University architect and architectural historian, reaches this position like the builder he is: by his keen critique of what is wrong but even more by an affirmation of what is right.

Human beings need architecture as environment for the "natural" actions of our domestic life and even more for the "unnatural"—the culturally rather than biologically constrained—circumstances of work. Heat and cold, a breath of air (odors too), light and sound, space and time and gravity, structure itself and the social arrangements of integration and plan shape Fitch's chapters. Throughout he keeps his attention on the aesthetic, the "essence of the human experience." To design a building that is both beautiful and good, "as Gropius once so movingly put it, is an authentic act of love." Here is no mere specialist's scoffing; Fitch includes the graphs of human tolerances and arm reach not to promote this or that solution but to make plain that the architect's true client is the whole person, not simply the viewer, the lathe hand, the housewife or the copywriter.

For Fitch good architecture must share responsibility for public health. Falls still cause most domestic injuries, but they are not mere random accidents; they center on steps and in the bath, just as burns and cuts cluster in the kitchen. It is manifest that searching design could help, but architects study home safety too little. It is no excuse to argue that matters such as fixtures, maintenance and sources of smog are out of reach of the architect. "Legally, his responsibility may be limited; morally, it is very great."

Why is the igloo so fine a winter shelter? Why is the adobe house—together with its pueblo outgrowth—so fine a desert one? The designers, by cultural selection over a long time, work under the severe discipline of a rude imperative from the environment and an austere choice of materials. Their necessity is their freedom. The result is precise understanding of microclimate and a profound grasp of the nature of materials. The modern architect is the reverse: he consistently underestimates the impact of environmental forces and he does not even choose the forms and materials of his components.

Bridge builders and the designers of yachts and aircraft make few such errors; theirs are the only structural triumphs of our world to match the igloo. Such architects as Ralph Knowles of Los Angeles and the Anglo-Swedish designer Ralph Erskine are Fitch's paradigms. These men seek to rework material and form. Knowles projects a skyscraper wall at 30 degrees north latitude in which the structural material is organized perpendicular to the wall plane, its thousands of cellular spaces rationally arranged to control the sunlight over the day and season, allow maximum transparency from within and yet admit wide spans. The complex honeycomb is surely beautiful, and yet it is the opposite of a manufactured glassy curtain. Erskine deals with arctic cities, dark and frozen all winter. His town project is a bowl tipped southward with two levels for pedestrian circulation, one of them sheltered and lighted.

In the past the building consumer was also in part a producer. Design once rested in the hands of a million anonymous artisans. The external conditions "established the limits of satiety and survival alike." Now we are all but freed from those limits by a powerful remote technology over which architect, client and even the decisive dollar retain only meager control. Take the theater, at once architectural vessel and artistic content. Fifth-century Greece made it a bowl whose cunningly designed roofless seats were symbiotic to the sights and sounds of the great plays that grew there by daylight. Early Christian polyphony "relinquished the intelligibility of speech" for its melody, made "both possible and necessary by the long reverberation period of the Romanesque basilica."

Today film has lost the magic of feedback to the living performer, but the assembled film audience, if it is well served by modern audiovisual technology, can still richly enjoy a viable art form. Fragmented into sealed automobiles, however, the drive-in audience must stare at the distant screen through a convex and foggy windshield as the sound blares from a tiny box. Their experience is "a travesty" of technical virtuosity. Finally, the "indecent privacy" of the television screen reduces the public participant in the high art of theater to a passive spectator straining at a flickering knothole.

Vision alone remains, in part. And that bias for visual appearance over sound, feel, smell, touch, community is also the occupational disease of the architect. (Photons, alas, are too cheap.) His remedy must be to serve knowingly all the senses, to defend his clients, to form a rapport of mutual interest with the ordi-

nary consumer of buildings; in parable, to stop working on the planner's scale of one inch to 100 feet. On such a model "decisions affecting 68-inch people are made from an altitude of 3,500 feet." Architects must come back to earth.

VACUUM MANUAL, edited by L. Holland, W. Steckelmacher and J. Yarwood. E. & F. N. Spon Ltd., distributed in the U.S. by Halsted Press, a division of John Wiley & Sons, Inc. ($33.50). A handsome book, this volume is so workaday that it carries advertisements. It is a brand-new overview of the industry of vacuum making and using, its editors are three Greater London experts, its viewpoint is international, firsthand and skeptical. More than half of the book is a section on the available equipment, listing the products of 150 manufacturers by category, with tabular comparisons and brief descriptive text. (There are no illustrations in this section, and readers new to the art need to strain for hints about what is being described.) The volume is meant for people with some minimum of sophistication in the field. Besides the equipment survey, the manual offers a very good chapter on basic vacuum data, itself most up to date in measurements of the properties of new materials and systems, and reviews centering on the latest thing (since 1968) in three active domains of vacuum technology: pumps, surface-analysis systems and the deposition of controlled films by sputtering.

It was just 25 years ago that this magazine carried a piece on the art of vacuum. The biggest change since then has been the growth in the complexity and variety of approach. Not much that used to work has become passé; the newest developments are simply added on. Everything finds some use today in the very wide range of the technology. Television tubes and simulated cislunar environments, freeze-dried strawberry ice cream, well-made new dentures and a wild proliferation of the cunningly controlled surface films forming the electronic devices that are so ubiquitous in our world—all of these represent modern vacuum uses. They stretch far beyond the physicist's orbiting accelerator beams and the chemist's clean surface.

The little oil-sealed piston pumps are still popping steadily in the laboratories of the world but there are no longer only a few manufacturers. They spread instead from Liechtenstein to Van Nuys and offer a model for every nuance. Want to produce a lot of rough vacuum quickly? Consider a big steam ejector pump. The steam jet entrains the air and you can bring a full-sized room down to a millimeter of mercury in three minutes. It will require a five-stage steam ejector and will cost you about a ton of steam. "The most outstanding recent development" in pumps, however, is the rotary pump based on the Wankel principle. Like its cognate automobile engine, it is quieter, vibrationless—no popping—and performs somewhat better than a typical little piston pump.

Another new-old device is the molecular drag pump, now in the variant called turbomolecular. Since 1912 this principle has been used by a very few laboratories with virtuoso machining abilities. The gas is pumped by spinning steel disks with very low clearances. Caught between disks, a molecule is dragged along bouncing from disk to disk, picking up a velocity in the direction demanded by the spin. Such a pump is truly "clean," or free of any gas of its own making, because it is built of special steels and often comes fitted with a built-in heater that keeps it all but red-hot until the dissolved and adsorbed gas has left it. The new version is commercial, a product of the formidable aircraft-engine industry, with 10 or 20 rotor and stator disks: a miniature axial-flow compressor designed for pumping. It is suitable for ultrahigh vacuums, nowadays down to the limit of measurement at about 10^{-13} millimeter of mercury (the manual uses the MKS pressure unit, the pascal, almost throughout). Its forte is heavy, slow molecules, such as those of air; it does much less well with the far nimbler hydrogen. (To hydrogen the turbine seems slow.)

Does the residual hydrogen bother your clean surface? Such a reactive gas is nicely swept out by titanium sublimation. Steadily evaporate an unreeling titanium wire—say by electron bombardment—and let the metal vapor condense on a cooled surface in a pretty good vacuum. The newly formed active metal film will rapidly combine with any loose hydrogen or nitrogen around. Yttrium metal is even better (and dearer); you can finish up with that. Catch titanium film on a liquid-nitrogen cooled surface and you can expect to trap more than 80 percent of the air molecules that strike it, and even hydrogen atoms every few times they hit.

Helium disdains such chemical lures. Noble-gas pumps, however, are not scarce in this ingenious technology. Use a titanium cathode, a strong electric field and a discharge to ionize the helium, and the newly formed ion will be forced to plunge into the active cathode, there to bury itself beyond return. Panels internally cooled by liquid helium or by deeply chilled helium gas catch almost everything that strikes. These pumps are pumps of last resort, so to speak, and can be had so far from only two manufacturers.

Pumps are the active elements of vacuum technique. The passive side is quite as important. Gas leaks out of steel and glass, boils in torrents out of softer materials and even makes its way past tight seals from the ocean of the air. The standard vacuum container for ultrahigh vacuum is cobalt-nickel stainless steel, well machined, degreased with vapor and even polished. Pump it, bake it out for days—and repeat for a couple of hours every time the surface is exposed to the air again. Even then hydrogen and carbon monoxide will continue to emerge slowly from the depths of the steel. Any polymer surface will outgas (a "virtual leak" into your clean volume) two orders of magnitude more than steel. Small areas of gasket or insulator or the like are all that is tolerable. For ultrahigh vacuum, valves may well be sealed with stainless-steel bellows; their gates are stainless-steel knife-edges homing on beds of gold-plated soft copper.

In the old days there were special mineral-oil distillates with low vapor pressure and few volatile impurities for the sealing greases and pumping fluids. It is interesting to see that these are still being made and sold nearly 50 years after their early development (under the same trade name but now by Shell). There are many competitors, one made by the Lion Fat & Oil Company of Tokyo (not from the fat of the lion). There is much new chemistry, of course. Silicones and various fluorocarbons rule the roost for low pressures. (If these expensive stuffs break down, the familiar dark carbon films we once saw may instead be silica deposits. Sand in the works?)

There is no room to tell of many new instruments. One neat device is a film-thickness monitor, which works by measuring the resonance of a high-frequency quartz-crystal oscillator that grows heavier and a little slower as a film forms on it in the vacuum. These too are commercially packaged systems, with everything else supplied for depositing a few controlled layers of materials to make square feet of logic chips, the heart of a good many thousand pocket calculators.

Only prices are missing from this admirable guide to a world of high-technology commerce. String and sealing wax are gone (although not picein wax,

quite). One imagines that Ernest Rutherford would be intrigued but would feel strange in this place of new combinations and logical developments of the physicist's tricks of a generation or two ago.

ENCYCLOPEDIA OF MINERALS, by Willard Lincoln Roberts, George Robert Rapp, Jr., and Julius Weber. Van Nostrand Reinhold Company ($69.50). The atom fixes the minimum scale of crystals. Once 100,000 atoms or so have struck a collective bargain between bond energies and thermal motion, a crystal is born. Its surface may be strange, its imperfections notable, but these remain oddballs in a conventional atomic union. (A mere local union of atoms may not even reach crystallinity: it is all surface and eccentric voids, perhaps even with a foreign nucleus.) Only time and a rare tolerant environment can enable a crystal to attain the enormous population size that becomes a hand specimen for the collector or even an example for the jeweler's craft. Individual mineral crystals on the human scale are rare and are therefore prized.

Nowadays, with zoom optics, long working distances and color correction, the large-scale production of convenient stereoscopic microscopes in the magnification range of one or two powers of 10 is commonplace. (It is the microcircuit industry that has provided the incentive for such developments.) With them the collector can enter the crystalline world of minerals with complete satisfaction to the eye, down to a scale millions of times below the museum-case specimen in mass. Through the lens he sees crystal habit and true color, sharp facet and contrasting substratum as well as he could ever see them in a larger specimen held in his hand. His domain of collection, however, has wonderfully expanded as the sample size has shrunk. Rarities become available and unique specimens become realistic goals.

Micromounts of minerals are not new. They were known among microscopists of the 19th century, and the current revival began just after World War II. Now it has reached a high degree of refinement in the hands of such connoisseurs as Julius Weber, virtuoso of the photomicrograph, about 1,000 of whose knowingly chosen, always sharp and stunningly colorful photographs the size of a business card illustrate this book of reference.

Weber's personal collection includes some 9,000 micromounts; he not only persuaded a few dozen people to lend him this or that single example but also got two other redoubtable collectors of micromounts to pool their samples, the finest in America, for the purposes of this volume. (The transparencies from which the reproductions were made are, somewhere, a hidden resource for the entire studentship of minerals.) The pictures are presented on sized paper insets in alphabetical order, 15 or 20 pages at a time, scattered through the thick rice-paper volume of text.

It is hopeless to convey a visual museum. Turquoise? No massy sky blue gem but here a set of sharp crystals gleaming in the familiar hue. Pyrites? A dozen examples, the most remarkable a sharply bent wire form looking like a piece of three-dimensional printed circuit. Rare? Examine the inky blue pyramids of callaghanite, a hydrated double copper carbonate that occurs only in the "pits of Gabbs Refractories, Inc." in Nye County, Nev. Cinnabar? Again in twinned crystals. Emerald, as a kind of starry vein. For the admirers of nature's gem design, a cuprite specimen from the famous Tsumeb mines of southwestern Africa, Chinese red crystals dotted with tiny green studs of malachite.

The text lists alphabetically more than 2,200 minerals and gives for each the chemical formula, crystallographic information (including lattice constants, the main powder X-ray-diffraction lines, refractive indexes and crystal habit), customary mechanical properties, color, luster and a laconic summary of occurrence. The best reference known to the authors and available in English is appended to each listing. All these data have been critically compiled by the first two authors, both expert professional mineralogists.

The text is not for beginners, but this is a quick, complete, authoritative reference work in one big volume for "a very broad audience," wider than those familiar with more refined systems of classification. The authors hope the volume will find its home as much on "the encyclopedia shelf of the small-town library as on the desk of the established mineral collector." Given a real need to go well beyond the many excellent primers, libraries and collectors would be wise to act.

WATER: A PRIMER, by Luna B. Leopold. W. H. Freeman and Company ($2.95). WATER: A VIEW FROM JAPAN, text by Bernard Barber, photographs by Dana Levy. John Weatherhill, Inc. ($30). The primer is about water, not the molecule or even the hydrogen-bonded fluid but the water of the hydrologist: whence it comes, how it runs off or soaks in, the nature of rivers, floods and moist soil and how we use them. Professor Leopold is one of the best-known hydrologists, and his simple paperback text for undergraduates is full of insight, even though he allows himself only one equation. (It is spelled out in words: Rate of outflow = rate of inflow ± change of storage.) Most terrestrial water is of course seawater (97 percent), but we live by the residual. Of that 3 percent the ice caps hold more than 2; what cycles each year is only one part in 3,000 of the total. Most of this falls back into the sea; the amount flowing in stream channels is only a couple of weeks' worth of rainfall on land, although lakes hold a three-year supply. Underground water (very poorly known) is perhaps comparable in quantity to the store held as ice.

The U.S. Geological Survey monitors the channel flow in America from more than 10,000 stations that record the discharge over time (called the hydrograph) for just about every stream of consequence. Here are photographs of Seneca Creek, which flows into the Potomac some 20 miles above the Watergate, in stages from low flow to flood, the supporting quantitative material given real texture by the pictures. We learn how flood crests grow from many upstream contributions, suitably delayed and then summed. The floodplain is built up and taken away again over centuries by the shifting stream but remains about the same in size, filled only by rare floods too great for the channel. There are graphs of streamflow over the years for eight U.S. rivers; they show strong peaks and dips, yet a pattern stubbornly refuses to disclose itself.

Leopold offers no panacea. Ground water is not easy to increase by soil change; flood control cannot always be achieved by sound upstream management. With heavy rain the runoff from saturated soil is about the same for farm and for uninhabited woodland. ("Many of the highest floods known occurred before man began logging or farming.") Large and small dams each have their uses. Zoning to require rational occupation of the floodplain (with foreknowledge of the risk of flood damage) is a hopeful method and is not yet generally practiced. A couple of pages of annotated references would have made this excellent book much more valuable for those who read it without an instructor.

Water: A View from Japan is an unusual ethnography with a text by an

acute resident observer of Japan and photographs by another longtime resident whose extraordinary images were described here four years ago. (Then they were of bamboo.) Here is watery Japan: water pouring in feathery fall as a ritual of purification on a willing victim in his white robe; water dotted with great rafts in Ago Bay, where cultured pearls are secreted by oysters in baskets tied under the floats; water simulated by meticulously raked garden sand or done in line by masters of the woodblock cut or of golden lacquer; water feeding the nation in the myriad mirrored squares of paddy rice, green with the promise of the young plants; water trickling its melody in every garden. "As we traveled around Japan we saw . . . a lot of water which was disgustingly polluted. . . . But here and there we saw efforts being made and evidence that the long journey back has been started. . . . Carp are once again living in Tokyo's Sumida, a river that has long been little better than a cesspool . . . a good omen."

A small neat boy walks in the rain under a cheerful yellow parasol. In English, children sing, "Rain, rain, go away," but the Japanese children's song says, "Rain, rain, please come." This is a happy evocation of a subtle and complex way of life, a superior ethnography in image and text. No visitor can forget either the hot tubs or the cool water splashed daily at house, inn and shop entrances. The water cleanses and allows floors of rock and stone to gleam to their best advantage, but it is ritual as well.

Kappa are the most Japanese of imaginary beings; they dwell in every mountain stream, wry and quarrelsome. They love cucumbers and raid gardens for them. They cannot survive out of water; during their sketchy forays onto land they carry the life-giving water in a concave hollow in their head. This volume somehow makes much of Japan visible in a small compass; in a way the Japanese themselves are like *kappa*, they "have water in their heads and they cannot live without it."

SYNTHESIS OF LIFE, edited by Charles C. Price. Dowden, Hutchinson & Ross, Inc. ($22). The title promises more than the book can deliver. Nevertheless, in these 26 facsimile "Benchmark Papers in Organic Chemistry" the editor not only has collected solid results (such as the historic first total synthesis of a protein, insulin, in 1965 by a group of Peking chemists "holding aloft the great red banner of Chairman Mao," and the same synthesis with a lower yield by

Brookhaven and German groups at almost the same time) but also has conveyed the sound of the huntsmen's horn as they pursue one of the most remarkable quarries of science: the synthesis of the quick from the dead.

The oldest paper here is a modern classic: Stanley L. Miller's report, now just 20 years old, of how he produced amino acids by sparking water vapor, ammonia, methane and hydrogen. Two other reviews discuss the origin of such biomonomers. The next topic is the assembly of biopolymers synthetically: proteins, nucleic acids and by 1970 a functional gene, the 77-unit double-helical DNA tape specifying a particular transfer RNA in yeast.

Then we have a fascinating interlude. What is the basis of biological activity, given the right polymer sequence? A 1972 jubilee lecture by C. B. Anfinsen tells how he dissects a protein catalyst and promotes its refolding into the naturally active form. Another group of investigators, Sidney Fox's, relates the nature of its proteinoid droplets, which can bud, a little like primitive organisms. Optical activity, that uniform handedness of life given significance by Pasteur himself, is surveyed in an older paper by George Wald, whose last paragraph has the fine remark, "We are the products of editing, rather than of authorship."

The next step is postmolecular. Here we read of the synthesis of organelles, such as the ribosomes, and even of cells in new editions—or better, anthologies—that put together "any desired combination" of cytoplasm, nucleus and membrane taken from other cells. J. F. Danielli is cited here in two papers, a brief technical account and a 1972 speculative essay on the expected impact of such subcellular domestications of new species.

Finally, most difficult and perhaps deepest, are two papers, mainly theoretical, that describe conceptually and mathematically how molecules might by random processes become self-organizing, given a flow of free energy and the tendency to replicate a successful catalytic sequence. The long paper by Manfred Eigen, with a detailed theoretical statement, and another by Hans Kuhn that presents, at a somewhat simpler level of discussion, a model constructed along plausible historical lines close the volume. There are a couple of good textbooks on this topic, but the display of original papers, well selected and given brief context by the editor, adds something irreplaceable to both immediacy and richness.

BIBLIOGRAPHY

Readers interested in further reading on the subjects covered by articles in this issue may find the lists below helpful.

THE FOREIGN MEDICAL GRADUATE

FOREIGN MEDICAL GRADUATES IN THE UNITED STATES. Harold Margulies and Lucille S. Bloch. Harvard University Press, 1969.

FOREIGN TRAINED PHYSICIANS AND AMERICAN MEDICINE. Rosemary Stevens and Joan Vermeulen. Department of Health, Education, and Welfare Publication No. (NIH) 73–324, June, 1972.

WHAT HAPPENS TO FOREIGN-TRAINED PHYSICIANS WHO COME TO THE UNITED STATES? R. A. Stevens, L. W. Goodman and S. S. Mick in *Inquiry,* Vol. 11, pages 112–122; June, 1974.

FOREIGN MEDICAL GRADUATES AND THE MEDICAL UNDERGROUND. R. J. Weiss, J. C. Kleinman, U. C. Brandt, J. J. Feldman and A. C. McGuinness in *The New England Journal of Medicine,* Vol. 290, pages 1408–1413; June 20, 1974.

FOREIGN PHYSICIANS: THEIR IMPACT ON U.S. HEALTH CARE. T. D. Dublin in *Science,* Vol. 185, pages 407–414; August 2, 1974.

COMPUTER-MANAGED PARTS MANUFACTURE

AM ON COMPUTERS. *American Machinist.* McGraw-Hill Book Company, 1971.

CAD/CAM REVOLUTION. George Schaffer in *American Machinist,* Vol. 116, No. 5, page 72; March 6, 1972.

DNC LINES LINK CUTTING TO NEW FUTURE. James E. Sandford in *Iron Age,* Vol. 210, No. 17, pages 46–49; October 26, 1972.

PROGRAMMABLE AUTOMATION: THE BRIGHT FUTURE OF COMPUTERS IN MANUFACTURING. Robert H. Anderson in *Datamation,* Vol. 18, No. 12, pages 46–52; December, 1972.

THE MOST PRIMITIVE OBJECTS IN THE SOLAR SYSTEM

METEORITES. Brian Mason. John Wiley & Sons, Inc., 1962.

METEORITES AND THE ORIGIN OF PLANETS. J. A. Wood. McGraw-Hill Book Company, 1968.

METEORITES AND THE EARLY SOLAR SYSTEM. Edward Anders in *Annual Review of Astronomy and Astrophysics,* Vol. 9, pages 1–34; 1971.

CONDENSATION IN THE PRIMITIVE SOLAR NEBULA. Lawrence Grossman in *Geochimica et Cosmochimica,* Vol. 36, No. 5, pages 597–619; May, 1972.

EARLY CHEMICAL HISTORY OF THE SOLAR SYSTEM. L. Grossman and J. W. Larimer in *Reviews of Geophysics and Space Physics,* Vol. 12, pages 71–101; 1974.

CHROMOSOMAL PROTEINS AND GENE REGULATION

THE BIOLOGY OF ISOLATED CHROMATIN. James Bonner, Michael E. Dahmus, Douglas Fambrough, Ru-chih C. Huang, Keiji Marushige and Dorothy Y. H. Tuan in *Science,* Vol. 159, No. 3810, pages 47–56; January 5, 1968.

THE STRUCTURE AND BIOLOGICAL FUNCTION OF HISTONES. L. S. Hnilica. CRC Press, 1972.

ACIDIC PROTEINS OF THE NUCLEUS. Edited by Ivan L. Cameron and James R. Jeter, Jr. Academic Press, 1974.

NONHISTONE CHROMOSOMAL PROTEINS AND GENE REGULATION. G. S. Stein, T. C. Spelsberg and L. J. Kleinsmith in *Science,* Vol. 183, pages 817–824; March 1, 1974.

CHROMOSOMAL PROTEINS AND THEIR ROLE IN THE REGULATION OF GENE EXPRESSION. Edited by G. S. Stein, J. S. Stein and L. J. Kleinsmith. Academic Press, in press.

DUAL-RESONANCE MODELS OF ELEMENTARY PARTICLES

DUAL RESONANCE MODELS. Paul H. Frampton. W. A. Benjamin, Inc., 1974.

DUAL THEORY. Edited by M. Jacob North-Holland Publishing Co., in press.

AN INTRODUCTION TO THE THEORY OF DUAL MODELS AND STRINGS. J. Scherk in *Reviews of Modern Physics,* in press.

BIOLOGICAL CLOCKS OF THE TIDAL ZONE

THE BIOLOGICAL CLOCK: TWO VIEWS. Frank A. Brown, Jr., J. Woodland Hastings and John D. Palmer. Academic Press, 1970.

THE PHYSIOLOGICAL CLOCK: CIRCADIAN RHYTHMS AND BIOLOGICAL CHRONOMETRY. Erwin Bünning. Springer-Verlag, 1973.

TIDAL RHYTHMS: THE CLOCK CONTROL OF THE RHYTHMIC PHYSIOLOGY OF

MARINE ORGANISMS. John D. Palmer in *Biological Reviews,* Vol. 43, pages 377–418; November, 1973.
BIOLOGICAL CLOCKS IN MARINE ORGANISMS: THE CONTROL OF PHYSIOLOGICAL AND BEHAVIORAL TIDAL RHYTHMS. John D. Palmer. Wiley-Interscience, 1974.

A CARTHAGINIAN FORTRESS IN SARDINIA

MONTE SIRAI: VOLS. I–IV. F. Barreca, S. Moscati et al. Institute of Near Eastern Studies, University of Rome, 1964–1967.
I FENICI E CARTAGINESI IN SARDEGNA. S. Moscati. Utet Editore, Modena, 1973.
LA SARDEGNA FENICIA E PUNICA. F. Barreca. Chiarella Editore, Sassari, 1973.

ALFRED WEGENER AND THE HYPOTHESIS OF CONTINENTAL DRIFT

ALFRED WEGENER. Else Wegener. F. A. Brockhaus, Wiesbaden, 1960.
MEMORIES OF ALFRED WEGENER. J. Georgi in *Continental Drift,* edited by S. K. Runcorn. Academic Press, 1962.
CONTINENTAL DRIFT: THE EVOLUTION OF A CONCEPT. Ursula B. Marvin. Smithsonian Institution Press, 1973.
A REVOLUTION IN THE EARTH SCIENCES: FROM CONTINENTAL DRIFT TO PLATE TECTONICS. A. Hallam. Oxford University Press, 1973.

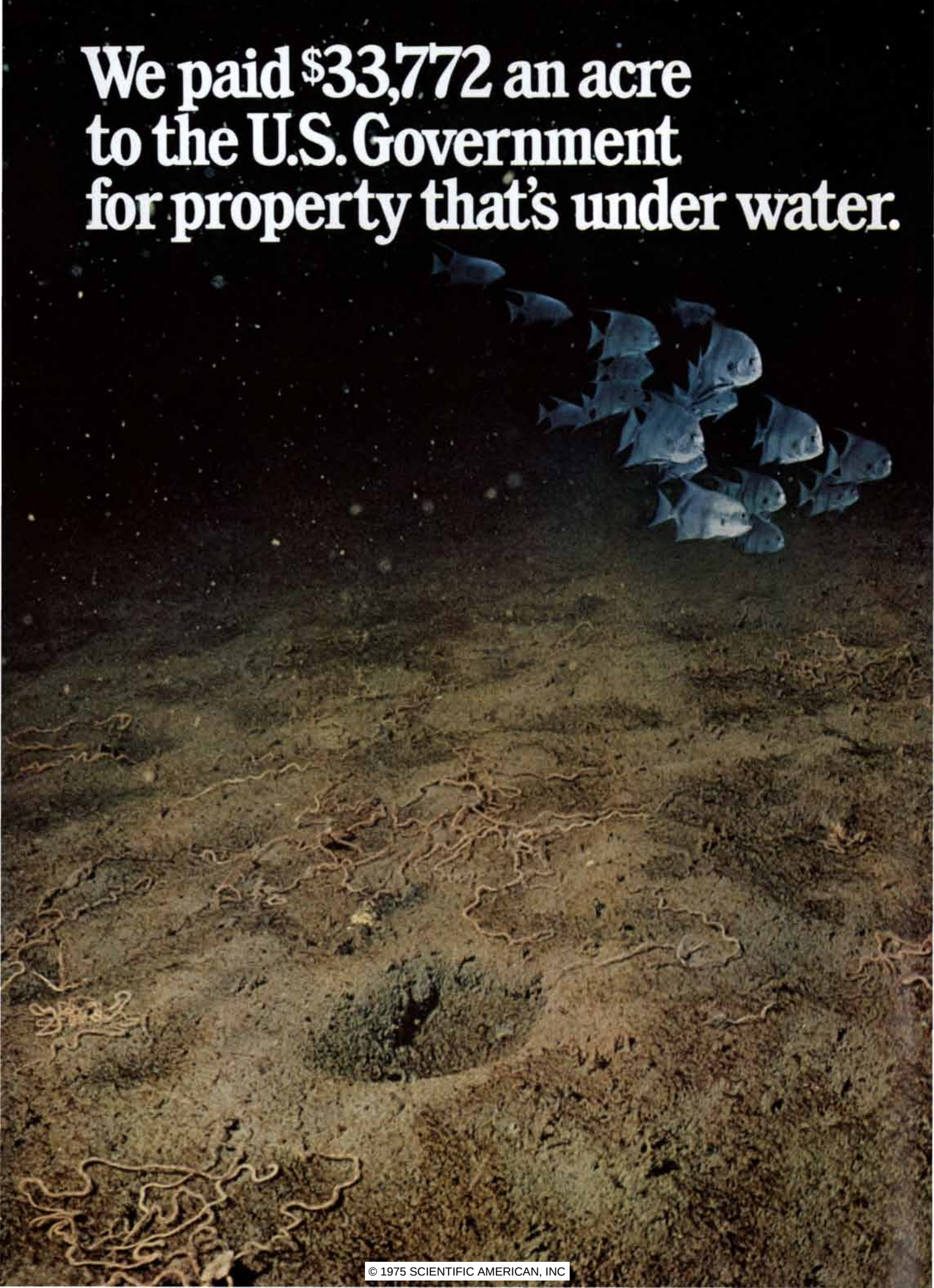
MATHEMATICAL GAMES

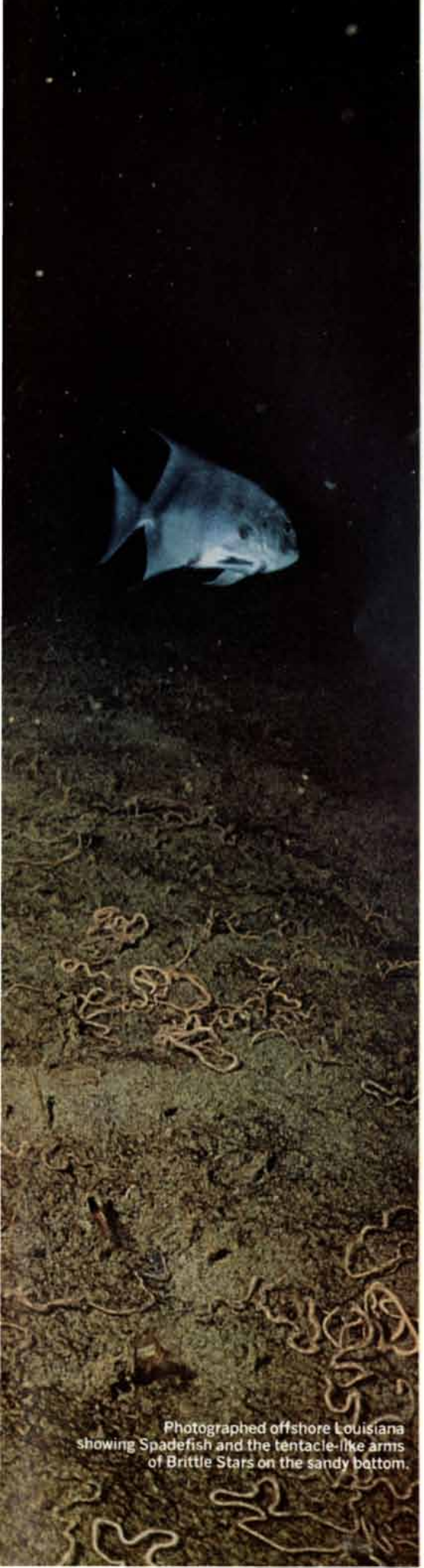
THE ANGUISH OF BEING in *An Introduction to Existentialism.* Robert G. Olson. Dover Publications, 1962.
NIL: EPISODES IN THE LITERARY CONQUEST OF VOID DURING THE NINETEENTH CENTURY. Robert Martin Adams. Oxford University Press, 1966.
NOTHING. P. L. Heath in *The Encyclopedia of Philosophy,* edited by Paul Edwards. The Macmillan Company and The Free Press, 1967.
WHY. Paul Edwards in *The Encyclopedia of Philosophy,* edited by Paul Edwards. The Macmillan Company and The Free Press, 1967.

THE AMATEUR SCIENTIST

HYMAN SERIES IN INVERTEBRATE BIOLOGY: ZOOLOGICAL SCIENCES. VOL. 6: MOLLUSCA 1. Libbie H. Hyman. McGraw-Hill Book Company, 1967.
A SNAIL'S PACE. Lilo Hess. Charles Scribner's Sons, 1974.

We paid $33,772 an acre to the U.S. Government for property that's under water.





Photographed offshore Louisiana showing Spadefish and the tentacle-like arms of Brittle Stars on the sandy bottom.

Photo by Jerry Greenberg.©1974.

Texaco is willing to go almost anywhere to get you the oil you need—even as deep as 1,000 feet under the Gulf of Mexico.

America needs energy. We're working to see that you get it.

But first, we've got to look for it.

And looking for oil requires a lot of looking around, even when you're on dry land.

Imagine how difficult it becomes when you're looking around for it under water—in depths that can go as deep as a thousand feet.

But America does need oil, and experts agree there's plenty of it right here in this country. Buried offshore.

So far, the oil industry has paid the U.S. Government close to 10 billion dollars for offshore leases. Just for the right to look for oil.

Texaco alone, since December 15, 1970, has paid the U.S. Government over 800 million dollars. And has 77 very promising offshore properties—owned fully or partially—to show for it. That's more than 200,000 acres.

With this much underwater acreage, we look forward to finding large new reserves of crude oil, and of natural gas, too, to meet your energy needs.

TEXACO

We're working to keep your trust.

THE SILVER MARTINI.
FOR PEOPLE WHO WANT A SILVER LINING WITHOUT THE CLOUD.

SMIRNOFF SILVER. NINETY POINT FOUR PROOF. SMIRNOFF LEAVES YOU BREATHLESS®

SMIRNOFF SILVER® VODKA. 90.4 PROOF. DISTILLED FROM GRAIN. STE. PIERRE SMIRNOFF FILS. (DIV. OF HEUBLEIN, INC.) HARTFORD, CONN.